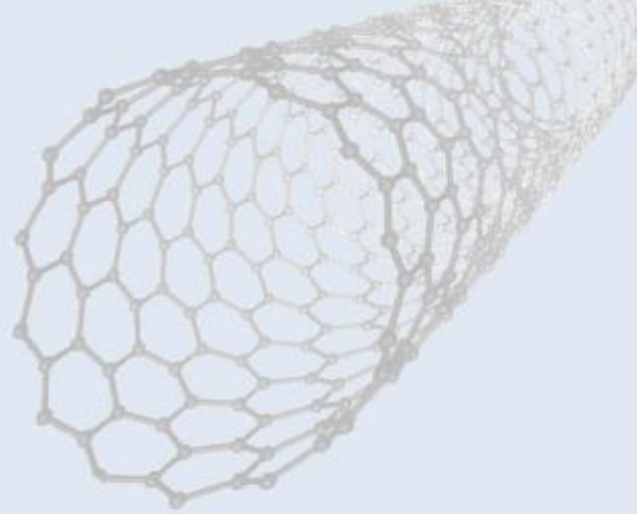
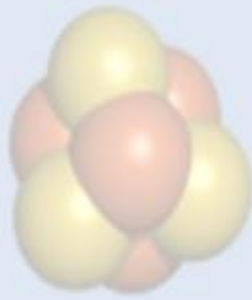


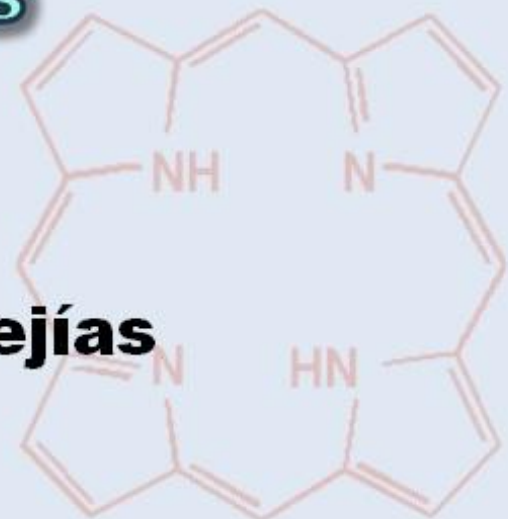
Ph.D thesis



**Engineering repeat proteins as
building blocks for
functional nanostructures and
materials**

**Ingeniería de
proteínas de repetición
como bloques estructurales
para la generación de
nanoestructuras funcionales
y materiales**

**Sara Hernández Mejías
Madrid, 2016**





FACULTAD DE
CIENCIAS
UNIVERSIDAD AUTÓNOMA DE MADRID

instituto
imdea
nanociencia

Ingeniería de proteínas de repetición como bloques estructurales para la generación de nanoestructuras funcionales y materiales

Tesis doctoral presentada por:

Sara Hernández Mejías

Para optar al grado de DOCTOR EN BIOFÍSICA
por la Universidad Autónoma de Madrid

Directora de tesis:

Aitziber López Cortajarena

Madrid, Octubre 2016



FACULTAD DE
CIENCIAS
UNIVERSIDAD AUTÓNOMA DE MADRID

instituto
imdea
nanociencia

Engineering repeat proteins as building blocks for functional nanostructures and materials

Ph.D Thesis presented by:

Sara Hernández Mejías

for the degree of DOCTOR IN BIOPHYSICS
by Universidad Autónoma de Madrid

Thesis supervisor:

Aitziber López Cortajarena

Madrid, October 2016

The results presented in this thesis have been published and listed below:

- 1) Controlled nanometric fibers of self-assembled designed protein scaffolds. **Mejias SH**, Sot B, Guantes R, Cortajarena AL. *Nanoscale*. **6**(19):10982-10988. (Appendix 1).
- 2) Assembly of designed protein scaffolds into monolayers for nanoparticles patterning. **Mejias SH**, Couleaud P, Casado S, Granados D, García MA, Abad JM, Cortajarena AL. *Colloids Surf B Biointerfaces*. 2016. **141**:93-10. (Appendix 2).
- 3) Repeat protein scaffolds: ordering photo-and electroactive molecules in solution and solid state. **Mejias SH**, Andarias JL, Sakurai T, Yoneda S, Erazo KP, Seki S, Atienza C, Martín N, Cortajarena AL. *Chem Sci*. 2016. **7**: 4842-4847. (Appendix 3).
- 4) AuNPs conjugated with CTPR16-4Cys for AuNPs patterning by protein. **Mejias SH**, Couleaud P, Martin-Lasanta A, Extremiana C, Romera D, Abad JM, Casado S, Latorre A, Rodríguez M, Gonzalez T, Cortajarena AL. *In preparation*.
- 5) Porphyrin-protein conjugates and SWCNTs: a captivating partner. Andarias JL, **Mejias SH**, Sakurai T, Matsuda W, Seki S, Atienza C, Martín N, Cortajarena AL. *In preparation*.

Table of contents

Summary	1-7
Resumen	9-15
Introduction	
I.1. Hierarchically organized natural systems	19-20
I.2. Nanotechnology: inspiration from the Nature	20-21
I.3. Self-assembly of DNA and peptides for mimicking natural structures	21-23
I.4. Proteins as scaffold for self-assembly	23-25
I.5. Understanding protein structure	25-28
I.6. Repeat proteins	29-31
I.7. TPR repeat proteins: ideal scaffold for bottom-up design	31-33
General objectives of the thesis	35-37
1. Chapter 1: CTPR protein design for protein self-assembly	
1.1. Motivation	41-46
1.2. Background	47-50
1.3. Objectives	51-54
1.4. Results and discussion	
1.4.1. Controlled nanometric fibers of self-assembled CTPR	
1.4.1.1. Motivation and objectives	57
1.4.1.2. Results and discussion	57-65
1.4.1.3. Conclusions	65
1.4.2. Tightly packed CTPR monolayers	
1.4.2.1. Motivation and objectives	66
1.4.2.2. Results and discussion	66-74
1.4.2.3. Conclusions	74-75
1.4.3. CTPR nanotubes: 3D structure	
1.4.3.1. Motivation and objectives	76
1.4.3.2. Results and discussion	77-89
1.4.3.3. Conclusions	89-90

1.5.	Chapter 1: conclusions	91-94
2.	Chapter 2: CTPR protein based bio-hybrid functional structures	
2.1.	Motivation	97-103
2.2.	Background	105-109
2.3.	Objectives	111-114
2.4.	Results and discussion	
	2.4.1. Templating of gold-nanoparticles by CTPR proteins	
	2.4.1.1. Motivation and objectives	117-118
	2.4.1.2. Results and discussion	
	2.4.1.2.a <i>Performing AuNPs monolayers templated by CTPR proteins</i>	118-121
	2.4.1.2.b <i>Performing AuNPs conductive films templated by CTPR proteins</i>	122-130
	2.4.1.3. Conclusions	130
	2.4.2. Generation of electron donor-acceptor systems	
	2.4.2.1. Motivation and objectives	131-132
	2.4.2.2. Results and discussion	
	2.4.2.2.a <i>Organizing porphyrins in solution and solid state</i>	132-143
	2.4.2.2.b <i>Generation of different sizes CTPR-porphyrin arrays</i>	143-146
	2.4.2.2.c <i>Donor-acceptor nanohybrids based on CTPR and CTPR-porphyrin conjugates and SWCNT</i>	146-157
	2.4.2.3. Conclusions	158
	2.4.3. Mimicking natural photosystems	
	2.4.3.1. Motivation and objectives	159-161
	2.4.3.2. Results and discussion	161-169
	2.4.3.3. Conclusions	169
2.5.	Chapter 2: conclusions	171-173
	General conclusions	175-178
	Conclusiones generales	179-182
	Materials and methods	
M.1.	Chapter 1. CTPR protein design for protein self-assembly	
	M.1.1. Controlled nanometric fibers of self-assembled CTPR	185-187
	M.1.2. Tightly packed CTPR monolayers	188-191
	M.1.3. CTPR nanotubes: 3D structure	191-194

M.2. Chapter 2. CTPR protein based ordered bio-hybrid functional structures

M.2.1. Templating of gold-nanoparticles by CTPR proteins 195-202

M.2.2. Generation of electron donor-acceptor systems 202-210

M.2.3. Mimicking natural photosystems 210-214

References 215-230

Appendix 1. Controlled nanometric fibers of self-assembled designed protein scaffolds

Appendix 2. Assembly of designed protein scaffolds into monolayers for nanoparticles patterning.

Appendix 3. Repeat protein scaffolds: ordering photo-and electroactive molecules in solution and solid state.

Resumen

Resumen

La generación de nano-dispositivos y materiales funcionales para aplicaciones en nanotecnología y biomedicina requiere el control preciso de elementos funcionales en diferentes escalas de tamaño. Para definir las propiedades finales de los materiales y nano-dispositivos es necesario controlar la disposición de los elementos a diferentes escalas de tamaños: nanométrica, micrométrica y macrométrica. Sin embargo, la tecnología actual carece de la precisión necesaria para obtener ese control y, normalmente, la generación de dispositivos con propiedades definidas se hace mediante “prueba y error”. El mayor obstáculo para el desarrollo de estas tecnologías de precisión es el poco conocimiento a nivel atómico que se tiene de los materiales que se utilizan.

Para intentar solucionar este problema, últimamente se está investigando en la generación de materiales mediante el proceso de auto-ensamblaje. El proceso de auto-ensamblaje se basa en interacciones específicas entre elementos pequeños y simples para formar estructuras supramoleculares complejas. La simplicidad de sus elementos hace más sencillo conocer y controlar el sistema.

En esta tesis, se utiliza el proceso de auto-ensamblaje para crear estructuras funcionales basadas en proteínas en las que los elementos están ordenados en diferentes escalas de tamaños. Las proteínas son moléculas muy versátiles y robustas que tienen un gran potencial en la generación de estructuras y materiales. Sin embargo, las interacciones entre proteínas no son fáciles de controlar debido a la gran cantidad de interacciones que hay que tener en cuenta para definir sus estructuras. No obstante, las proteínas de repetición ofrecen una nueva visión en el diseño de proteínas para la generación de estructuras ordenadas funcionales debido a su simplicidad. Las proteínas de repetición están compuestas por un número variable de unidades básicas que se repiten a lo largo de la proteína y su plegamiento está dominado por interacciones regulares de corto alcance (Figura R1). Estas características hacen que el sistema sea más sencillo de entender y de controlar.




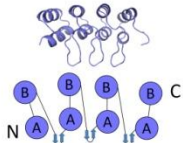


	ANK	CTPR	LRR
Unidad básica de repetición			
Proteínas de repetición			

Figure R1. Representación de diferentes proteínas de repetición. Para cada familia de repetición se muestra, en la primera fila de la tabla, la estructura y la representación esquemática de cada unidad básica que componen las diferentes proteínas de repetición: en azul el módulo “ANKyrin repeat” (ANK), en naranja el módulo “Tetratricopeptide Repeat” con la secuencia consenso (CTPR) y en verde el módulo “Leucin Rich Repeat” (LRR). En la segunda fila de la tabla, la estructura cristalina de las proteínas de repetición compuestas por 4 repeticiones de cada unidad básica, utilizando el mismo código de colores que en la fila de arriba. Estas estructuras muestran diferentes empaquetamientos que dependen de la unidad básica de repetición que las componen. Las estructuras se acompañan de esquemas que representan el empaquetamiento en dirección N-terminal C-terminal de las diferentes estructuras cristalinas.

Concretamente, en esta tesis se utilizan las CTPR, basadas en la secuencia consenso del módulo “Tetratricopeptide Repeat”, para la generación de estructuras funcionales. La estructura y la estabilidad de las CTPR están muy bien caracterizadas. Las CTPR tienen únicamente unos pocos aminoácidos conservados que definen su plegamiento característico. Los demás aminoácidos admiten variaciones sin perturbar la estructura de estas proteínas. Esta característica de las CTPR aporta flexibilidad a la hora de introducir nuevas reactividades a lo largo de la proteína para generar proteínas modificadas para diversas aplicaciones.

Esta tesis consta de dos capítulos principales. En el capítulo 1, “Diseño de proteínas CTPR para la generación de estructuras auto-ensambladas”, se explica el diseño, generación y caracterización de diferentes estructuras supramoleculares modificando, de forma controlada, el proceso de auto-ensamblaje de las CTPR. En el capítulo 2, “Estructuras funcionales bio-híbridas basadas en las CTPR”, se explica la generación de estructuras funcionales usando como molde las proteínas CTPR. Para ello, se introducen diferentes reactividades a lo largo de la proteína con resolución atómica donde se van a conjugar diferentes elementos activos. Una vez tenemos formados los conjugados entre la proteína y los elementos activos, se generan las estructuras funcionales mediante el auto-ensamblaje de dichos conjugados.

Capítulo 1. Diseño de proteínas CTPR para la generación de estructuras auto-ensambladas.

En este capítulo, se explica la generación y caracterización de estructuras ordenadas formadas a partir del auto-ensamblaje de proteínas CTPR. En concreto, se generan estructuras de diferentes dimensionalidades: nanofibras de proteína, monocapas 2D compactas y nanotubos de proteínas 3D (Figura R2).

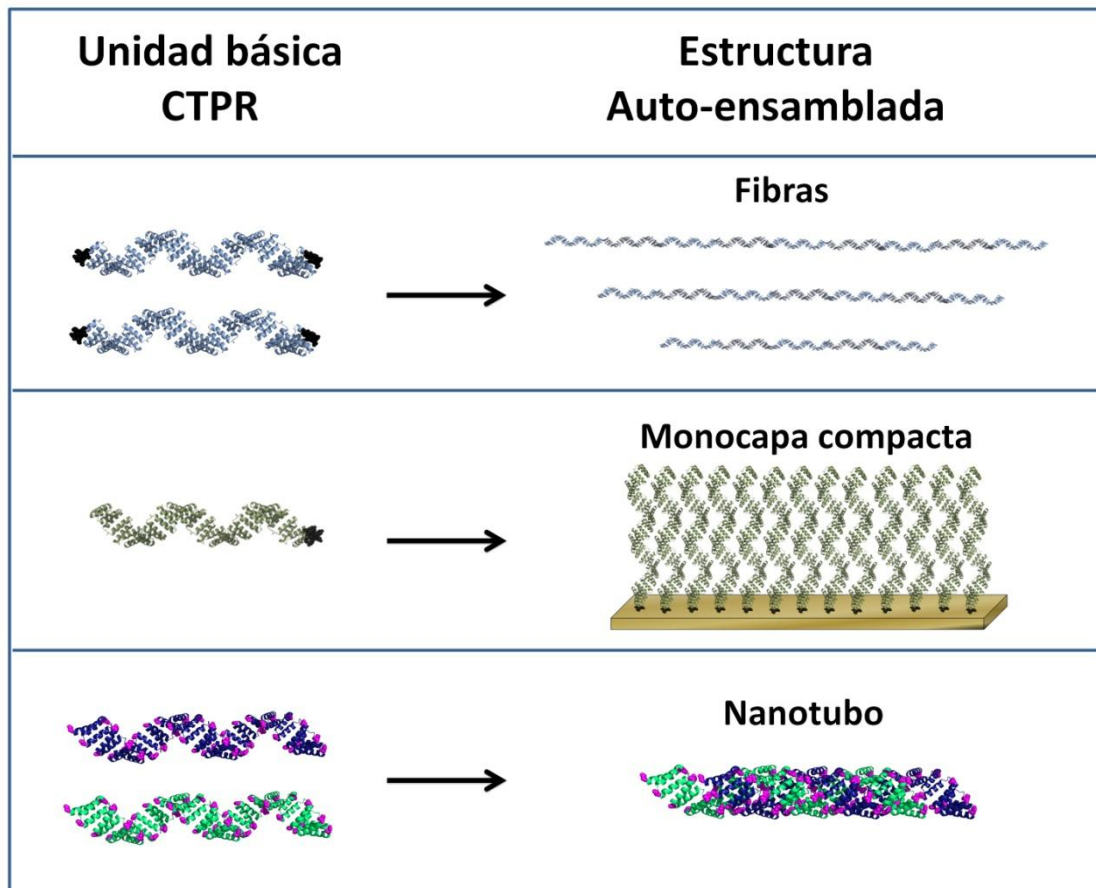


Figura R2. Diferentes tipos de estructuras auto-ensambladas formadas a partir de la modificación de las interacciones entre las proteínas CTPR. A la izquierda, se representa la proteína CTPR señalando la modificación que da lugar a la interacción entre proteínas para la generación de la estructura auto-ensamblada. A la derecha, se representa la estructura obtenida debido a esa interacción entre las proteínas. De arriba hacia abajo: fibras, monocapas compactas y nanotubos de proteína.

Las proteínas CTPR son modificadas para codificar interacciones específicas entre ellas para la generación de las diferentes estructuras. En el capítulo se diseñan y se sintetizan los diferentes variantes de proteína que dan lugar a las diferentes estructuras. Además, se caracteriza, en detalle, el proceso de auto-ensamblaje de los diferentes variantes y las propiedades biofísicas de las estructuras obtenidas.

Capítulo 2. Estructuras funcionales bio-híbridas basadas en las CTPRs

En este capítulo, se generan estructuras funcionales bio-híbridas usando como molde las proteínas CTPR. Se explota el conocimiento a nivel atómico de la estructura de las CTPR para introducir reactividades a lo largo de la proteína (Figura R3) para usarla como molde para ordenar elementos activos a escala nanométrica. Los elementos activos que se ordenan son: partículas de oro, moléculas dadoras-aceptoras, nanotubos de carbono, y clústeres metálicos (Figura R3.B).

Una vez se obtiene el control en el proceso de conjugación, se combina con el control adquirido en el capítulo 1 de generar las estructuras auto-ensambladas para generar estructuras bio-híbridas donde los elementos activos están ordenados a diferentes escalas de tamaños: escala nanométrica, micrométrica y macrométrica. Como ejemplo, se generan monocapas de partículas de oro, films ordenados de partículas de oro y films ordenadas de moléculas dadoras-aceptoras. Además, se genera un sistema multifuncional imitando los fotosistemas naturales (Figura R3.C). Las estructuras funcionales generadas tienen propiedades que dependen de la disposición de los elementos activos en el sistema.

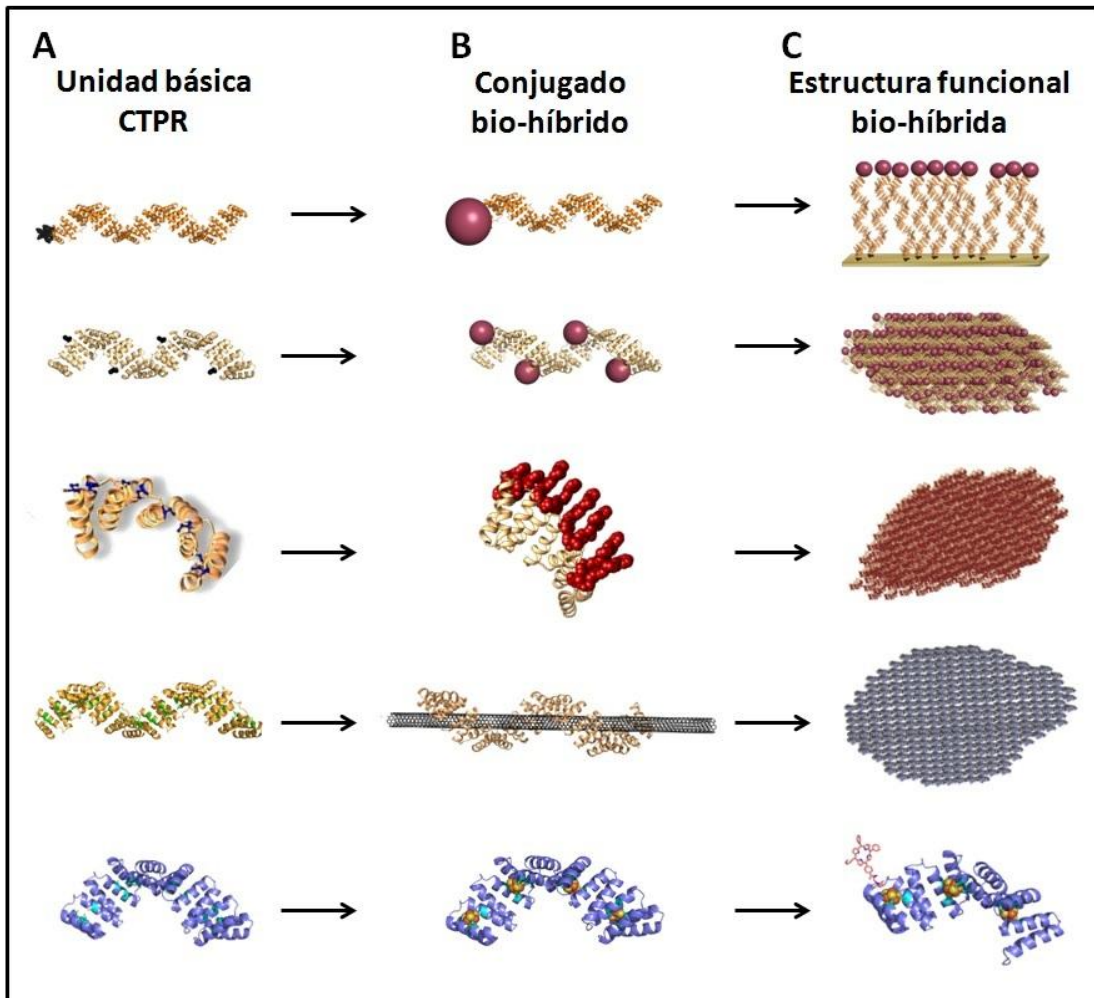


Figura R3. Estructuras funcionales bio-híbridas generadas utilizando como molde las proteínas CTPR. A) Proteínas CTPRs modificadas donde se resaltan las diferentes reactividades que se han metido a lo largo de la estructura de la proteína. B) El conjugado obtenido entre la proteína y el elemento activo que se une de forma específica a las diferentes reactividades metidas en la proteína. De arriba abajo: partículas de oro, porfirinas, nanotubos de carbono y clústeres [4Fe4S]. C) Estructura auto-ensamblada generada a partir de los conjugados.

Los estudios realizados en esta tesis, establecen las bases para la generación de futuros materiales y estructuras funcionales con diferentes morfologías y propiedades basadas en proteínas de repetición con aplicaciones en diferentes campos de la nanotecnología.

Summary

Summary

The precise synthesis of nano-devices and materials with tailored complex structures and properties is a requisite for their use in nanotechnology. The development of these devices and materials requires methods to control the order and structure along different size scales. However, nowadays, the technology for the generation of hierarchically organized materials lacks of precision to determine their structure and properties in a control manner, and is accomplished mostly by “trial and error” experimental approaches. The major obstacle in the development of this technology is the scant understanding of the atomic structure of biopolymer materials that hinders the selectively introduction of reactive functionalities. In this sense, bottom-up self-assembly that relies on highly specific biomolecular interactions of small and relatively simple components is an attractive approach to generate ordered materials, because it offers facile means for organizing molecules into complex supramolecular structures.

In this thesis, a self-assembly bottom-up strategy is used to generate ordered functional materials based on proteins. Proteins have long been recognized as the most versatile and robust building blocks with the great potential in the generation of novel materials and nanostructures. However, this versatility comes with the cost of complexity. The countless number of non-covalent interactions makes the protein systems hard to control in a predictive manner. In this sense, repeat proteins are promising scaffolds for protein design due to their structural simplicity. Repeat proteins are composed by a variable number of tandem repeats of a basic structural motif, being their structure is dominated by short-range regular interactions (Figure S1). These characteristics make the system easier to understand and to control.




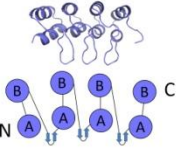
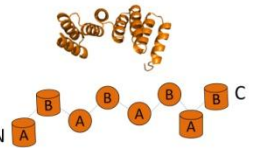

	ANK	CTPR	LRR
Building blocks			
Repeat proteins			

Figure S1. Representation of repeat protein scaffolds. For each repeat protein family, in the first line of the table, the structure of an individual repeated unit is shown together with a schematic representation of each building block: ANKyrin repeat in blue (ANK), TetratricoPeptide Repeat (TPR) in orange, and Leucin Rich Repeat (LRR) in green. In the second line of the table, the crystal structures of repeated proteins composed by 4 repeats of each building block are represented using the same color code as in the top line. The structures illustrate the different packing arrangements between the building blocks as it is also displayed in the schematic representations of the packing from N-terminal to C-terminal of the proteins below the crystal structures.

Specifically, in this thesis CTPR repeat proteins are used for the generation of functional structures. The structure and stability of the CTPR proteins are very well-characterized. These CTPR proteins present only few highly conserved amino acids, being involved in intra- and inter-repeat packing interactions that maintain the CTPR fold. The amino acids at the other positions admit variations, giving the flexibility to introduce novel functionalities such as different chemical reactivities and ligand-binding specificities.

This thesis contains two main chapters. The chapter 1, “*CTPR protein design for protein self-assembly*”, explains the generation and characterization of different supramolecular structures modifying the self-assembly properties of the CTPR proteins. The chapter 2, “*CTPR protein based bio-hybrid functional structures*”, explains the formation of functional structures using the CTPR proteins as template. For this purpose, different reactivities are introduced along the CTPR protein structure with atomic resolution to conjugate desired active elements. Once the conjugates are formed functional supramolecular structures are generated by self-assembly of the hybrid conjugates.

Chapter 1. *CTPR protein design for protein self-assembly*

In this chapter, the generation and characterization of higher order structures through hierarchical self-assembly of CTPR proteins is presented. In particular, structures with

different dimensionalities have been generated: protein fibers, tightly packed 2D protein monolayers and 3D protein nanotubes (Figure S2).

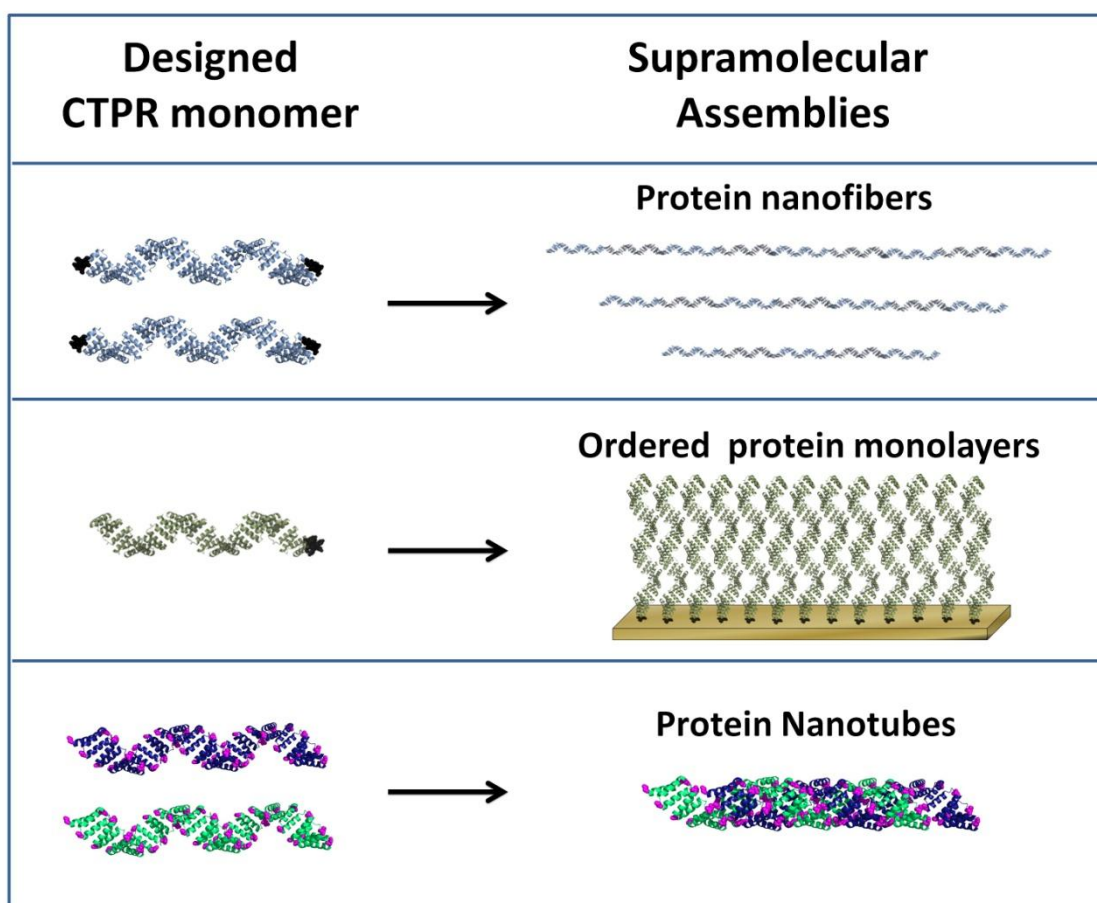


Figure S2. Different types of protein assemblies formed by CTPR proteins by the design of novel interactions between the proteins. On the left column the modified CTPR protein structure highlighting the modifications introduced in the CTPR protein. In the right column, the assembly formed from the novel interactions. From top to the bottom: Protein nanofibers, tightly packed monolayers, protein nanotubes.

The CTPR proteins are modified to encode novel interactions between them. These interactions lead the generation of different geometries, based on the structure of the CTPR proteins. Different variants of proteins are designed and synthesized and the self-assembly properties of the different variants are characterized.

Chapter 2. CTPR protein based bio-hybrid functional structures

In this chapter, the generation of bio-hybrid functional structures using as a template CTPR proteins is presented. The atomic-resolution knowledge of the CTPR protein structure is exploited to specifically introduce reactivities along the proteins (Figure S3) to use them as a template to organize at nanoscale different active elements including gold

nanoparticles, donor-acceptor molecules, carbon nanotubes, and redox-active clusters (Figure S3.B).

Once the control over the conjugation process is obtained, it is combined with the control over the assembly of CTPR proteins that is described in the Chapter 1 to form bio-hybrid structures where the active elements are organized in different range of scales: nano, micro and macroscale. As examples, monolayers of gold nanoparticles, films of gold nanoparticles and films of donor-acceptor molecules are formed. Also, a multifunctional system that mimics natural photosystems is presented (Figure S3.C). The generated functional structures have different properties depending on the disposition of the active elements.

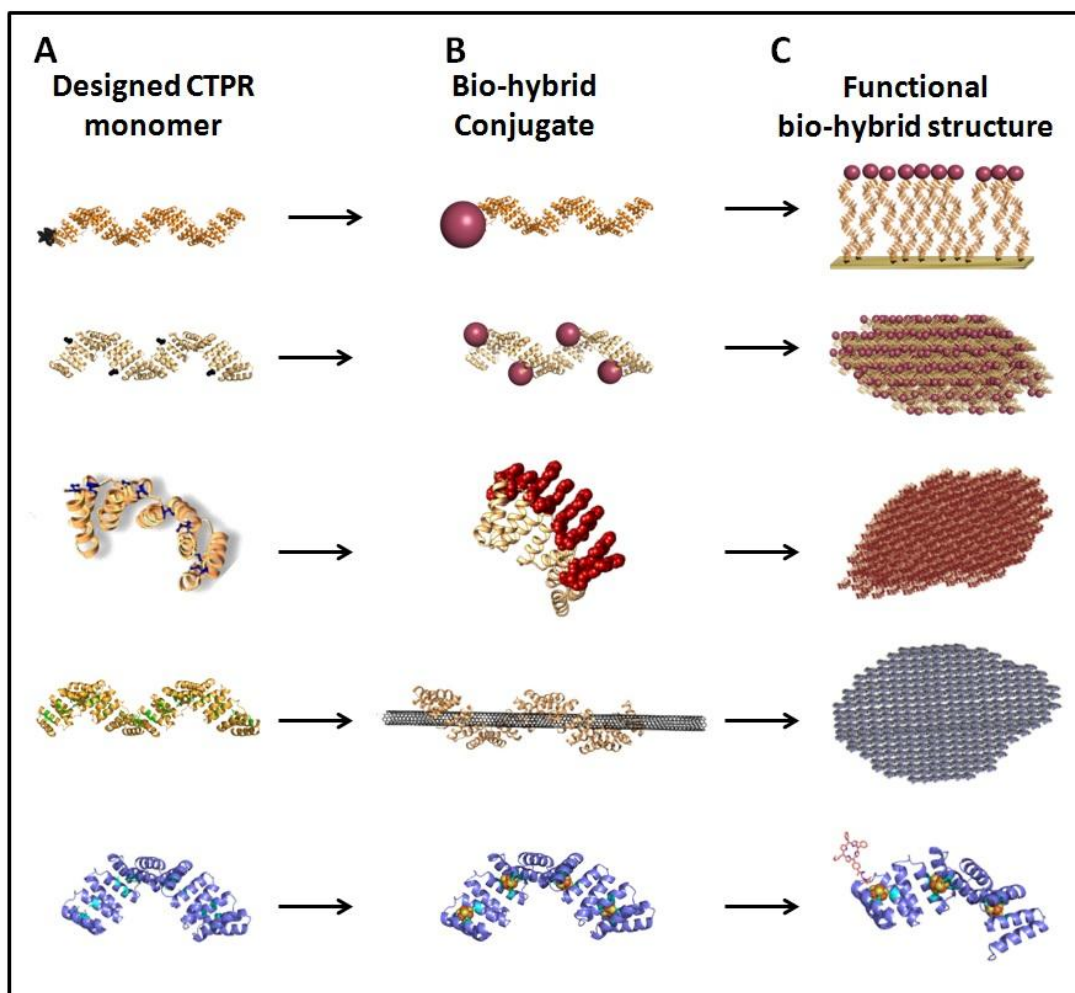


Figure S3. CTPR as templates to generate functional bio-hybrid structures. A) Modified CTPR protein structure highlighting the modifications introduced in the CTPR protein. B) Conjugation between the protein and different active components. From top to the bottom: gold nanoparticles, porphyrins, carbon nanotubes, and [4Fe4S] clusters. C) Assembly of the conjugates to form different functional hybrid structures based on CTPR proteins.

The studies performed in this thesis project establish the basis for the generation of future functional materials and structures with different morphologies and properties based on repeat proteins with many potential applications in nanotechnology.

Introduction

Introduction

I.1. Hierarchically organized natural systems

Since the life started on the earth, 3500-3800 million of years ago, natural selection and evolution have developed a huge amount of biological materials in different environments. These materials offer incredible properties that have attracted the human interest such as light-weight, high-toughness [1], mechanical efficiency [2], flexible-switch attaching and detaching [3], self-cleaning properties [4].

The growing possibility of studying the materials at the nanoscale gives the chance of understanding better the materials that Nature offers. Nowadays, it is known that many of these interesting natural materials have structural hierarchy. This fact means that the materials are made by small modules organized at different length scales into defined structures where the organization of components defines the properties of the material. There are plenty of examples of these materials in Nature. For example, gecko lizard has the ability to walk on walls and ceilings thanks to some hair-projections on the feet that contain nano-seized structures that act as a dry adhesive (Figure 1.A) [5, 6]. Lotus leaf contains minuscule crystals, just a nanometer across, that act as integrated vacuum cleaners. The crystals help water droplets to carry dust and dirt off the surface, keeping the leaf clean at all times (Figure 1.B) [7]. Nacre is a strong and durable material, which structure is formed by aragonite platelets and organic matrices and exhibit two-level crossed lamellar micro-architectures (Figure 1.C). The function of the platelets is to increase the structural stiffness and hardness, whereas, the function of organic matrices between layers is to control the nucleation and growth of the inorganic phases in the biomineralization process of these structures [8]. These are some of a large list of examples of self-assembled and hierarchically organized materials existing in Nature [9].

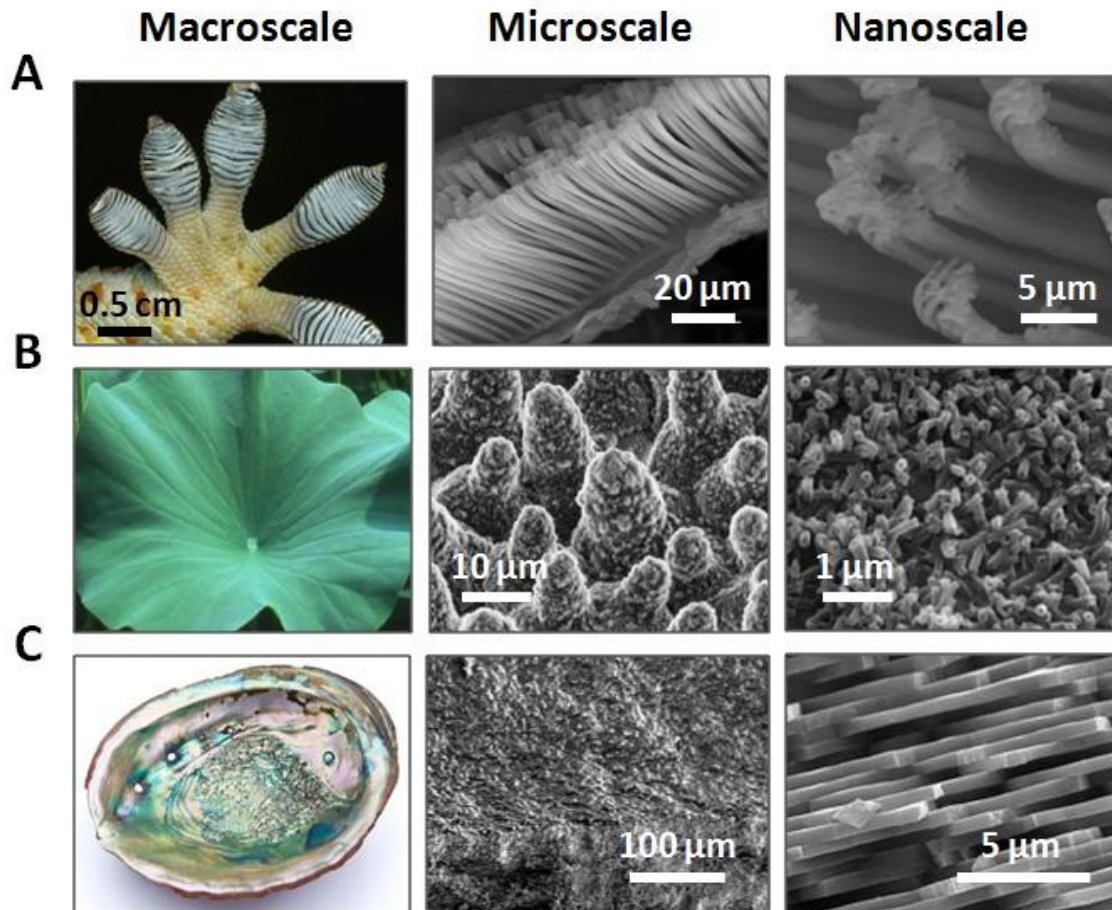


Figure 1. Example of different natural material with hierarchical organization. Organization at macroscale, microscale and nanoscale. A) Gecko foot at different magnifications. Figure from [10]. B) Lotus leaf at different magnifications. Figure from [11]. C) Nacre structure at different magnifications. Figure from [12].

These biological materials possess many inspiring properties and fascinating mechanisms that encourage to investigate them deeply and broadly. Understanding the effects of hierarchical structures can guide the synthesis of new materials with physical properties that are tailored for specific applications. As we learn more about Nature, we also find ways to integrate these new discoveries into modern technologies. The ability of mimicking Nature could lead to an array of new technologies and devices.

1.2. Nanotechnology: Inspiration from Nature

Nowadays, there is a growing interest in manipulating matter in a controlled manner, to create sophisticated devices for applications in different fields. The precise synthesis of nano-devices and materials with tailored complex structures and properties it is a requisite for their use in nanotechnology.

Nanotechnology is defined as the manipulation and, more importantly, the control of matter on an atomic, molecular and, consequently, supramolecular scale. Indeed, the final aim of nanotechnology is to create new materials and devices for applications in different fields including medicine, electronics, biomaterials, energy production, and consumer products [13].

However, nowadays the technology for the generation of these devices and materials still lacks of precision to determine their structure and properties in a control manner, and it is accomplished mostly by “trial and error” experimental approaches. In this way, Nature has much to teach nanotechnology. As explained before, in Nature there are materials with very interesting properties in which the constitutive elements are organized hierarchically. So, it is not surprising that many efforts in this research area are directed to mimic the natural materials regarding structure, properties or function.

In order to mimic the hierarchical organization observed in Nature, it is essential to control the structure of the elements in the material at different length scales. For this purpose, a bottom-up approach is a good approximation in which the properties of small elements are used to form more complex structures. Bottom-up self-assembly offers facile means for organizing molecules into complex supramolecular structures [14-16]. This approach represents an extraordinary source of innovation with strong potential impact in material sciences [17].

1.3. Self-assembly of DNA and peptides for mimicking natural structures

Self-assembly, in the context of material science, implies that the components are programmed to create ordered functional structures. Most of the research in this area has been focused on the design and creation of relatively simple small molecules that act as building blocks. Those molecules are provided of reactive groups that lead the interaction between the components in a specific way to form organized structures. Specifically, the most promising results have been obtained using nano-sized biological entities as building blocks to engineered devices as substrates with desired physicochemical and biomimetic properties [18-21].

Many of the current examples of self-assembly based on biomolecules exploited for nanotechnology are based on DNA origami. DNA is the carrier of the genetic information in cells composed by only four deoxynucleotide chemical building blocks. Because of the high chemical stability and predictable folding of DNA its assembly properties are easy to control by rational design [22]. DNA nanotechnology refers to the construction of supramolecular objects with robust topological or geometrical features from DNA components. The great potential of the control of the DNA architectonics has been

reflected by the variety of two and three dimensional shapes and patterns with sizes from 20 to 200 nm generated by DNA [23]. In this strategy, a long single stranded DNA scaffold is folded with complementary oligonucleotides that act as staples. Once synthesized and mixed, the staples and scaffold strands self-assemble in one single step. Some examples of possible structures that can be programmed using DNA origami are shown in Figure 2 [24].

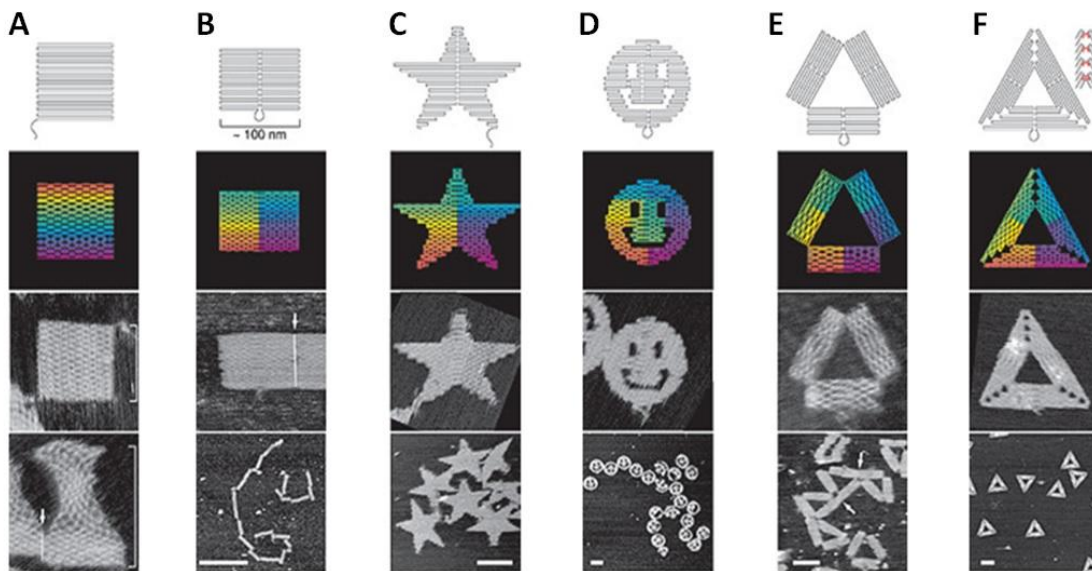


Figure 2. Nanoshapes created by DNA origami. A) Square, B) rectangle, C) star, D) disk with three holes, E) triangle with rectangular domains, and F) sharp triangle with trapezoidal domains. Figure from [23].

DNA origami has been used to create nanomechanical devices with considerable success [25, 26]. However, functionalization of nucleotide-based nanostructures is still challenging [27]. The major obstacle in the development of DNA-based templates is the fact that DNA assemblies are non-covalent and the post-assembly functionalization may destroy the structure of the system. In addition, there is a lack of understanding of the atomic structure of the final DNA-based materials.

In order to solve the aforementioned functionalization problems, peptides are increasingly investigated as building blocks. Peptides are small chains of amino acids (50 or fewer) linked by peptidic amide bonds. The control in solid-phase peptide synthesis enables to introduce different reactive groups in the peptide chain in a specific manner. The self-assembly of short designed peptides into molecular nanostructures, mimicking natural self-assembly processes, is becoming a growing interest in nanobiotechnology. There are a lot of examples where peptides are used as a building blocks for nanostructures [28, 29] and functional bio-materials (Figure 3) [30-33]. A particular

emphasis is placed on the versatility of peptide self-assembly in terms of modularity, responsiveness and functional diversity.

Despite of the advances, it is not possible to easily control the order of peptide-based materials at different length scales (nano, micro and macroscale) because of their small size and the lack of stable three dimensional structure. Also, peptides are expensive to synthesize at large scale, that is a limitation to use them as nanotechnological tools [34].

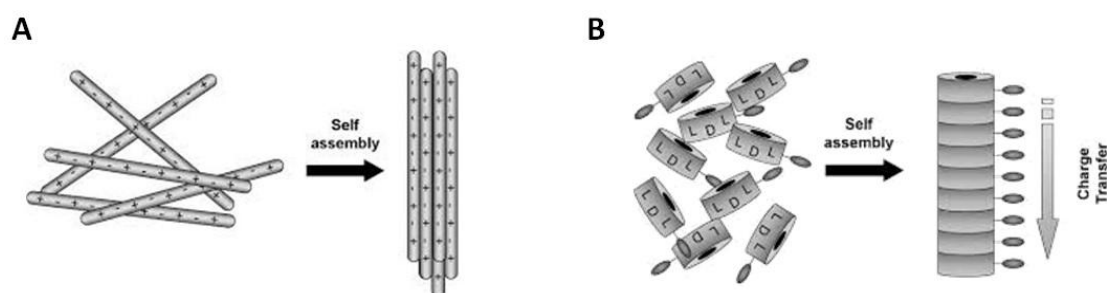


Figure 3. Peptides as building blocks to form nanostructures. A) Self-assembly of peptides by charge complementarity. The design of peptide building blocks with opposing charges allows the efficient self-assembly of peptide monomers into well-ordered structures at the nano-scale. These peptide-based structures can be used to form macroscopic hydrogels that have nano-scale order. B) Self-assembly of peptide nanotubes. These tubes are based on the precise assembly of alternating D- and L-amino acid cyclic peptides to form elongated and hollow nanotubes. Decoration of the external parts of the nanotubes with functional moieties makes it possible to engineer them into functional nanoassemblies. For example, decoration with aromatic moieties that allow charge transfer is envisioned for applications in molecular electronics. Figure from [35].

1.4. Proteins as scaffolds for self-assembly

As described in the previous section, it is still a challenge to control the organization of active components over different length scales using DNA and peptides. However, from the study of natural systems it is clear that the more complex examples of self-assembled structures are made by proteins [36]. Proteins are organized in structures of different sizes from nanometers to millimeters. There are a lot of examples of protein assemblies in Nature where the properties of the final material rely on the organization of the proteins components. For example, viral capsids consist of different proteins that self-assemble to form cages with different structures and sizes from 10 nm to hundreds of nanometers (Figure 4.A). These capsids have different mechanical properties depending on the assembly of the proteins [37]. In the bacterial flagellum, depending on the assembly of the proteins, the obtained structures have different specific functions such as an ion driven motor, which can rotate in either direction; the hook, which transmits the rotation from the motor to the filament; and the filament, a very long structure of several micrometers which acts as a propeller and behaves differently depending on the direction

of the motor turn (Figure 4.B) [38]. The spider silk is a macroscopic fibrous biomaterial made mainly by proteins. Fibers also contain sugars, lipids, ions, and pigments that might affect the aggregation of the proteins and act as a protection layer of the structure. In the spider silk, the proteins are organized at the nanoscale forming crystalline regions that are separated by amorphous linkages (Figure 4.C) [39]. The disposition of the proteins in the assemblies provides very interesting properties to the final material, having a tensile strength comparable to steel and yet more elastic than rubber. Because of the large number of examples of natural protein assemblies in which the elements of the systems are organized at different length scales, there is a growing interest in using proteins as scaffolds to form hierarchical materials.

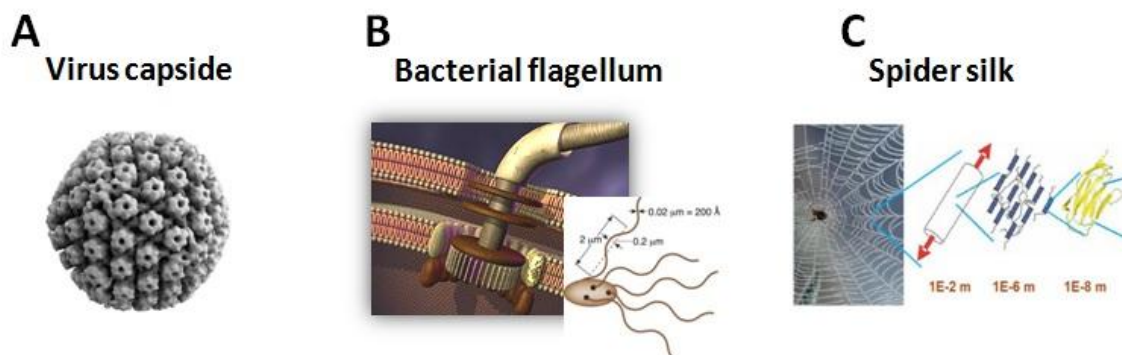


Figure 4. Hierarchically organized proteins at different length scales. A) 3D reconstruction of Herpes simplex virus capsid (EMDB ID: 3288). The micrographs for the reconstruction were acquired using cryoTEM. The capsid is made by several proteins organized in an icosahedral shell of 125 nm in diameter. B) Representation of the bacterial flagellum located in the cell membrane. The flagellum is composed of about 30 different proteins with copy numbers ranging from a few to a few thousands. The complex structure is made by self-assembly of those proteins. Inset shows a representation of the whole bacteria highlighting the size of the filament. C) From left to right, the representation of the spider silk at different magnifications, showing the organization of the molecules in the silk at different length scales (*Image: Dr. Buehler, MIT*).

Apart from the versatility to create structures made by self-assembly, in Nature, proteins act as scaffolds to locate different kind of active components, such as metals, fluorophores, peptides or even DNA [40-44]. There are a lot of proteins in this category, such as metalloproteins that contain metal cofactors [45]. An example of metalloprotein is hemoglobin, which is the principal oxygen carrier in humans and has four sub-units with a heme group in which the iron (II) ion is coordinated (Figure 5.A). Another example is nitrogenase, which performs the fixation of atmospheric nitrogen. The nitrogenase complex consists of two metalloproteins: Fe protein, which contains one $[\text{Fe}_4\text{S}_4]$ cluster and MoFe protein, which contains two iron-sulfur clusters, $[\text{Fe}_8\text{S}_7]$ known as P-clusters and MoFe cofactor. Fe protein transfers the electrons from the reducing agent to the P-clusters, which then transfer the electrons to the MoFe cofactor. Each MoFe cofactor then acts as a site for nitrogen fixation, with N_2 binding in the central cavity of the

cofactor (Figure 5.B). These are some examples of the many bio-hybrid systems that exist in Nature. Moreover, there are some protein-based systems in which the self-assembly properties of the proteins are combined with their ability to conjugate other active components to form a bio-hybrid systems that are very interesting systems for nanotechnology [46].

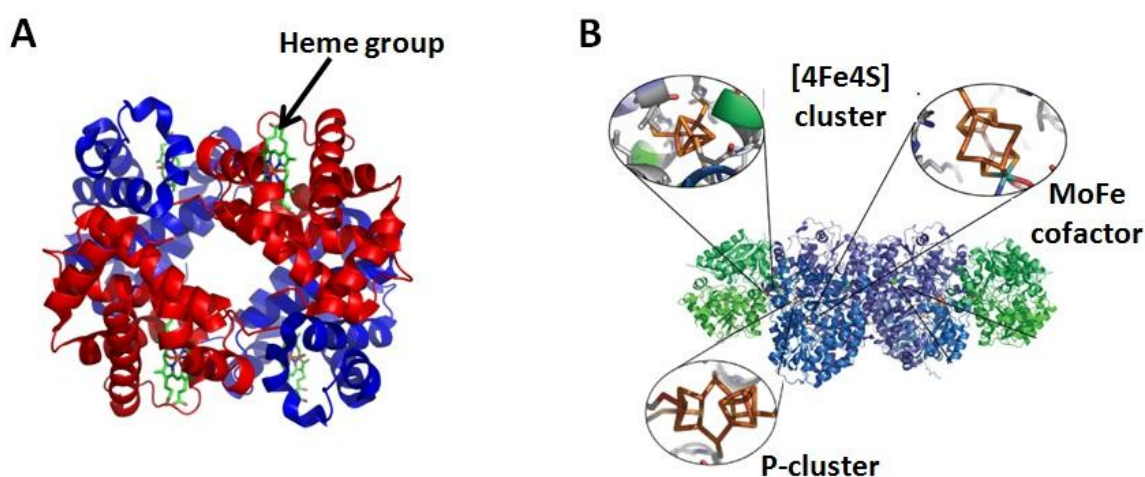


Figure 5. Examples of metalloproteins. A) Hemoglobine protein structure (PDB ID: 1A3N). In red and blue are colored the 4 protein sub-units alternatively. In green is colored the heme group where the iron (II) is coordinated. B) Nitrogenase complex structure (PDB ID: 1N2C). In green is colored the Fe protein, where is highlighted the [4Fe-4S] cluster and in blue is colored the MoFe protein where is highlighted the MoFe cofactor and P-clusters.

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 in the generation of novel materials and nanostructures [47]. Proteins offer enormous diversity in three dimensional architectures and lots possibilities for modification and functionalization [48]. However, this versatility comes with the cost of complexity. The countless number of non-covalent interactions makes the protein systems hard to control in a predictive manner and it is still a challenge to design protein-based systems, despite decades of research and undeniable progress.

1.5. Understanding protein structure

Proteins are linear sequences of α -amino acids that are linked together by peptide bonds, a covalent bond between the carboxyl group of one amino acid and the amino group of another. The existence of an amino group (N-terminal) at one end of the chain and a carboxyl group (C-terminal) at the other end imposes directionality to the chain.

The peptide chain itself is called the backbone and the chemical functionalities associated with each amino acid are known as the side chains. The conformation of polypeptide backbones is restricted by steric clashes between backbone and side-chain atoms [47, 49, 50]. Thus the linear sequence of amino acids determines the secondary structure and, consequently, the three dimensional structure or fold of the protein (Figure 6).

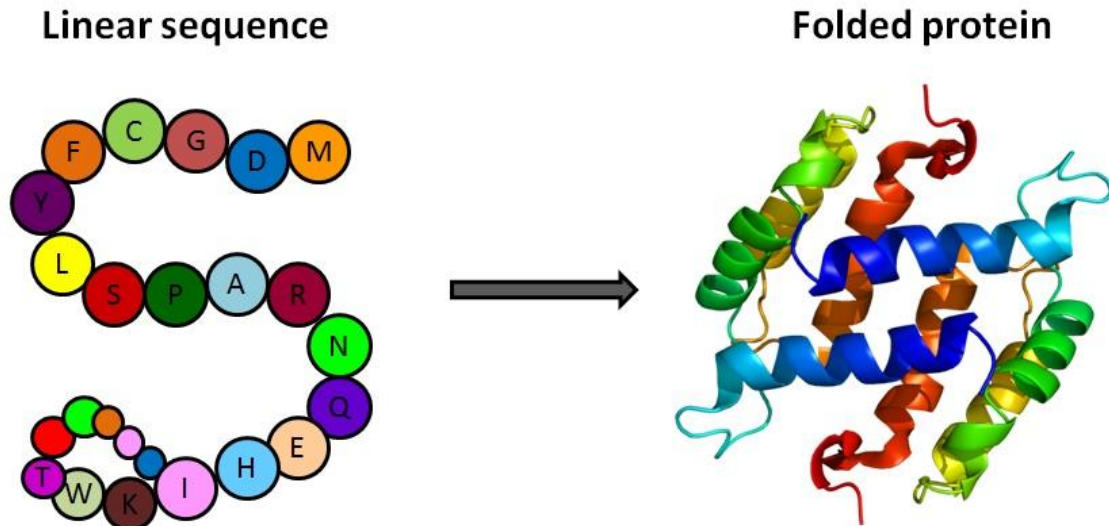


Figure 6. The linear sequence of amino acids defines the three dimensional structure of a folded protein. On the left, a linear sequence of amino acids that form the primary structure of a protein. On the right, representation of the three dimensional structure of a folded protein that depends on the linear sequence of amino acids.

Taking into account steric restrictions, the two major possibilities of secondary structure in proteins are α -helix, where the conformation is obtained through internal hydrogen bonding along the polypeptide chain, and β -stands, where internal hydrogen bonding is not possible. Two or more β -strands form extended structures that are called β -sheets (Figure 7.A). With the appropriate placement of hydrophobic and polar side chains along the polypeptide sequence, both secondary structures have distinct hydrophobic polar faces that guide the folding of the polypeptides to form the tertiary structure (Figure 7.B). Once a protein is folded can function as it is or it may interact with other proteins or cofactors such us carbohydrates, vitamins or minerals forming the quaternary structure (Figure 7.B.C).

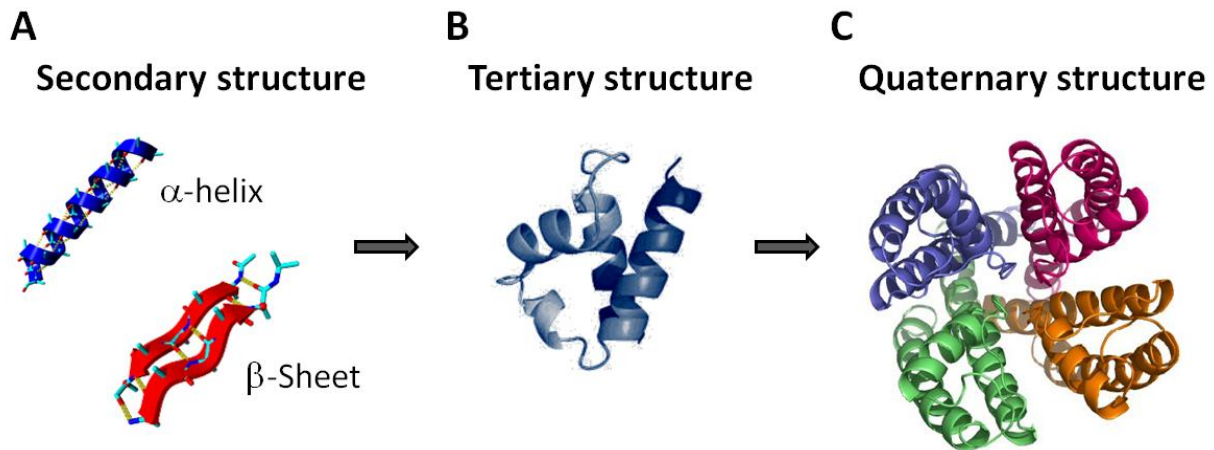


Figure 7. Protein structure. A) The two main secondary structure elements found in proteins: α -helix and β -sheet. B) The secondary structure elements interact with themselves in a particular arrangement to form the tertiary structure. C) A protein that consists on one amino acid strand, once folded, can interact with other proteins or cofactors such us carbohydrates, vitamins or minerals forming the quaternary structure.

In the last decade, our understanding on how to manipulate the structure of the proteins to create artificial constructs with properties has increased exponentially [51]. Proteins can be produced on an industrial scale from bacterial expression systems relatively cheaply. Proteins are codified by DNA in living systems. The DNA sequence encoding the protein of interest is usually introduced in the bacteria through a plasmid for bacterial expression. Plasmids are circular DNA fragments that generally occur in bacteria. While the chromosomes are big and contain all the essential genetic information for living under normal conditions, plasmids usually are small, with sizes from 3 to 10 kbp. Plasmids contain additional genes that provide information that allows the survival of the organism, such as antibiotic resistance.

Artificial plasmids are used in molecular biology as basic tools for expressing proteins in bacteria. The plasmid with the gene of interest is transformed into the bacteria. The bacteria carrying the plasmid will produce the recombinant protein using their protein expression machinery (Figure 8).

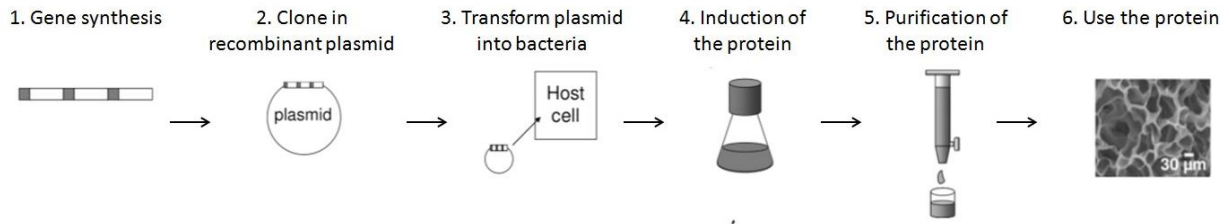


Figure 8. Recombinant protein production. 1. Synthesis of gene of interest. 2. Cloning the gene in a recombinant plasmid. 3. Transformation of the plasmid into bacteria. 4. Bacterial expression of the protein. 5. Purification of the protein. 6. Use the purified protein for specific application.

Nowadays, the utilization of bacteria for producing proteins is widely extended. Recombinant proteins can be made with great accuracy, reproducibility and sequence control. Thus, taking advantages of the control in the expression of recombinant proteins and using the knowledge that we have on protein science, up to some extent it is possible to rationally design proteins to create desired structures and functions, with good stability in different environments.

Combined advances in experimental and theoretical methods to understand protein materials increasingly open up potential applications in material design and nanotechnology [52]. Proteins have a number of characteristics that lend themselves to be useful materials for a range of applications: they are synthesized under ambient conditions, without toxic product unlike other methods of nanofabrication [48], and do not require a clean room. The future for protein nanotechnology seems bright. However, the vast potential of proteins in nanofabrication reveals the difficulty that exists nowadays to rationally redesign or *de novo* design proteins with defined properties and features to form a functional material.

The main limitation in the rational protein design is the limited understanding about how protein sequence-structure-function relate. Thus, it is critical to understand the fundamental principles that underlie protein structure, stability and function, and how the structure and function of the proteins are defined by their sequence. Many efforts have been dedicated in the last decades in the fields of protein folding and protein design to reach the current stage at which protein design has already achieved some success including the design of new protein folds [53], and enzymatic activities [54]. However, in the emerging field of nanomaterials based on designed proteins there are only few promising works from the protein design perspective [55, 56] and it is still a challenge to use proteins in a controlled way.

I.6. Repeat proteins

The control of protein structure, stability and function is not easy because of the number of interactions that are involved in defining the protein fold. In this sense, repeat proteins, are interesting scaffolds for protein design because, due to the simplicity of the system, it is possible to control how protein sequence-structure-function relate in these proteins. Repeat proteins are non-globular structures involved in essential cellular processes acting typically as scaffolds for the mediation of protein–protein interactions. Repeat proteins are composed by a variable number of tandem repeats of a basic structural motif from 18 to 47 amino acids, and their structure is dominated by short-range and regular interactions [57, 58].

There are a variety of repeat protein families composed of units with different structures, being α -helix, β -strand or a mixture of these two secondary structures. Some of the more abundant and well-studied classes of repeat proteins are formed by the repetition of the following building blocks: tetratricopeptide repeats (TPR) which consists of 34 amino acid sequence that folds in helix-turn-helix motif [59], ankyrin repeats (ANK) which consist of 33 amino acid sequence that folds in helix-loop-helix motif [60], leucine rich repeats (LRR) which consist of 20-30 amino acids that fold in a beta-turn-helix motif (Figure 9) [61], armadillo repeats (ARM) [62], and transcription activator-like (TALE) [63]. As shown in Figure 9, the repeated units form elongated structures with different twists due to the distinct packing between the units leading to structures with different shapes.

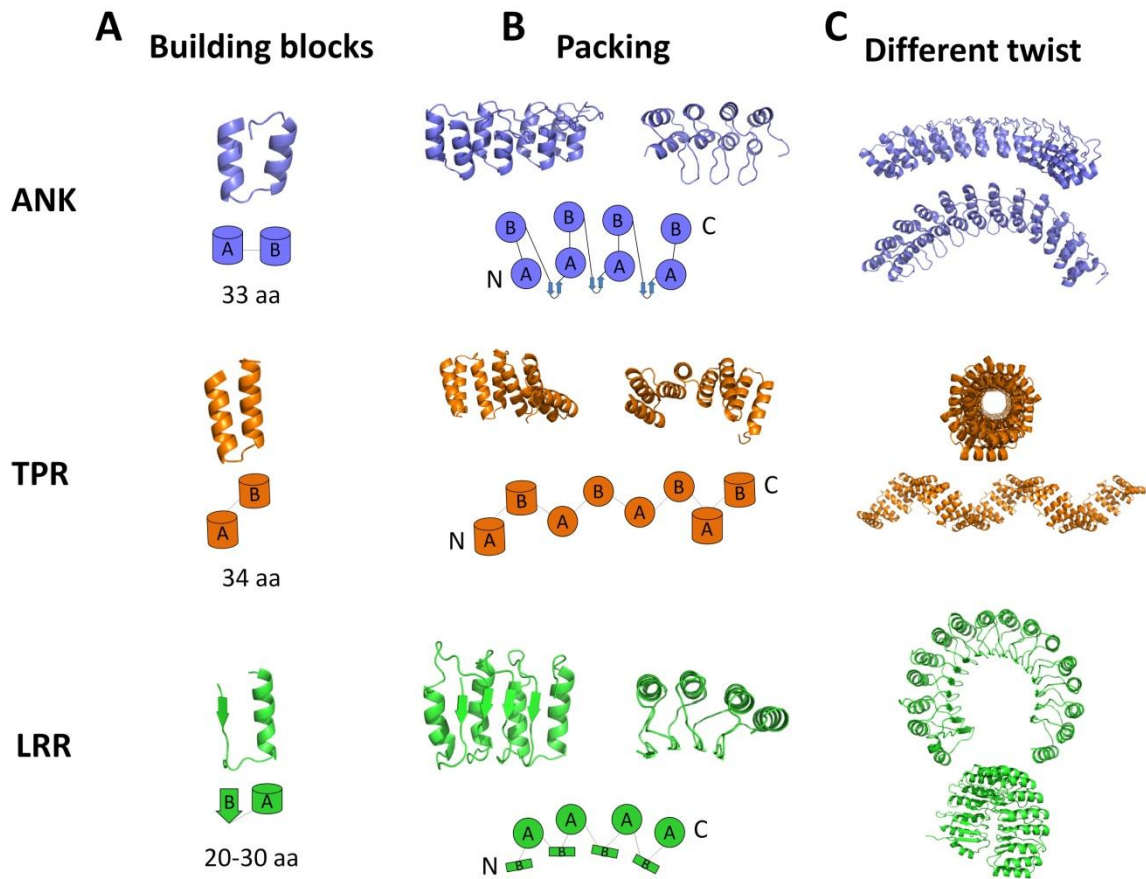


Figure 9. Representation of repeat protein scaffolds. A) For each repeat protein family, the structure of an individual repeated unit is shown together with a schematic representation of each building block: ANKyrin repeat in blue (ANK), Tetratricopeptide Repeat (TPR) in orange, and Leucin Rich Repeat (LRR) in green. B) The crystal structures of repeated proteins composed by 4 repeats of each building block are represented using the same color code as in the panel A (front view on the left side and top view on the right side). The structures illustrate the different packing arrangements between the building blocks as is also displayed in the schematic representations of the packing from N-terminal to C-terminal of the proteins below the crystal structures. C) Crystal structures of long repeat arrays. A repeat protein formed by 12 ANK repeats (PDB ID: 2XEE); a repeat protein formed by 20 TPR repeat (PDB ID: 2AVP); and a repeat protein formed by 16 LRR repeat (PDB ID: 1A4Y). Depending on the packing of the building blocks the repeat proteins shows different twist and therefore different shapes within the assemblies.

The modular structure of repeat proteins simplifies the design problem to the level of simple units and to the interactions between the neighboring units, which are local and predictable. Stability, length, binding affinity and specificity of the modules are defined by design. Thus, each repeat unit can be used as a building block with individually engineered properties (stability, function, and interactions between modules). Individual modules can be combined in a rational way to generate different protein molecules, depending on the application. The understanding and control over the structure and stability of the repeat proteins open the door to use them as scaffolds to generate self-assembled functional structures in which it would be possible to fix the position of the functionalities in a specific manner [64-66].

Considering the main features of repeat proteins previously described it is evident that they are ideally suited for protein design and nanobioengineering. Indeed, some recent works confirm the level of understanding on those repeated systems, in which is possible a rational computational design to engineer proteins with different properties that expand the sequence and structure space observed in Nature [67-69].

1.7. TPR repeat proteins: ideal scaffold for bottom-up design

As introduced in the previous section, TPR repeat consists a 34 of amino acid helixA-turn-helixB motif (Figure 10.A) [59]. In Nature, TPRs occur in tandem repeat arrays and the role is to mediate protein-protein interactions in cells and to assemble of multi-protein complexes. To create new TPR proteins that capture the sequence-structure relationship, a consensus TPR (CTPR) sequence was designed by Dr. Regan's group (Yale University) from the statistical analysis of natural TPRs (Figure 10.A/bottom) [70]. CTPR sequence presents only few highly conserved small and large hydrophobic amino acids, being involved in intra- and inter-repeat packing interactions that maintain the TPR fold [59, 71, 72]. The amino acids at the other positions admit variations, giving the flexibility to introduce novel functionalities such us different chemical reactivities and ligand-binding specificities [73, 74]. In Nature, TPRs occur in tandem repeat arrays, from 3 to 20. Similarly, CTPR repeats can be combined in tandem (Figure 10.B) to form CTPR proteins that present a continuous right-handed superhelical structure with eight repeats per one full turn of the superhelix (Figure 10.C) [70].

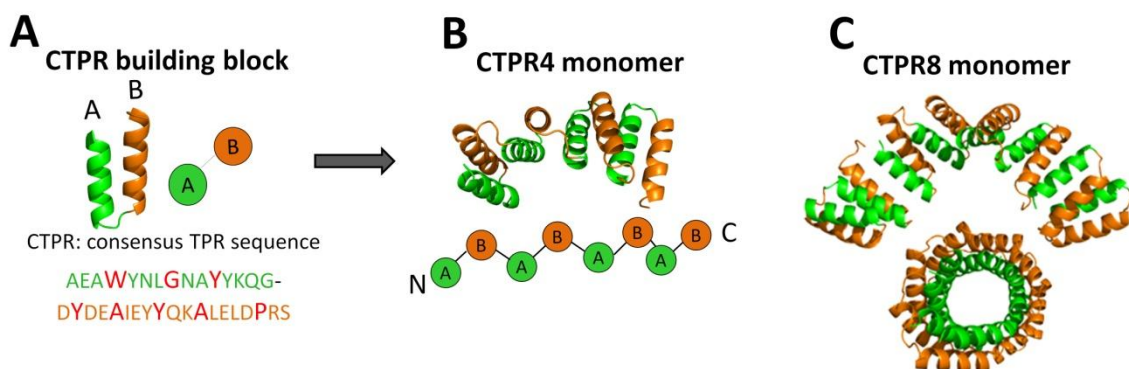


Figure 10. CTPR repeat as a building block for repeat proteins. A) CTPR repeat unit structure is represented showing helix A in green and helix B in orange. On the right, a schematic representation of the structure of the CTPR building block using the same color code. Below, it is shown the consensus sequence highlighting in red the conserved amino acids. B) The crystal structure of a repeat protein composed of 4 CTPR repeats using the green color for the A helices and orange for the B helices. Below, it is shown the schematic representation of the CTPR packing from N-terminal to C-terminal. C) The crystal structure of a repeat protein composed of 8 CTPR repeats in two different orientations, showing the full turn of the superhelix. The green color is used for the A helices and orange for the B helices (PDB ID: 2AVP).

The structure and stability of CTPR proteins are very well characterized [75]. In fact, the crystal structures of CTPR proteins were obtained with different number of identical repeats (Figure 11.A) [76]. It is demonstrated that CTPRs are more thermodynamically stable than their natural counterparts, which make them more tolerant to the destabilizing effects of mutations. Furthermore, thermodynamical stability of CTPR proteins can be predicted using theoretical models (Figure 11.B). These studies make possible to modulate the stability of the CTPR proteins in a predictable manner by changing either the sequence of the repeats or the number of repeated units [77, 78].

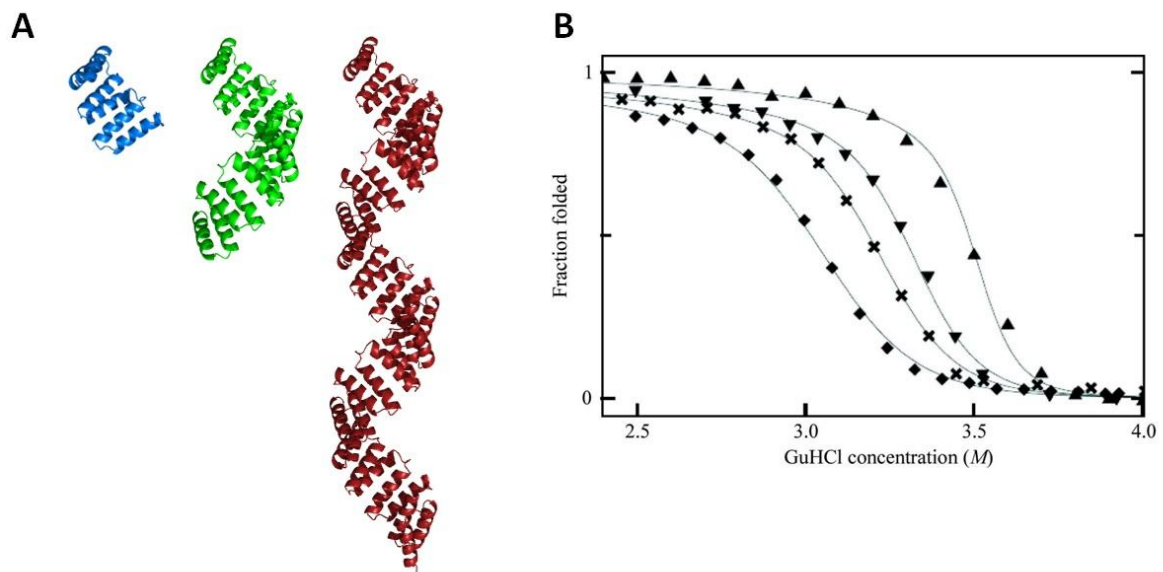


Figure 11. Characterization of the structure and stability of the CTPR proteins. A) Crystal structure of the CTPR proteins with different number of repeats: CTPR3, in blue, CTPR8 in green, and CTPR20 in red. B) Denaturation curves of the CTPR6 (squares), CTPR8 (cross), CTPR10 (inverted triangles) and CTPR20 (triangles) proteins as a function of GuHCl monitored by circular dichroism and fit to the theoretical model (lines). Figure from [76].

The aforementioned properties confirm the understanding over the system and allow a good control at the molecular level. This knowledge opens the door to design a variety of modified CTPR modules with desired amino acid composition for selected applications (Figure 12.A). By combining these building blocks into long arrays, it could be possible to generate proteins with modules that encode different structural and functional properties (Figure 12.B).

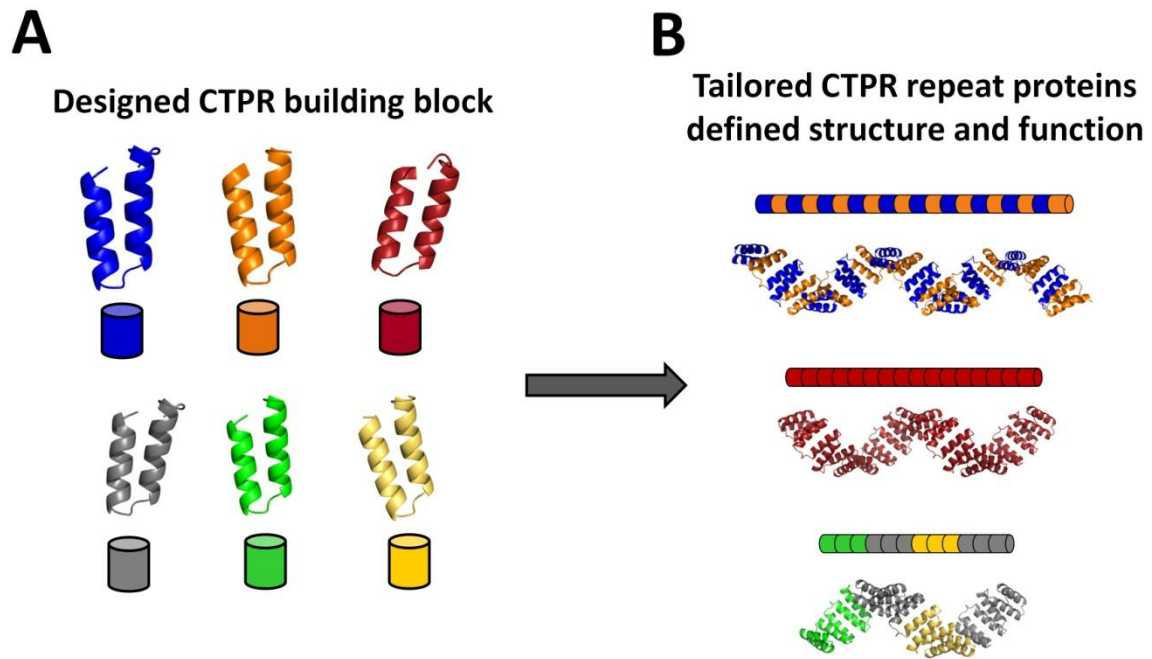


Figure 12. Tailored CTPR repeat proteins formed by modified CTPR-based building blocks. A) CTPR-based repeat units with variable sequences and properties are represented in different colors. Below: schematic representation of the blocks as cylinders. B) Examples of the formation of CTPR proteins by combination of different CTPR repeat units. On the top, schematic representation of the protein and below the three dimensional models based on the structure of CTPR8 and CTPR20 (PDB ID: 2AVP). From the top to the bottom: CTPR20 protein formed by 20 repeats of two different CTPR units colored in blue and orange alternatively; a CTPR16 formed by 16 identical CTPR repeats; a CTPR12 protein formed by 12 repeats of three different CTPR variants combined in groups of three.

General objectives

General objectives

In this thesis, we explore the possibility of using repeat proteins as nanotechnological tools. Specifically, the main objective of this work is using proteins as scaffolds to form functional materials in which the molecules are organized at different length scales: nano, micro and macroscale. Because of the knowledge and control over the stability and structure of the CTPR proteins, we use this class of repeat protein as model system to develop this main objective. We will use a bottom-up strategy inspired in natural systems where the elements of the system self-assemble to form a hierarchical organized materials. To achieve the final objective it is necessary to have a full control of the protein system. Thus, we define 3 specific objectives in order to accomplish the general objective of the work:

1. Control of the self-assembly process of the CTPR proteins. We aim to design proteins that encode novel interactions between the molecular units in a controlled manner that will lead to the formation of different protein structures.
2. Use the CTPR proteins as scaffold to organize different active elements at the nanoscale. We aim to introduce functionalities at defined positions in the CTPR proteins for the conjugation of different active elements with particular patterns and controlled distances along the structure with atomic resolution.
3. Formation of functional structures. We aim to combine the control over the assembly process with the control over the functionalization to generate structures with defined properties, in which the elements are organized at different length scales: nano, micro and macroscale.

The specific objectives are covered in two main chapters. The first chapter covers the first specific objective describing the use of CTPR proteins to form structures where the elements self-assemble in a controlled way to form higher order structures such as fibers, protein monolayers, and protein nanotubes. The second chapter covers the second and the third specific objectives. The chapter describes the introduction of reactive moieties in the CTPR proteins, taking advantage of the structural knowledge and the control over the stability of the template. Modified CTPR proteins are used as scaffolds to organize active elements with different properties such as gold nanoparticles, donor-or-acceptor molecules and clusters at nanoscale. Finally, once the hybrid conjugates are obtained, by taking advantage of the control over the assembly and the control in the functionalization of the CTPR proteins functional bio-hybrid structures and materials will be created. These functional nanostructures and materials have defined properties such as conductivity, photoactivity and electrochemical activity, which depend on the order of the elements.

Chapter 1.

CTPR protein design for protein self-assembly

1.1. Chapter 1: Motivation

1.1. Chapter 1. Motivation

As we explained before, in Nature, proteins provide examples of complex structures such as large arrays of proteinaceous materials including viral capsids, spider silk, and the bacterial flagellar motor. The complexity and sophistication of protein-based structures and materials in Nature hints to the great potential of designed protein-based materials and nanostructures [47, 79, 80]. Therefore, there is a growing interest in generating novel protein structures by exploiting the same set tools and interactions that govern protein structures in Nature [81].

There are a lot of examples in Nature, in which complex protein structures have evolved through combinations or self-assembly of smaller independently folded domains [55, 82, 83]. Self-assembly is a spontaneous process of organization of molecular units into ordered structures as a result of intra-molecular and/or inter-molecular interactions [84]. Self-assembly relies on highly specific biomolecular interactions. Thus, using bottom-up approaches based on these interactions is an attractive strategy to design complex structures from simple molecular units [85].

However, up to date, most examples described on self-assembly are based on peptides, especially in the generation of fibrillar materials (Figure 13). Filamentous assemblies are usually classified in two main groups: α -helix-based and amyloid-like assemblies, as detailed below. On one hand, the designs based on interactions of alpha-helical peptides are usually obtained from *de-novo* sequences. These designs usually rely on association of amphipathic helices in the alpha helical coiled-coil motif [24]. These structures can be engineered to form offset dimers with complementary sticky ends to promote longitudinal assembly into fibers. Sequence-to-structure relationship tends to be better defined for these kinds of assemblies (Figure 13.A). On the other hand, the designs based on amyloid-like peptides can be obtained from naturally occurring and designed sequences [86]. The design of amyloid-like fibers relies on the general tendency of *beta*-strands to aggregate. There are few examples in which interactions at the molecular-level can be prolonged to a macroscopic material using these assemblies [22]. The application of this approach results in materials that have a hierarchy along different length scales: nanometer ordering within the fibrils and micrometer-scale ordering in the stacking of the fibrils into films (Figure 13.B). Moreover, these peptide-based nanostructured films possessed mechanical properties similar to that of the most rigid proteinaceous materials found in Nature, such as keratin or collagen. The downside of amyloid-like assemblies is that the assembly is not specific and can not be modified in a controlled way, since all the sequences form similar assemblies.

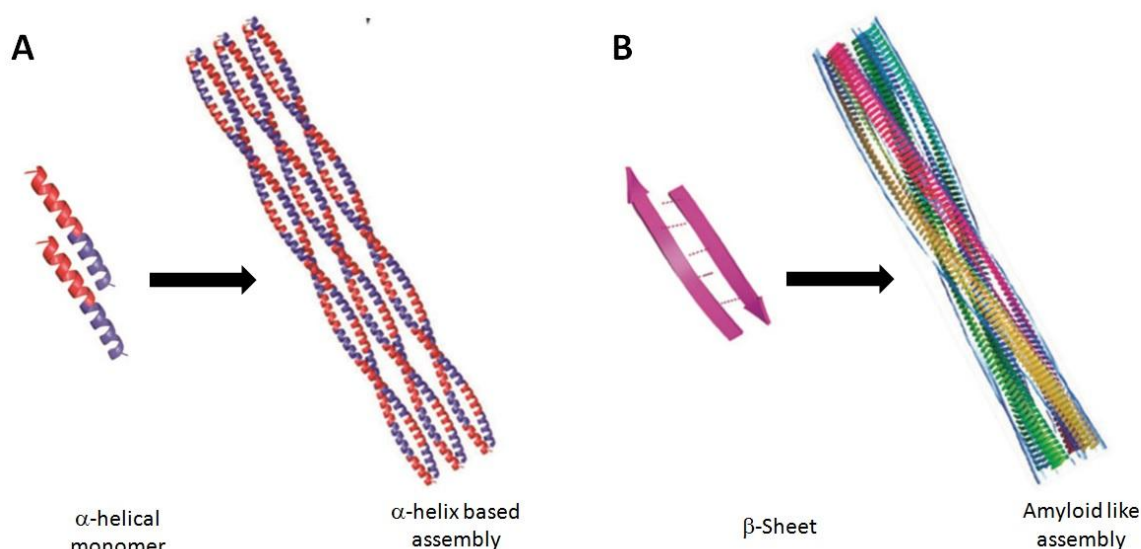


Figure 13. Peptide based filamentous assemblies. A) α -like assembly formation. On the left, small peptides form alpha-helix structure. On the right, filaments formed by self-assembled α -helical coiled-coils. B) Amyloid like assembly formation. On the left, small peptides from a β -strand that hydrogen bond into α -sheets. On the right, self-assembly of the β -sheet to form cross- β -structured fibrils. Figure from [87].

Despite of the advances in the peptide self-assembly, nowadays, there is a growing interest in adding a degree of chemical control and robustness to the obtained structures. This objective can be pursued using proteins, as explained by Fegan et al. [88] who reflect on the role of protein assembly in biological structures to suggest tools to use in “the 1-100 nm niche” 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 of the rational engineering of protein assemblies for nanotechnology [34, 89-91]. The field focuses on the understanding of the design principles inherent in natural proteins and how these might be exploited for the generation of different structures by bottom-up approaches for different application in nanotechnology for biomaterial design, biocatalysis, and synthetic biology. For example, rods and cylinders offer a potential for formation 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, perhaps with controlled release. Planar assemblies suggest applications in protection, molecular filtration, and immobilization of useful functionalities such as enzymes.

One characteristic example on protein design are the tubules created with Hcp1 protein [92]. Hcp1 protein presents a ring shaped structure, with an outer diameter of 9 nm and inner diameter of 4 nm [93]. By introducing cysteine residues on the top and bottom surface of the ring, a protein nanotube was stabilized by disulfide bonds (Figure 14). Biophysical characterization of the Hcp1 architectures showed that the nanotubes can disassemble without perturbing the structure using reduction conditions, thereby

providing a control over its assembly process. This approach can be further exploited for fabrication of nanocontainers.

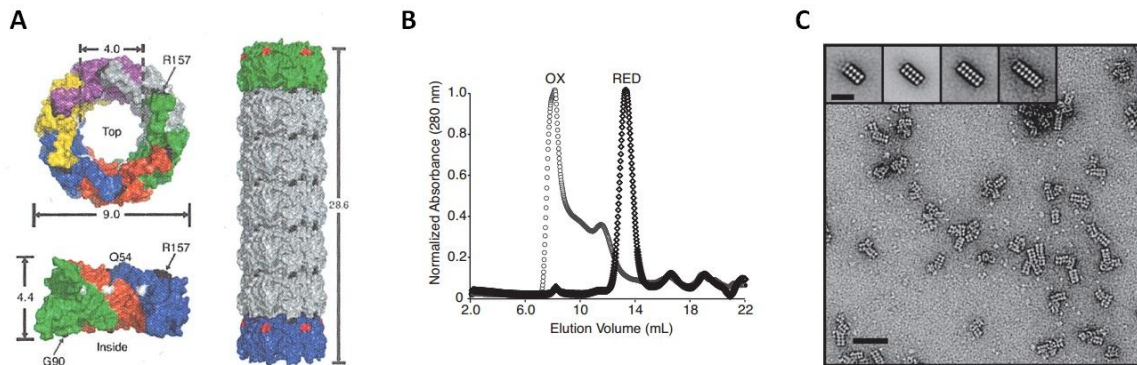


Figure 14. Design and characterization of Hcp1 nanotubes. A) Hcp1 protein genetically modified to assemble into tubules. B) SEC of purified Hcp1CC under oxidizing (circles) and reducing (diamonds) conditions. The oxidized sample peak eluted at the column void volume (7 ml), whereas the reduced sample peak coeluted with wild-type Hcp1 at 14 ml (wild-type data not shown). C) Negative stained TEM and single-particle analysis of Hcp1 from the void volume fraction of the SEC in B). (Scale bar, 45 nm.) (Inset) Averaged structures for Hcp1 nanotubes containing 4–7 ring subunits. (Scale bar, 20 nm). Figure from [94].

During the last years, increasing efforts are being invested in designing *de novo* protein-protein association to create new nanoarchitectures from proteins. The analysis of natural interfaces between proteins has established the formulation of some generic rules that govern this association. As example, Grueninger et al. in 2008 [95] have produced a number of novel assemblies, demonstrating that a given protein can be engineered to form contacts at various points on its surface (Figure 15.B.C.D.E). They performed comparisons between the designed contacts and the resulting assemblies, analyzed by size exclusion chromatography (SEC), dynamic light scattering (DLS) and crystallization. They observed that, depending on the design of the contact, the designed structures or a different ones were obtained in each case. From these results, they concluded that symmetry is an important factor in protein association because it enhances the multiplicity of the designed contact and therefore minimizes the number of required mutations (Figure 15.A). Moreover, they observed that the mobility of the side-chains responsables of the interaction is an important factor in contact design. They demonstrate that the production of a particular contact is feasible, whereas high precision seems difficult to achieve. This work provides useful guidelines for the development of future architectures.

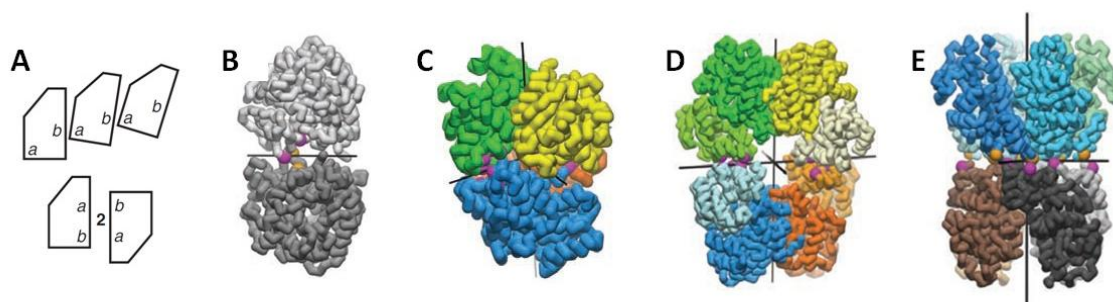


Figure 15. Design of protein assemblies. The proteins in B) to E) are depicted as thick-lined with mutated residues as colored spheres. A) Sketch of an asymmetric interface between patches a and b, which, in general, gives rise to an infinite helix (top). A C₂-symmetric interface also between patches a and b doubles the numbers of contacts and forms a globular complex (bottom). B) Designed assembly in crystal contact a-a. C) Designed assembly as D₂ tetramer at a rotation angle of 86° around a common molecular twofold axis (vertical). D) Designed D₂ tetramer around a common twofold axis (vertical). E) Designed assembly with head-head contact. Figure from [96].

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 tissue engineering, drug-delivery systems, sensors, storage devices, and catalysts [97-100]. However, the application of these materials requires good control over self-assembly and material properties. In spite our increased understanding of interactions that govern assemblies in Nature, in the laboratory there is still a challenge to make controlled assemblies that span scales over several orders of magnitude. Designing biomolecules that assemble into specific, well ordered arrays has been challenging [101] and proof of concept works have just began to arise from protein design perspective [56].

1.2. Chapter 1: Background

1.2. Chapter 1. Background

Structural characterization of the CTPRs revealed some interesting inherent self-assembly properties of these building blocks. In the crystal form, CTPR proteins showed “head-to-tail” and “side-to-side” intermolecular interactions that resulted in different packing geometries of the same molecules [76, 102]. In the crystal form it is observed that individual CTPR molecules stack to form virtually continuous superhelices along the crystal, in which the inter-molecular packing interface is identical to the intramolecular repeat-repeat interface (Figure 16) [76]. These properties can serve as models to develop synthetic systems in which the assembly of the units is controlled by design.

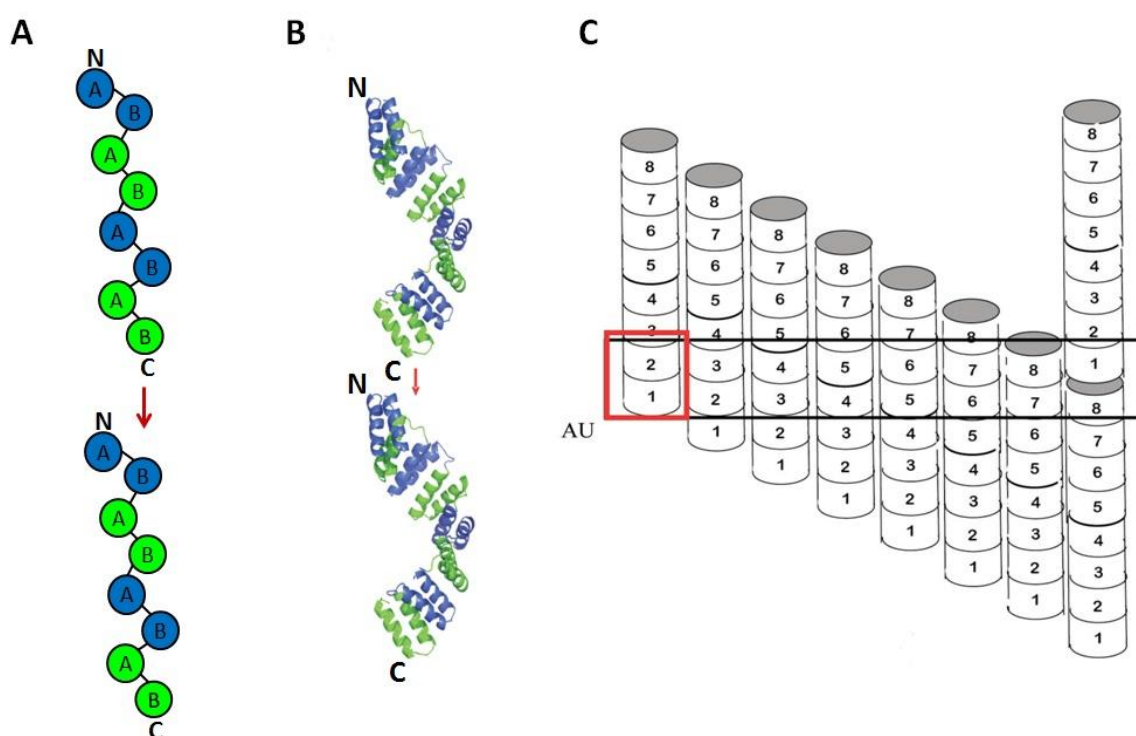


Figure 16. Schematic representation of the crystal-packing interactions between superhelical molecules. A) Schematic illustration of the stacking of helices between repeats of individual molecules. Each repeat is colored in green and blue alternatively. The first repeat (in blue) of the next molecule must always pack against the last repeat of the previous molecule (in green). B) Ribbon representation of the superhelical stacking with each single repeat colored green or blue. C) As an example, in the $P4_12_12$ crystal form there are two repeats (numbered 1–8) within the asymmetric unit (indicated as AU; red box). For this arrangement there are eight equally possible two-repeat arrangements for the asymmetric unit. Figure from [76].

In order to use these proteins as building blocks for supramolecular assembly, is important to control the interactions that will drive the self-assembly and the environment that will affect the assembly process. In this sense, the modular structure of the CTPR repeat proteins and the basic understanding of their sequence-structure

relationships open the possibility to use these proteins for obtaining this control. Thus, it would be possible the formation of different protein assemblies in a controlled manner.

Additionally, it has been described that CTPR proteins retain their original structure and functionality in the solid state [103, 104]. There was demonstrated that, under certain conditions, the CTPR proteins form solid macroscopic films after solvent evaporation of a CTPR aqueous solution. The proteins self-assembled through “head-to-tail” and “side-to-side” interactions similar to the ones that lead the crystallization process [75]. The macroscopic organization within the film was confirmed by different techniques: fluorescence anisotropy, X-ray diffraction (XRD) and atomic force microscopy (AFM). Additionally, circular dichroism (CD) spectroscopy verified that CTPR proteins retained their characteristic α -helical secondary structure, as CD spectrum of the film is practically indistinguishable from the spectrum of the protein in solution (Figure 17). In the work is shown that, as CTPR protein molecules retain their secondary structure in the film, they also retain their function, in particular ligand recognition. These results provide clear evidence that CTPR proteins are ideal models to design novel biomaterials and devices in which molecular order can be modulated by design of the CTPR proteins sequence.

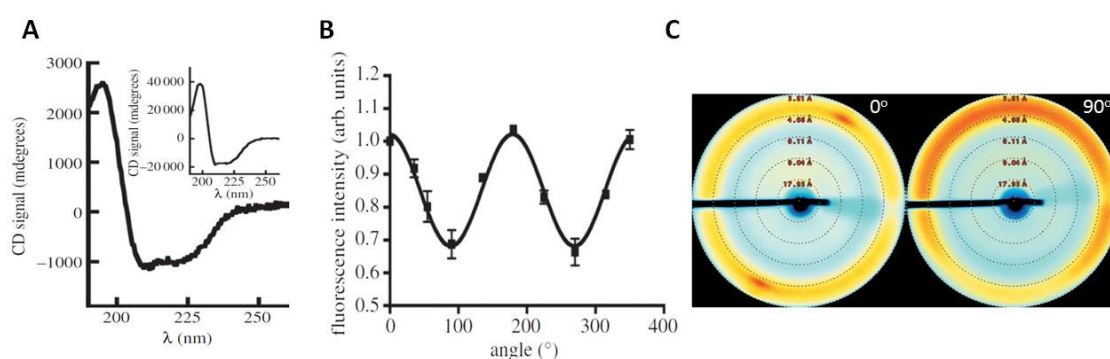


Figure 17. Structural and functional characterization of protein films. A) CD spectrum of a CTPR18 protein film. Inset: CD spectrum of the same protein in solution. B) Change in fluorescence intensity of a peptide ligand bound to the film, demonstrating its anisotropy. C) X-ray diffraction patterns solid protein film at 0° and 90° . At 0° the diffraction arcs corresponding to an order film appear while at 90° the arcs disappear, showing the directional organization of the film. Figure from [105].

1.3. Chapter 1: Objectives

1.3. Chapter 1. Objectives

The main objective of the work described in this chapter is to manipulate the self-assembling properties of CTPR proteins to create higher order structures with defined architectures using bottom-up approach. Self-assembled peptide based nanostructures have demonstrated considerable potential as biomaterials [106-108]. Thus, the self-assembly of the peptides have been exploited for generating bio-inspired nanostructures, including nanofibers, hydrogels and nanotubes [35, 109-111]. What we propose in this chapter is to generate structures similar to those that have been exploited for applications with peptides but using proteins. In contrast to peptides, proteins offer more versatility and robustness because of their stable three dimensional structures and broad range of functionalities.

In this chapter, we present the design of nanostructured assemblies through hierarchical self-assembly of CTPR proteins. We aim to obtain a collection of nanostructures with different dimensionality, in particular we show the generation of protein fibers, 2D protein monolayers and 3D protein nanotubes (Figure 18). For the generation of these structures, we engineered novel proteins predicted to arrange into defined geometries and we will characterize the self-assembly properties of the different protein variants, including the morphology and thermodynamics of the novel complex assemblies.

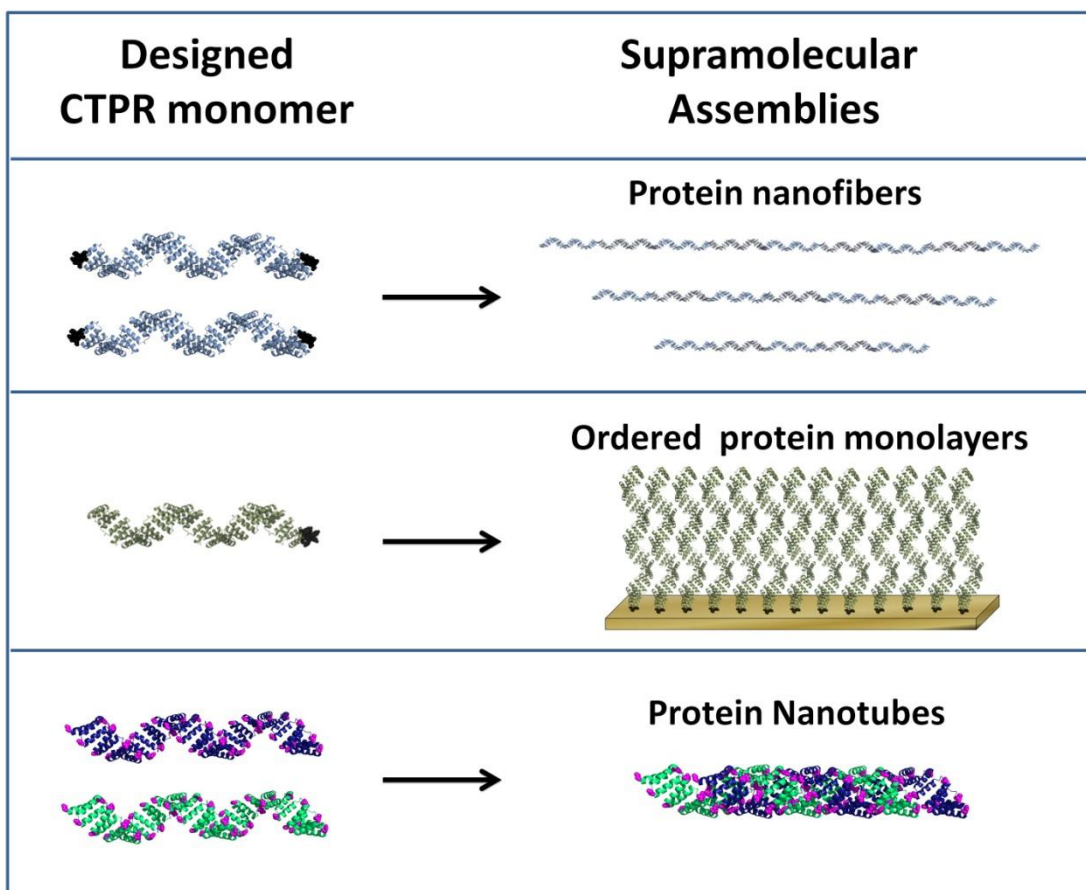


Figure 18. Different types of protein assemblies formed by the introduction of novel interactions between CTPR modified proteins. In the left column the modified CTPR protein structure highlighting the modifications introduced in the CTPR protein. In the right column, the assembly formed from the novel interactions. From top to the bottom: Protein fibers formation led by head-to-tail interactions; CTPR protein monolayer formation led by side-to-side interaction; CTPR protein nanotubes formed by the introduction of a novel interface between the proteins.

1.4. Chapter 1: Results and discussion

1.4. Chapter 1. Results and discussion

1.4.1. Controlled nanometric fibers of self-assembled CTPR

1.4.1.1. Motivation and objectives

In this first approach, we use CTPR proteins to generate fibrillar nanostructures by controlling the self-assembly properties of the CTPR unit. This is the simplest higher order structure based on CTPR blocks. We characterize the kinetics and thermodynamics of the assembly and describe the polymerization process by a simple model that captured the features of the structured formed under defined conditions.

Nanofibers are defined as fibers with a diameter of less than 100 nanometer. There are different examples where peptide based fibers are used in applications such as tissue engineering [112, 113], drug delivery [108], nanosensor and electronics [107, 114], nanoreactors [115] and imaging [111, 116, 117]. However, most of those fibrillar structures are generated using small peptides, which do not have the versatility of functionality exhibited by proteins, as we explained in section *Introduction: 1.4*. In this section, we address the use of proteins to form controlled reversible nanofibrillar structures, with the preservation of the structure, stability and function of the proteins, making the protein based fibers more robust than peptide based fibers.

1.4.1.2. Results and discussion

As it is explained in section *1.2.Chapter 1: background*, in the crystal structures, individual CTPR molecules stack “head-to-tail” to form virtually continuous superhelices along the crystal, in which the inter-molecular packing interface is identical to the intramolecular repeat-repeat interface (Figure 16/ Figure 19.A) [76]. Taking advantage of these interactions, it is hypothesized that linear higher order structures could be assembled by specific contacts between superhelices, similarly to the ones present in crystalline forms, as we show in the schematics in Figure 19.B. This interface is inherent to the repeat units so the unit sequence does not need to be modified ensuring the preservation of the structure, stability and function within the arrays.

In the present study CTPR20 protein is used, the longest array constructed with 20 TPR consensus units, as our molecular unit (Figure 19.B) [76]. Combining the “head-to-tail” interactions present in the crystals with specific design of the functional properties of the CTPR building blocks that form the repeat proteins CTPR linear nanofibers are formed (Figure 19.C).

To generate the nanofibers, cysteine residues are introduced in each terminal end of the CTPR20 protein that act as staple of the “head-to-tail” interaction in solution (Figure 19.C) (*Materials and Methods M.1.1.1*) forming the Cys-CTPR2-Cys protein. If the “head-to-tail” interaction is not fixed using the cysteine, it is not possible to see any polymer formation, even if the head-to-tail interaction happens between the CTPR proteins, because the dissociation of the complex occurs faster than the association when the proteins are in solution. Cysteine-modified interfaces have been already applied to generate nano-scale assemblies [118, 119]. Our designed variant is made to encode directional “head-to-tail” packing according to the schematic picture shown in Figure 19.C without perturbing the structure, stability and functional capabilities of the CTPR domains.

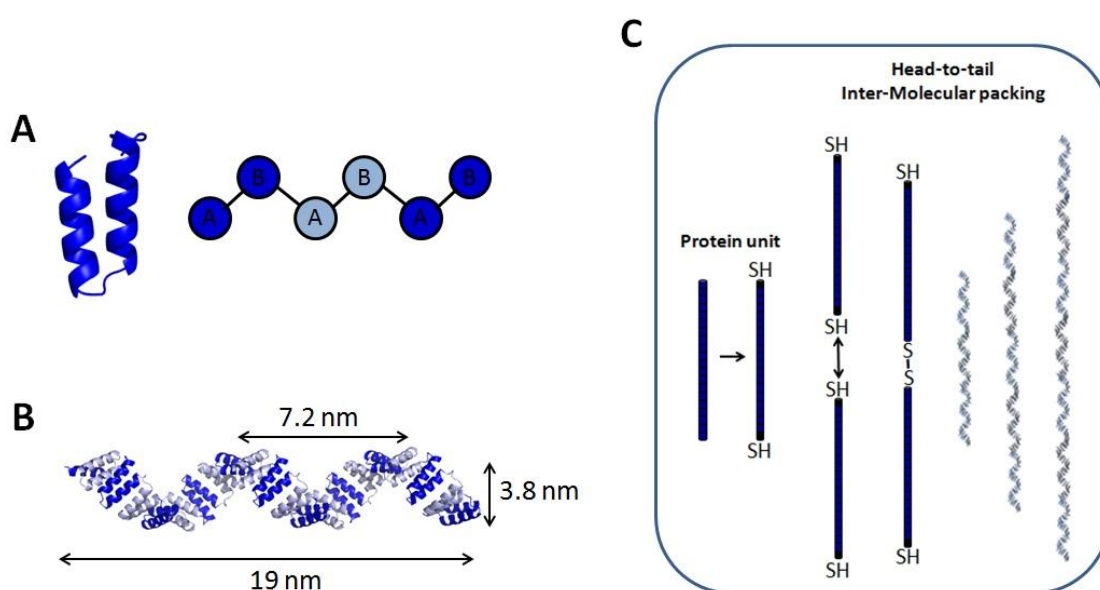
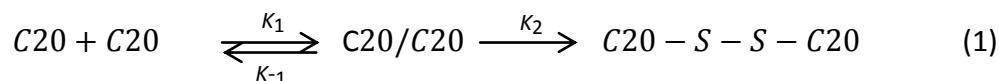


Figure 19. CTPR proteins as building blocks for linear nanofiber formation. A) Structure of a single helix-turn-helix TPR domain (34 amino acids). Schematic representation of TPR packing interactions in which the first repeat composed of two helices is highlighted in dark blue, and alternating repeats are colored in dark and light blue. B) Crystal structure of CTPR20 in which each repeat is colored either orange or blue, alternatively (PDB ID: 2AVP)[76]. The superhelical repeated structure of the 20 repeats is clearly shown in this lateral view of the protein. The protein dimensions and the dimension of one superhelical turn are indicated. C) Schematics showing the bottom-up strategy to generate protein-based polymeric nanofibers. Cys-CTPR20-Cys protein units are formed by 20 identical CTPRs and the units are modified to present two unique cysteine residues at the N- and C-terminal ends (left). As a result of the “head-to-tail” inter-molecular interactions the protein units are hypothesized to polymerize with cysteine mediated di-sulfide bonds acting as staples of the interaction (right). The intermolecular packing interfaces in the polymers are the same as the intramolecular interfaces between two repeats in the same molecule, as shown in panel A.

The formation of Cys-CTPR20-Cys polymers is facilitated by the kinetics of association between the N-terminal and C-terminal subunits (i.e., the “head-to-tail” interaction). The formation of Cys-CTPR20-Cys polymers is led by the kinetics of “head-to-tail” interaction, including the recognition of two molecules through their packing interfaces

and the formation of a disulfide bond between the cysteines. The polymerization includes the recognition of two molecules through their packing interfaces and the formation of a disulfide bond that acts as a staple of the interaction to generate longer polymers. Therefore, the polymerization is expected to follow the mechanism described in Equation (1):



Where k_1 and k_{-1} are the rate constants of the “head-to-tail” interaction and dissociation respectively, and k_2 is the rate constant of the formation of the disulfide bond. The polymerization in solution of Cys-CTPR20-Cys proteins is monitored by the increase in the average hydrodynamic size by dynamic light scattering (DLS) (*Materials and Methods M.1.1.2 and M.1.1.3*). When a freshly reduced protein sample is placed in the DLS instrument at defined protein concentration and temperature the polymerization kinetics of the Cys-CTPR20-Cys proteins can be clearly observed by the increase in the hydrodynamic diameter with time (Figure 20.A).

For CTPR20 proteins with no cysteines even though the “head-to-tail” interfacial recognition should take place, no significant polymer growth can be detected by DLS (Figure 20.A, circles) or fiber formation in the TEM image (Figure 21.A). The hydrodynamic diameter in Figure 20.A increases linearly with time as expected from the reaction scheme given by Equation (1), and can be fitted to estimate the polymerization rate to Equation (2):

$$D = k \times t + D_0 \quad (2)$$

Where D is the average hydrodynamic diameter of the polymeric sample in nm; k is the rate of polymerization; t is the time in minutes and D_0 is the hydrodynamic diameter at time 0. The obtained results show that the CTPR scaffold proteins in solution at relatively diluted protein concentrations (low μM) are able to polymerize and form larger structures. The Cys-CTPR20-Cys polymerization can be reversed by adding a reducing agent to the sample after polymerization saturation, and the hydrodynamic size of the sample returns to the initial value (13.94 ± 0.63 nm).

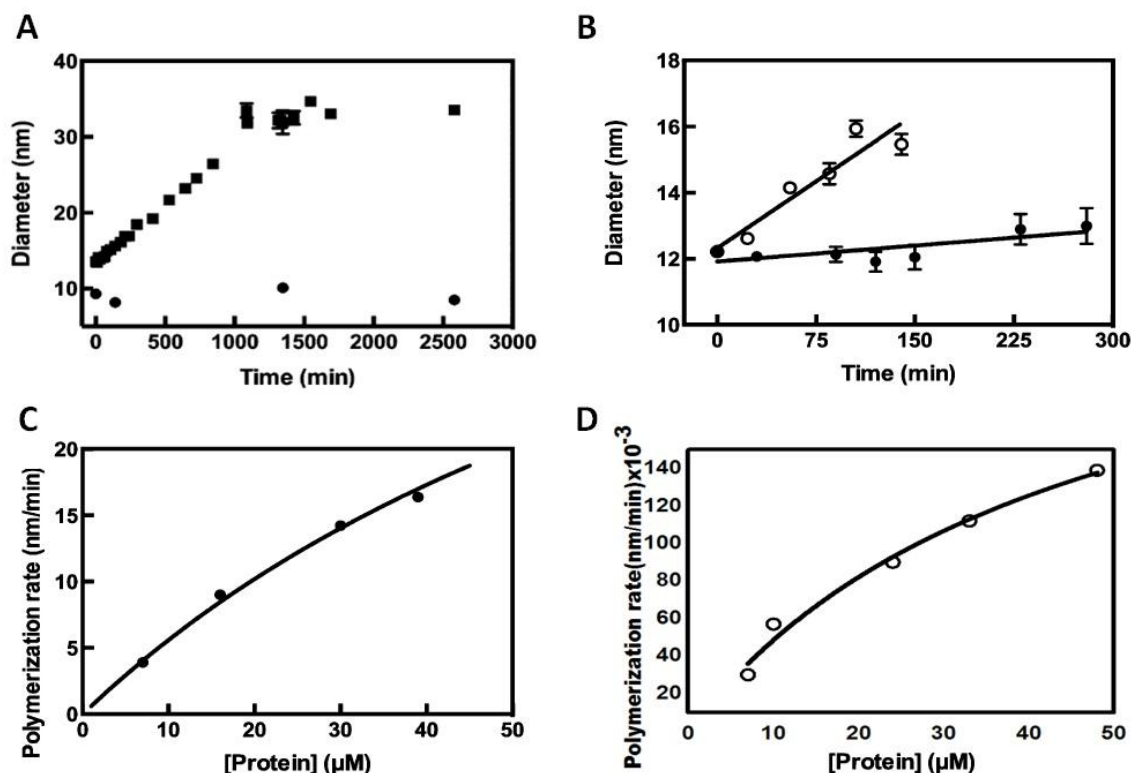


Figure 20. CTPR20 protein polymerization kinetics. A) Polymerization growth monitored by the increase in the size of the Cys-CTPR20-Cys polymers as a function of time at 40 μM protein concentration (solid squares). The control showing the size of CTPR20 without cysteines measured over time is shown in solid circles. The error bars show the standard deviation of three measurements, in some data points the error bars are not visible because they are smaller than the size of the symbols. B) Effect of the temperature on the polymerization kinetics. The increase in the size of Cys-CTPR20-Cys at 7 μM protein concentration is measured by DLS and plotted versus the time at two polymerization temperatures, 25°C (filled circles) and 42°C (empty circles). At 42°C, the polymerization is about 10 times faster than at 25°C. Error bars show the standard deviation of three measurements. C) Effect of the protein concentration on the polymerization kinetics. Polymerization growth rates at different protein concentrations (7 μM, 16 μM, 30 μM and 40 μM) (filled circles) can be well fitted to the saturation Equation (3) (solid line). D) Effect of the protein concentration on the polymerization kinetics at 42°C. Polymerization growth rates at different protein concentrations (7 μM, 10 μM, 24 μM, 33 μM and 48 μM) (empty circles) can be well fitted to the saturation Equation (3) (solid line).

To quantitatively characterize in detail the polymerization process, the effect of the protein concentration in the polymerization rate at 25°C is monitored. First, it is observed that the growth rates are dependent on the initial protein concentration, as expected (Figure 20.C). Since there is no measurable increase in the hydrodynamic size without cysteines (Figure 20.A, circles), it is assumed that the equilibrium in Equation (1) is shifted towards the monomeric forms, with the dissociation of the complex being faster than the association under our experimental conditions ($k_1 \ll k_{-1}$). The disulfide bond equilibrium under the experimental conditions can be assumed fully shifted to the disulfide bond formation, considering the described bond dissociation energy [120]. Additionally, the formation of cysteine bonds in vitro has been described to be relatively slow and the limiting step of many folding processes [121]. Therefore, it can be assumed that dissociation of the CTPR20 interface is much faster than the staple through the disulfide bond ($k_2 \ll \ll$

k_{-1}). Under these conditions, it can be assumed that the CTPR20 interface (complex C20/C20 in Equation (1)) is in the steady-state, and the velocity of growth is fitted as a function of the protein concentration to a Michaelis–Menten-like function (Equation (3)) to obtain the effective maximum velocity, V_{\max} and the dissociation constant $K_D = k_{-1}/k_1$ (since the Michaelis constant $K = (k_{-1} + k_2)/k_1 \sim K_D$ in the steady-state):

$$V = \frac{V_{\max} [P]}{k + [P]} \quad (3)$$

The values obtained from the fitting are shown in Table 1.

Temperature (°C)	V_{\max} (nm min ⁻¹)	K_D (μM)	ΔG_{exp} (kcal mol ⁻¹)
25	0.056	91.81	- 5.50
41	0.435	90.49	- 5.51

Table 1. Kinetic and thermodynamic parameters of CTPR building block polymerization. The maximum velocity of the polymerization process (V_{\max}), the dissociation constant for the inter-molecular interaction (K_D), and the difference in Gibbs energy (ΔG_{exp}) obtained from the Cys-CTPR20-Cys polymerization experiments at two different temperatures: 25°C and 42°C.

Next, the effect of temperature on the polymerization kinetics is analyzed, repeating identical polymerization experiments at 42°C. Figure 20.B compares the growth rates at 25°C and 42°C for an initial protein concentration of 7 μM. At 42°C, the functional dependence of the growth velocity is also of Michaelis–Menten form (Figure 20.D), with the fitted maximum velocity about ten times faster than at 25°C (Table 1).

The increase in the polymerization rate with the temperature is described by Arrhenius equation (Equation (4)):

$$\frac{K_2}{K_1} = e^{\left[\frac{E_a}{R} \left(\frac{1}{T_2} - \frac{1}{T_1}\right)\right]} \quad (4)$$

Where K_1 and K_2 are two reaction rates in m/s; T_1 and T_2 are two reaction temperatures in Kelvin; E_a is activation energy in Kcal; R is the ideal gas constant ($1.98 \cdot 10^{-3}$ kcal K⁻¹ mol⁻¹). From Equation (4) the activation energy (E_a) of the polymerization is estimated to be around 25 kcal/mol, in the same range of spontaneous natural fiber growth such as collagen formation [122]. From the equilibrium constant (K_D) we can calculate the change in free energy of the “head-to-tail” interaction, ΔG_{exp} (Table 1) is calculated, using Equation (5):

$$\Delta G = -RT \ln (K_D) \quad (5)$$

Where R is the ideal gas constant ($1.98 \cdot 10^{-3} \text{ kcal K}^{-1} \text{ mol}^{-1}$), T is the temperature in Kelvin and K_D the dissociation constant. These equilibrium parameters are important for the use of these scaffolds under experimental conditions that ensure the thermodynamic stability of the nanostructures.

The values obtained for the inter-molecular “head-to-tail” interaction can be validated with those expected from previous detailed structural and stability data on CTPR proteins [78, 123, 124]. As mentioned before, the new inter-molecular interface is identical to the intra-molecular repeat-repeat interface (Figure 19) and therefore, the expected ΔG for “head-to-tail” packing can be calculated in different ways. First, ΔG is estimated based on a calculation of the free energy associated with the hydrophobic surface area burial upon binding, using the transfer free energy scales from hydrophobic to polar media³⁶ for each amino acid on the interface. Taking into account the change in the free energy of the amino acids in the interface between two monomers, this ΔG is -5.59 kcal/mol . Additionally, the inter-repeat coupling interaction is estimated from the previously published linear 1D-Ising model description of the folding and stability of CTPR proteins [77, 78]. This model dissects the experimental thermodynamic stability of the repeated proteins in two parameters, the intrinsic repeat stability (H) and the coupling energy between two adjacent repeats (J) [77, 78, 125]. The repeat-repeat interfacial free energy ΔG is -5.45 kcal/mol considering a J coupling value of 2.3 [78]. These results show that the ΔG calculated from the polymerization experiments is in close agreement with the interfacial interaction energy calculated by two independent methods, and confirm that the driving force of the polymerization is the “head-to-tail” interaction between the C- and N-terminal repeats of two proteins.

In order to obtain structural information about the morphology of the higher order species observed by DLS the polymerized protein samples are imaged using transmission electron microscopy (TEM) (*Materials and Methods M.1.1.4*). First, as a control, CTPR20 protein monomers without cysteines are imaged. Figure 21.A shows the presence of individual rod-shaped CTPR20 monomers, of $18.8 \pm 1.3 \text{ nm}$ long and $3.6 \pm 0.4 \text{ nm}$ width, in close agreement with the dimensions of the CTPR20 protein calculated from the crystal structure [76]. The superhelical structure of the CTPR20 monomers is also observed. Then, protein samples after saturation of the polymerization growth are imaged. Figure 21.B clearly shows the presence of linear fibers. Their experimentally measured width is $3.6 \pm 0.6 \text{ nm}$, in agreement with fibers formed by the linear assembly of CTPR20 monomers. These fibers are significantly thinner than other protein-based amyloid and collagen fibers previously studied and comparable to hyperthin nanochains of 2–3 nm recently reported [126]. To determine the specificity of the polymerization process through the disulfide bond staple of the “head-to-tail” interaction, the polymerized

sample is reduced. The disruption of the polymeric fibers to monomers by the reducing agent is observed in Figure 21.C, indicating the reversibility of the assembly and that the polymers observed in Figure 21.B are the result of the designed controlled linear interaction.

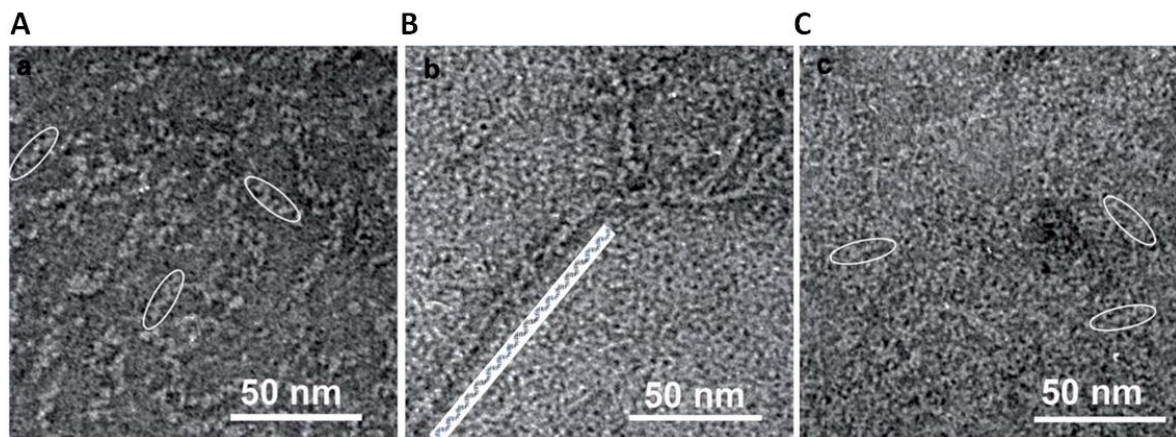


Figure 21. Negative stain TEM imaging of the protein polymerization process. A) Transmission electron micrograph of the CTPR20 monomer deposited on TEM grids. Individual CTPR20 monomers are circled. The particle size measured is the average of 5 molecules. B) Cys-CTPR20-Cys sample after polymerization saturation, where the linear polymeric structures are observed. The schematic arrangement of CTPR20 units in the linear fibers is shown with each CTPR20 unit displayed in different colors on approximately the scale of the fiber in the TEM image. C) Cys-CTPR20-Cys after polymerization treated with a reducing agent (1 mM DTT), where the polymeric structures dissociate, and monomers appear. Individual CTPR20 monomers are circled.

During the polymerization experiments the average hydrodynamic size of the sample is measured over the time by DLS. To have a quantitative description of the polymerization process it is necessary to correlate the average size with the number of building block units. The relationship between the number of repeats and the hydrodynamic size for CTPR proteins with different numbers of repeats is previously estimated (Figure 22.A) (*Materials and Methods M.1.1.5*) [127]. The hydrodynamic size of CTPR proteins with different repeats is experimentally measured by fluorescence correlation spectroscopy, and protein sizes are also estimated from the crystal structures using the program Hydropro [128]. This relationship, shown in Figure 22.A, can be fitted to the scaling equation $R_h = 1.48 N^{0.376}$ that is used to calculate the average number of repeats per polymer chain (N) from the values of average hydrodynamic radius (R_h). Additionally, these calculations allow to estimate the average polymer size that would be expected from a defined polymerization process.

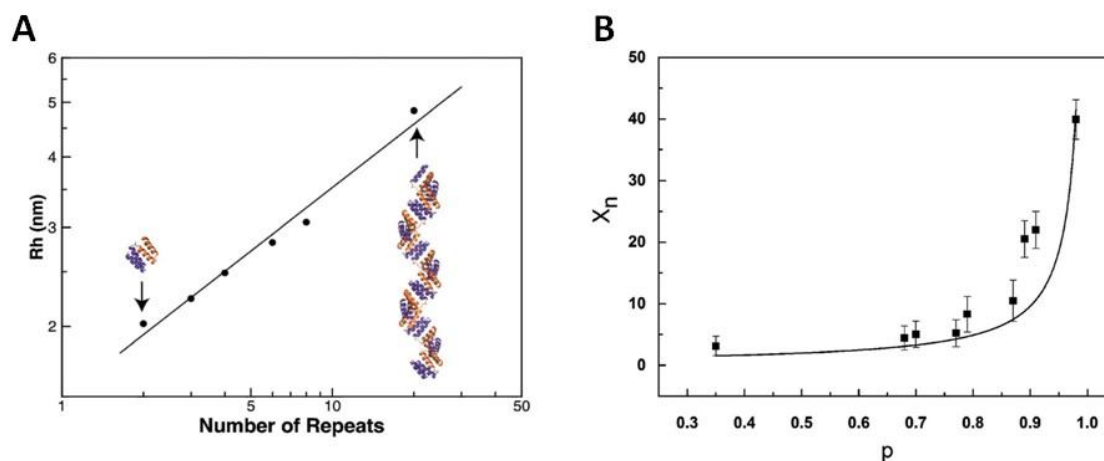


Figure 22. A) **Hydrodynamic radius (R_h) of CTPR proteins as a function of the number of repeated units.** The log–log plot is fitted to the relationship $R_h = 1.486N^{0.376}$. The structures of CTPR proteins with 2 and 20 repeats are showed with each repeat unit colored blue or orange, alternatively. B) **Step growth polymerization model.** The relationship between the degree of polymerization quantified as the average number of monomers in the polymer chains (X_n) and the reaction probability (or extent of reaction) quantified as the fraction of Cys ends bonded (p). Black squares represent the experimental data from DLS measurements. The solid line shows the step-growth model prediction ($X_n = 1/(1 - p)$). Error bars indicate the standard error of three measurements.

To quantify the self-assembly process and predict both the kinetics of chain growth and the chain length under different experimental conditions, a simple model based on 1D polymer-like structure growth is tested [129]. Each protein fiber can grow through two N- and C-terminal ends and the probability of adding an extra molecule at each end would be the same because of the identical interaction interface. Therefore, it is hypothesized that the process may occur following a simple statistical step-growth polymerization process. The step-growth model is a random process that proceeds via a step-by-step succession of elementary reactions between reactive sites and assumes equal reactivity of the functional group [129, 130]. This model has been successfully used recently to describe the polymerization of inorganic nanoparticles [131].

CTPR modules can be considered bifunctional monomers with two identical functionalities one at each end. At different times of the polymerization reaction several parameters that define the polymer growth are calculated: (1) the average degree of polymerization (X_n) defined as the average number of CTPR20 monomers per polymer chain, calculated from the average hydrodynamic size using the relationship described in Figure 22.A; (2) the fraction of functional groups bounded, or extent of reaction (p), estimated measuring the free cysteine groups as a function of polymerization time. For an ideal step growth polymerization these parameters are related by Equation (6) [129]:

$$X_n = \frac{1}{1-p} \quad (6)$$

The relationship between X_n and p are shown in Figure 22.B: as p increases, X_n becomes larger until the saturation of the polymerization process at $p=1$. The experimental values at different polymerization stages closely follow the behavior of the ideal model.

1.4.1.3. Conclusions

In conclusion, we have shown that we can generate thin long linear protein nanofibers with designed CTPR protein building blocks in solution at relatively low protein concentrations. In this work we took advantage of the intrinsic “head-to-tail” interactions encoded in the repeated sequence, and, therefore, no modification of the scaffold is required thus the structure and stability of the building blocks are preserved. Additionally, a simple disulfide bond staple is used to generate linear polymers. This approach is simpler than others described for polymerization of repeat protein scaffolds in which chemical ligation is used [132]. Our structures can reversibly disassemble into the monomeric units under reducing conditions while maintaining the structure of the building blocks.

We have described in detail the kinetics and thermodynamics of the interaction. This quantitative description provides the tools to rationally control the polymerization process by tuning experimental conditions such as protein concentration, temperature and polymerization time to achieve the desired size distributions of the polymeric chains and to ensure the thermodynamical stability of the higher order structures.

In this section, we set up the basis for fabrication and templating based on these protein scaffolds. We demonstrate that nanometric thin fibers can be built from simple building blocks with minimal modification of the units. The CTPR system provides several advantages compared to other protein based fibers reported [133-135]: the design is simple without mutational or chemical modification of the blocks; the fiber formation is covalent but reversible, what it has specially interest for applications [133]; the thin diameter and the superhelical conformation of the arrays grant a large exposed surface area and therefore potential functional binding sites; the structural periodicity of the CTPR superhelix can be used to define periodicity on the functional decoration; Finally, a predictive simple polymerization model can be used to describe the self-assembly of the building blocks.

1.4.2. Tightly packed CTPR protein monolayer

1.4.2.1. Motivation and objectives

In this section, we describe the assembly of the CTPR proteins into 2D monolayers by oriented immobilization. CTPR proteins interact by side-to-side interaction to form a high packed surface. To characterize the protein's specific adsorption on gold surface and their self-packing, we monitored the kinetics of the immobilization process and characterize the physical properties of the monolayers including topography, thickness and viscoelasticity.

Nowadays, there is a growing interest in the study of protein absorption on surface because of its wide application in nanotechnology, biological sensing, materials science and biology [136]. For example, in biomaterials field, protein adsorption is the first step in the integration of an implanted device or material with tissue [137, 138]. Also, in nanotechnology, the control of the protein–surface interactions that lead the assembly of the proteins is crucial to generate sensors, activators and other functional components at the biological/electronic junction [139].

However, to progress in the applications mentioned above is crucial to understand the molecular-level interactions of proteins and surfaces and how to control and tune the absorption in the desired degree. To achieve the goal of rationally control the absorption of proteins onto surfaces, it is necessary a deep understanding of the mechanism that govern the protein absorption. Also, to exploit the biological function of an adsorbed protein molecule, one requires some control of the orientation of the protein at the interface. However, the lack of experimental data leads to difficulties in investigating molecular level hypotheses and validating molecular model regarding protein absorption [140].

1.4.2.2. Results and discussion

In this chapter, it is proposed to use CTPR proteins to form self-assembled monolayers (SAMS) using the selectivity of thiol–gold interaction for an oriented immobilization on gold surface[141]. As it is explained before, under certain conditions, CTPR proteins self-assembled into ordered protein films (Figure 17) [103] mimicking the packing observed in the crystal forms of CTPRs (Figure 16) [77, 102, 124]. One key component for the assembly is the side-to-side lateral interactions between CTPR superhelices [103, 104]. Figure 23.A/B schematically shows the potential alignment of CTPR superhelices. In this section, the self-assembly of the CTPR20-Cys protein composed of 20 identical CTPR repeats with a single C-terminal cysteine is analyzed (*Materials and Methods M.1.2.1*).

CTPR20 is chosen since it is the longest CTPR that can be produced recombinantly with high yield. The longer the CTPR protein the higher the aspect ratio[26], thus CTPR20 is expected to have higher propensity to be oriented on the surface and form more stable side-to-side interactions. In addition, longer CTPRs show higher stability[40] which is important for future applications. CTPR20-Cys is immobilized on a gold surface through gold–sulfur bond and, driven by the lateral inter-molecular interactions, can generate ordered and oriented protein self-assembled monolayers (SAMs) (Figure 23.C) (*Materials and Methods M.1.2.2*).

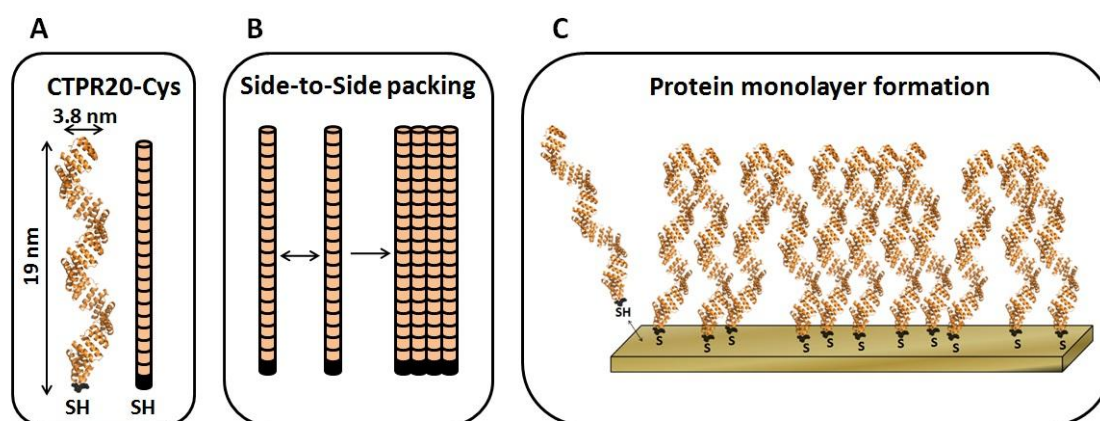


Figure 23. Schematic representation of CTPR building blocks and lateral monolayer assembly. A) CTPR unit showing the X-ray crystal structure of the CTPR20 protein with the characteristic superhelical conformation and a schematic representation of CTPR protein unit in orange. Colored in black is represented the cysteine residue added to the protein immobilization in gold surface. B) Schematic of the CTPR protein packing model where each protein is hypothesized to assemble non-covalently with other CTPR proteins by side-to-side lateral interactions. C) Schematic representation of CTPR20-Cys adsorption onto gold surface through thiol-gold interactions.

The first step of the study is the adsorption of CTPR20-Cys onto gold surface. Thus, first the immobilization on gold surface of two proteins is compared: CTPR20-Cys for oriented thiol mediated immobilization and CTPR20 without Cys as a non-oriented deposition control. The process is monitored by quartz crystal microbalance (QCM), which is a powerful tool to study interfacial process at surfaces (*Materials and Methods M.1.2.3*). After stabilization of the temperature, frequency and resistance, purified proteins solutions prepared as previously reported [71, 127] are flowed through the cell. Figure 24.A.1/2 presents the resulting frequency and resistance changes as a function of time after protein injection. As it is seen in Figure 24.A.1, upon flowing CTPR20-Cys protein, a rapid decrease in the resonant frequency is observed during the first 2 min as a result of the protein binding to the surface by chemisorption of cysteine on gold. Afterward, the frequency decreases more slowly until a steady state is reached. Simultaneously with the frequency decrease, an increase in motional resistance is also observed (Figure 24.A.2). It

indicates that the binding of the protein does not result in a rigid load but viscoelastic causing an energy loss by dissipation from the QCM sensor.

The ratio between the motional resistance change and the frequency shift can be used to evaluate the relative influence of viscoelastic and mass loading effects (see below). The shape of the frequency–time profile can be employed to study the kinetics of adsorption by fitting to a first-order kinetics equation (Equation (7) (Figure 24.A.1):

$$\Delta F = -\Delta F_{\max} (1 - e^{-kt}) \quad (7)$$

Where ΔF is the frequency change (in hertz), ΔF_{\max} the frequency change between the initial and the steady-state frequencies, and k the first-order rate constant (min^{-1}). From the data fitting, the obtained values for ΔF_{\max} and k are -73.2 Hz and 0.57 min^{-1} , respectively. From the ΔF_{\max} at the equilibrium, assuming that the frequency decrease is only due to the change in mass arising from the adsorption of the protein, the mass of CTPR20-Cys immobilized on the gold electrode surface can be calculated using Sauerbrey's equation (Equation (8):

$$\Delta m = - C_f \Delta F \quad (8)$$

Where Δm is the mass change (ng cm^{-2}), C_f ($17.7 \text{ ng Hz}^{-1}\text{cm}^{-2}$) the proportionality constant for the 5.0 MHz crystals used in this study, and ΔF the frequency change (in hertz). Using Sauerbrey's relation, the estimated mass of CTPR20-Cys immobilized on the gold surface is 1285 ng cm^{-2} . This value corresponds to a surface coverage of about $15.2 \times 10^{-12} \text{ mol cm}^{-2}$ considering a molecular mass of 85242 Da for the CTPR20-Cys ($9.0 \times 10^{12} \text{ CTPR20-Cys molecules cm}^{-2}$). This value is in agreement with the number of molecules for a compact monolayer of oriented CTPR20-Cys proteins, considering a projected area of about 11 nm^2 per molecule based on the dimensions from the crystal structure (Figure 23.A) [124]. Additional support to this fact is provided by the comparison with the extensively studied monolayers of *n*-alkanethiolates on to gold surface [142, 143] which result in a maximum coverage of $\sim 4.2 \times 10^{14} \text{ molecules cm}^{-2}$ for molecules that cover 0.217 nm^2 [144]. For an equally packed CTPR20-Cys molecule (11 nm^2) the theoretical maximum coverage would be reached with 50 times less molecules $8.4 \times 10^{12} \text{ molecules cm}^{-2}$, which is in the same range of the experimental value of $9.0 \times 10^{12} \text{ molecules cm}^{-2}$.

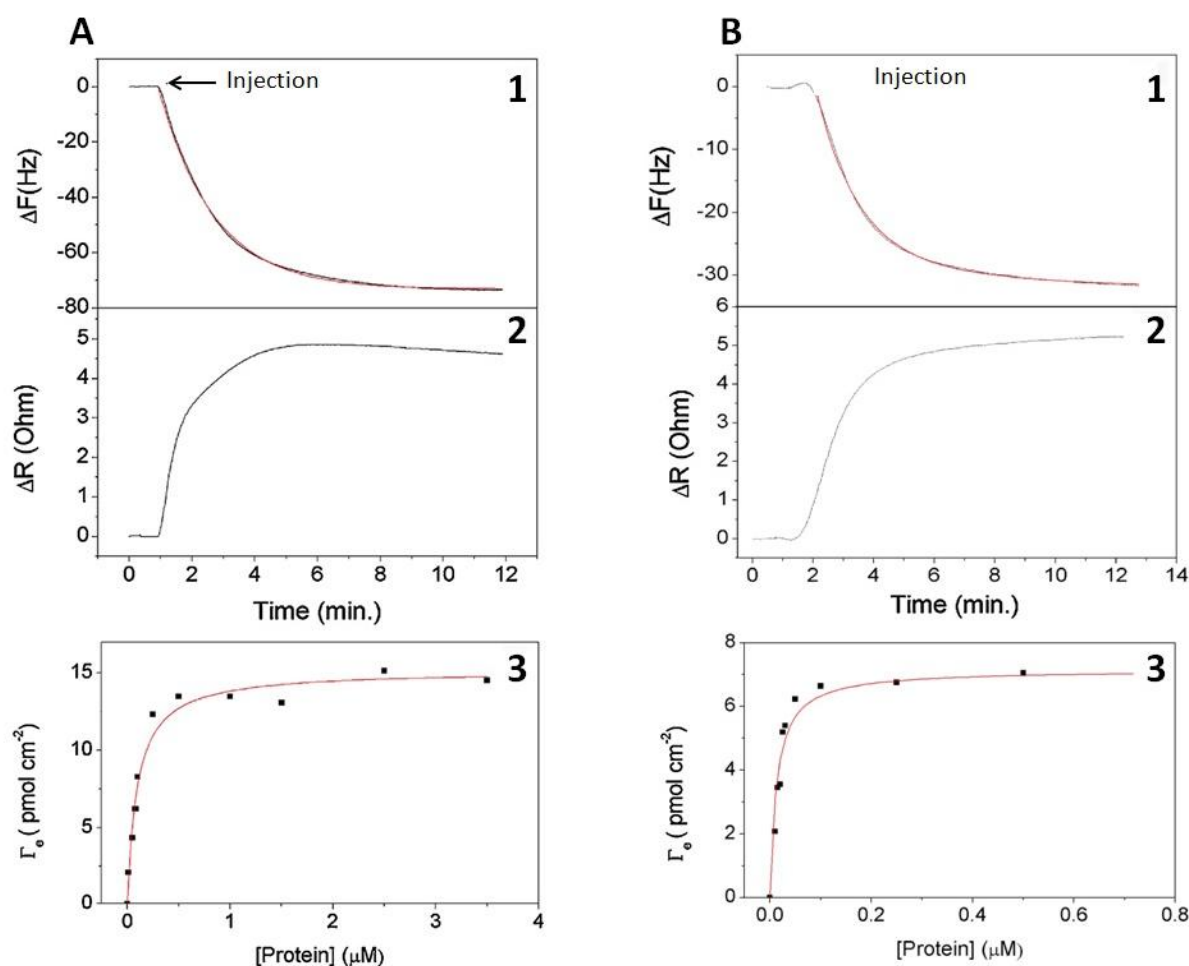


Figure 24. CTPR20 and CTPR20-Cys adsorption on gold surface monitored by quartz crystal microbalance. A) **1.** Adsorption kinetics of CTPR20-Cys onto quartz crystals with gold electrodes at 2.5 μM protein concentration. The frequency changes of the quartz crystal resonator are recorded upon protein injection. The curve is fit to a first-order kinetics equation (solid line), obtaining -73.2Hz and 0.57min^{-1} values for $-\Delta F_{\text{max}}$ and k , respectively. **2.** Simultaneous resistance increase recorded by QCM upon CTPR20-Cys protein injection. **3.** Equilibrium surface coverage by QCM. Equilibrium surface coverage of CTPR20-Cys adsorbed on gold surface represented versus the protein concentration and fitted using a Langmuir adsorption isotherm. The calculated value for Γ_s is $15.2 \times 10^{-12} \text{ mol cm}^{-2}$ for CTPR20-Cys. B) **1.** Adsorption kinetics of CTPR20 without Cys onto quartz crystals with gold electrodes at 0.25 μM protein concentration. The curve is fit to a first-order double exponential decay with two absorption processes (solid line), obtaining for the fast decay rate -32.2 Hz and 0.71 min^{-1} values for ΔF_{max} and k , respectively. **2.** Simultaneous resistance increase recorded by QCM upon CTPR20 protein injection. **3.** Equilibrium surface coverage by QCM of CTPR20 adsorbed on gold surface represented versus the protein concentration and fitted using a Langmuir adsorption isotherm. The calculated value for Γ_s is $7.3 \times 10^{-12} \text{ mol cm}^{-2}$ for CTPR20.

These results show that the protein units assemble in a compact manner on the surface by thiol chemisorption of cysteine residues onto the gold surface. As it is the case for more simple thiol derivatives, CTPR20-Cys proteins are not randomly deposited since the amount of protein per surface area would be, otherwise, smaller. To serve as comparison, QCM adsorption measurements of CTPR20 without cysteine are also carried out. As shown in Figure 24.B.1 flowing CTPR20 gives a final surface coverage of 6.5×10^{-12}

mol cm^{-2} , significantly lower than for CTPR20-Cys. This result can be associated to the fact that direct adsorption of CTPR20 proceeds through unspecific multi point attachment of the protein to the surface contrary to the oriented thiol-derivative protein that is tightly adsorbed on the gold surface by an unique single-point interaction [145, 146]. This fact is also put in evidence by studying the thermodynamics of adsorption. The calculated equilibrium proteins coverages from ΔF_{max} for different protein concentrations are fitted to a Langmuir adsorption isotherm (Equation (9)):

$$\Gamma_e = \frac{\Gamma_s K_a C_s}{1 + K_a C_s} \quad (9)$$

Where K_a is the thermodynamic binding constant, C_s is the bulk protein concentration, and Γ_e and Γ_s are the equilibrium and saturation protein coverage. The saturation protein coverage of CTPR20-Cys protein calculated with Equation (9) is $\Gamma_s = 15.2 \times 10^{-12} \text{ mol cm}^{-2}$ (Figure 24.A.3). In contrast, the saturation protein coverage for CTPR20 without cysteine is Γ_s is $7.3 \times 10^{-12} \text{ mol cm}^{-2}$ (Figure 24.B.3). The coverage at saturation for CTPR20 is half of the one obtained for CTPR20-Cys, presumably due to a random immobilization of CTPR20 in different orientations.

In order to confirm the validity of conversion the change infrequency to molecular mass from QCM data using Sauerbrey's equation, surface plasmon resonance (SPR) experiments are conducted, as detailed in *Materials and Methods M.1.2.4* [147, 148], to quantify the amount of protein bound. Figure 25.A shows a SPR experiment carried out employing a CTPR20-Cys protein concentration of $2.5 \mu\text{M}$ to saturate the surface. The mass measured by SPR from the resonance angle shift upon protein binding to the gold surface is approximately 1000 ng cm^{-2} . This mass is slightly lower than the value obtained from QCM (1285 ng cm^{-2}), probably due to the uncertainty of the refractive index of the adsorbed protein and to the effect of the hydration water in the QCM measurements. The SPR mass can be considered comparable to the QCM mass confirming that the frequency changes observed in the QCM experiments are mainly ascribed to mass changes as consequence of the protein adsorption since viscoelastic effects should induce a bigger difference between Δm_{QCM} and Δm_{SPR} , as previously reported [149].

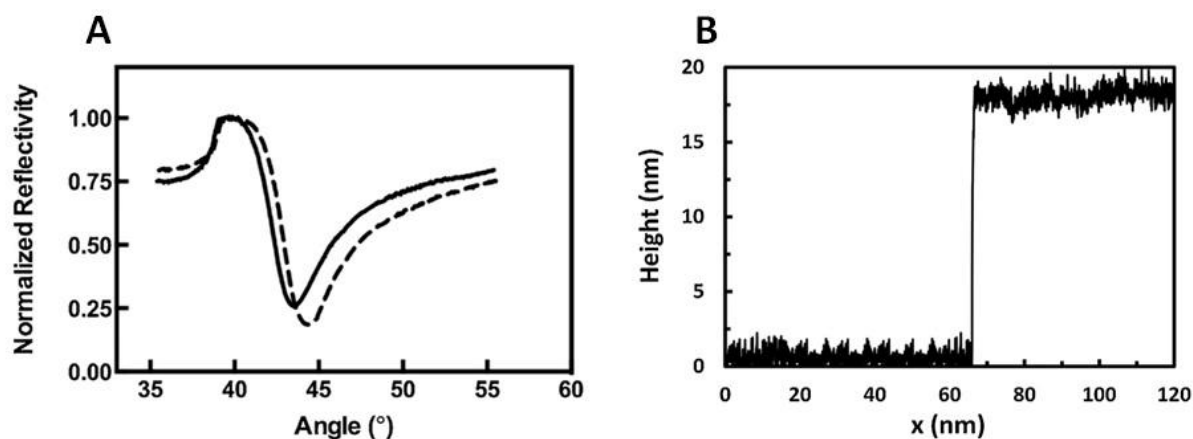


Figure 25. A) Surface protein coverage by surface plasmon resonance. SPR spectra of bare gold surface (solid line) and CTPR20-Cys deposited on the gold surface (dashed line). The protein spectrum fits with a gold film thickness of 32 ± 2 nm, a protein coverage of approximately 1000 ng cm^{-2} . B) Height profile of the CTPR20-Cys monolayer deposited on the gold surface acquired by a stylus profilometer. The thickness of the protein layer was 17.93 ± 1.05 nm.

AFM is used to image the CTPR20-Cys proteins immobilized onto gold surfaces (*Materials and Methods M.1.2.5*). Single CTPR20-Cys molecules are imaged first at a low protein concentration (10 nM) to characterize the features of the individual monomers (Figure 26.A). A single size distribution peak is obtained for individual CTPR20-Cys molecules with an average height of approximately 8 nm, indicating that each bright dot on the image corresponds to a protein monomer (Figure 26.A). The width measurements are not reliable due to the tip convolution [53]. The end to end distance of the CTPR20 superhelix calculated from the crystal structure is approximately 19 nm [76] (Figure 23.A). There are several reasons that will account for the smaller size measured by AFM: (1) CTPR20-Cys adsorption onto gold surface through its thiol group might be tilted with an angle relative to an orthogonal orientation to the surface; (2) vertical elastic tip compression of the elongated superhelical structure, as it has been described for other repeat proteins [150]; (3) AFM characterizes single protein molecules in solution, which might have some flexibility [151] compared to the dimensions determined from the rigid crystal structure of the protein [152].

Following, CTPR20-Cys monolayer assembly on gold surface is studied. CTPR20-Cys at $2.5 \mu\text{M}$ concentration, which guarantees full coverage of the surface according to the equilibrium immobilization studies by QCM (Figure 24.3), is deposited onto annealed gold surfaces. The AFM image shows a densely packed array of protein particles that completely covers the gold surface (Figure 26.B). Under the solution deposition conditions and in the range of protein concentration used (10 nM- $2.5 \mu\text{M}$), CTPR20-Cys protein is stable, folded, and fully monomeric [76, 123], thus is not expected the deposition of multimeric particles. The AFM images at saturation provide insights about the tight packing of the protein molecules on the monolayer and clearly show a full

coverage of the surface (Figure 26.B), corroborating a higher occupancy than expected from simple random distribution of individual CTPR20 monomers. The overall roughness analysis of the surface resulted in a RMS roughness of 0.86 nm with an average height of 3.42 nm, smaller than the one observed for individual CTPR20-Cys proteins (RMS roughness of 1.88 nm and average size of 6.36 nm), in agreement with a compaction of the molecules. The CTPR20-Cys covered surface shown in Figure 26.B is incubated with an excess of dithiothreitol (DTT) that will displace the previously adsorbed CTPR20-Cys, in order to study the specificity of the interaction between CTPR20-Cys and the gold surface. After the incubation and wash of the surface the AFM image (Figure 26.C) shows some dots corresponding to individual protein molecules, while most of the protein has been detached from the surface. This result confirms the specificity of the thiol-gold immobilization of the CTPR20-Cys.

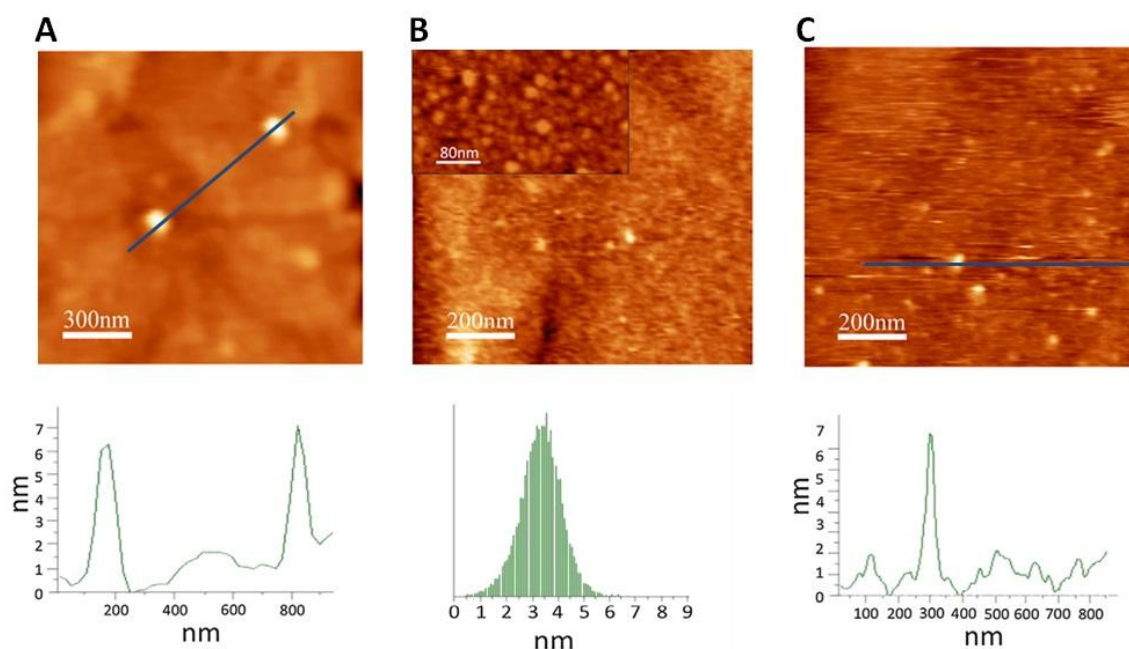


Figure 26. AFM image of CTPR20-Cys protein immobilized on gold surfaces. A) AFM image of individual CTPR20-Cys molecules deposited on gold surface at 10 nM protein concentration after 1 hour incubation at room temperature. Each bright spot corresponds to a single protein. The right panel shows the height profile of two single protein molecules. Note that the nanometer-scale widths obtained by AFM are not reliable due to the size and shape of the contacting probe. B) AFM image of a large area of completely saturated CTPR20-Cys monolayer deposited on gold surface at 2.5 μM protein concentration. The inset shows a zoom-in image in which individual protein units can be observed. The panel below shows the global roughness analysis of the image that showed an RMS roughness of 0.86 nm and an average size of 3.42 nm. C) AFM image of the same region as image B after the addition of DTT to the sample in order to remove the CTPR20-Cys. The panel below shows the height profile along the line to show a single protein molecule.

In order to determine the thickness of the monolayer imaged by AFM, profilometry experiments are performed on a CTPR20-Cys layer deposited on one half of a gold surface using 2.5 μM protein concentration. Three protein layers are measured to give a thickness

value of 17.93 ± 1.05 nm (Figure 25.B), consistent with the formation of a single CTPR20-Cys protein monolayer.

In addition, the ratio between the motional resistance (ΔR) and the frequency (ΔF) provides information about the viscoelastic properties of the deposited protein layer [153]. In Figure 27.A.1, the ΔR during the course of CTPR20-Cys protein assembly is plotted against ΔF . Two separate regions are observed with different $\Delta R/\Delta F$ slopes indicating that surface coverage has a strong influence on the viscoelastic properties of the protein film. The first region (ΔF from 0 Hz to -20 Hz) shows a linear relationship between $\Delta R/\Delta F$ presenting a steeper slope. It corresponds to the beginning of the adsorption process associated to lower protein densities and therefore less lateral interactions between protein molecules. The deposit tends to be more flexible and more dissipative at this stage. When higher protein densities are reached (ΔF from -20 Hz to -60 Hz), protein molecules can establish lateral interactions with neighbouring molecules, yielding finally to a more rigid and less dissipative film. The change in slope between two regions indicates the transition from individual proteins to a compact state. The intersection of the two straight segments from the linear fit occurs at 22.2 Hz and at a $\Delta R/\Delta F$ ratio $0.130 \Omega \text{ Hz}^{-1}$.

To confirm that the observed difference in $\Delta R/\Delta F$ between high and low protein density on the surface can be attributed to the protein compact state, experiments using different protein concentrations and recording ΔF and ΔR simultaneously are performed (Figure 27.A.2). It is observed that at low protein concentrations the $\Delta R/\Delta F$ ratio increases as the protein concentration increases, associated to a higher viscoelasticity of the protein film, reaching a maximum of $0.124 \Omega \text{ Hz}^{-1}$ at $0.075 \mu\text{M}$ of protein. It is noteworthy that this $\Delta R/\Delta F$ value is comparable to that obtained from the intersection in Figure 27.A.1. This result shows that at low concentrations there are not enough neighbouring CTPR20-Cys molecules immobilized for them to assemble, as previously suggested. When the protein concentration is increased, a subsequent $\Delta R/\Delta F$ decrease is obtained until reaching a stable value of $0.06 \Omega \text{ Hz}^{-1}$ at surface-saturating protein concentrations. The surface becomes more densely packed with CTPR20-Cys molecules, promoting lateral interactions and leading to an increase in the rigidity of the film. These results clearly suggest that the mode of attachment through the C-terminal Cys residue leads to a spatial arrangement, with parallel alignment between protein molecules, favoring the establishment of lateral protein interactions. In the case of CTPR20 protein without cysteine this analysis shows a different behavior with a linear increase in the viscoelasticity of the deposit as the protein is adsorbed (Figure 27.B).

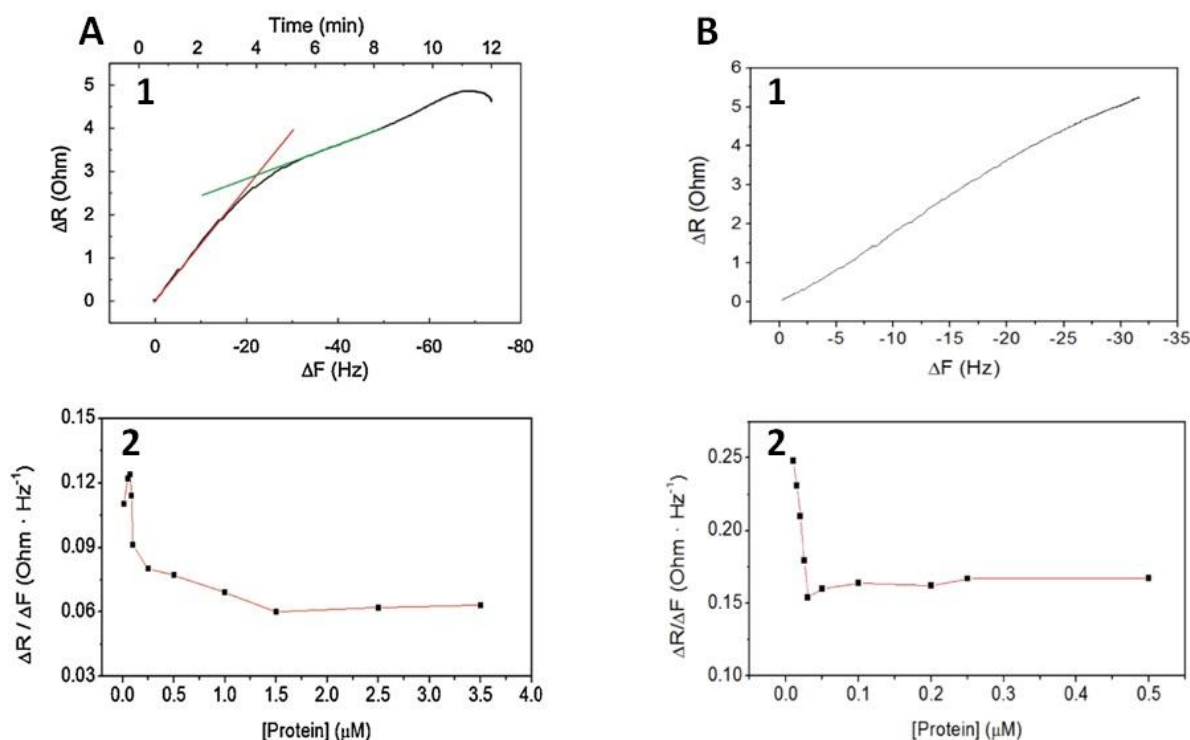


Figure 27. Viscoelastic properties of deposited protein material measured by QCM. A) **1.** The change in resistance during the course of protein immobilization and assembly kinetics vs. the change in frequency for CTPR20-Cys at 2.5 μM protein concentration. For CTPR20-Cys two regions are observed. The fit to linear equations with two different slopes are shown. **2.** The change in resistance divided by the change in frequency ($\Delta R/\Delta F$) for different protein concentrations of CTPR20-Cys measured after steady state was reached. B) **1.** The change in resistance during the course of protein immobilization and assembly kinetics vs. the change in frequency for CTPR20 at 0.25 μM protein concentration. **2.** The change in resistance divided by the change in frequency ($\Delta R/\Delta F$) for different protein concentrations of CTPR20.

1.4.2.3. Conclusions

In conclusion, we demonstrate that we can immobilize oriented long designed CTPR proteins on gold surface through thiol–gold specific interaction and that the protein modules present intrinsic lateral assembly properties. By controlled immobilization, we generated a tightly packed protein SAM and characterized the packing by QCM, SPR, and AFM. We monitored the assembly transition by the change on the viscoelastic properties of the immobilized materials. It is clear that the oriented immobilization of the long protein arrays allows the arrangement of the modules and the formation of a rigid protein film on the surface. Regarding the mechanism of interaction between the proteins and the forces that drive the assembly, we can speculate based on the previously observed lateral packing of superhelices in solid films [103] and in crystal forms [124]. Long CTPR proteins are structurally rigid with anisotropic shape (Figure 23.A), that allows for preferential alignment along one direction and lateral packing when the proteins are immobilized directionally oriented on a surface. The crystal packing provides hints that

CTPR protein superhelices have a tendency to interact through salt bridges between positively and negatively charged residues on the protein surface [76]. Similar interactions can mediate side-to-side packing within monolayer.

This work is an example where the molecular understanding of the biomolecular blocks and the control of the immobilization and assembly process on surfaces gives the possibility to generate monolayers with different packing grade. For example, by changing the protein packing density, the rigidity of the protein film can be tuned at will. This is very interesting for different applications, as we mentioned at the beginning of the section [137-139]. Also, because of the chirality of CTPR proteins, the formation of CTPR SAMS has great potential in the field of spintronics, that is an emerging field with tremendous applications in the field of nano-electronics [154-156].

1.4.3. CTPR nanotubes: 3D structure

1.4.3.1. Motivation and objectives

In this section, we describe the assembly of the CTPR proteins into a 3D protein nanotube. Looking at the superhelical structure formed by long CTPR proteins one can predict that by generating novel interactions between CTPR molecules, they could assemble into closed protein nanotubes. For this purpose we modify the CTPR proteins to engineer novel interfaces along the superhelix that will allow the interaction between two CTPR superhelices to assemble into a closed 3D nanotube.

Nanotubes are nanometer-sized tube-like structures. They are elongated nanostructures that have a defined inner hole. Nanotubes are attractive because of the large internal surface area and confined inner cavity that provide potential for their application in filtration, encapsulation and release of small molecules or drugs [157], catalysis due to the increased local concentrations [158], optics and electronics [159, 160], among others. Over the past few decades, researches have made significant progress in the generation of covalently bonded nano-structures. However, non-covalent nanotubes have become subject of major interest due to their significant advantages including facile synthesis, self-organization, control in diameter, size and high efficiency.

A number of protocols have been reported for the preparation of the non-covalently self-assembled nanotubes using different structures including rod-like units, helical structures and stacked rings. Up to the date, Fmoc-dipeptides [161], cyclic β -sheet stacking peptides [162], lock-washer α -helical bundles as the building blocks [163-165], and short peptides that self-assemble into spiral tapes [166, 167] have been used for nanotube assembly. Also, recently has been presented a generic modular approach to assemble nanotubes from α -helical barrels (α HBs) [168-170]. Moreover, there are recent works in which by self-organization of β -sheet stacking peptides are achieved systems with tailored inner diameters via the ring size of the peptide [160, 171].

Although these works represent an improvement on the control of the biomolecule-based nanotubes formation, most of them used small peptides as building blocks. However, there are no significant works in which proteins are applied as molecular units to generate synthetic nanotubes. Proteins provide more versatility than peptides, giving the opportunity to expand the applications of the use of formed nanotubes. In this chapter, we describe the use CTPR proteins to engineer protein-based nanotubes.

1.4.3.2. Results and discussion

CTPR arrays contain multiple repeats that interact through a single inter-repeat interface to form elongated superhelices (Figure 28.A). Previously it was hypothesized that for repeat proteins the introduction of a novel single flat interface that gets repeated along the molecule will allow the interaction between two units containing this interface as shown in Figure 28.C [172]. The CTPR repeat units twist between consecutive repeats forming a superhelix. Thus, the introduction of a second repeated interface along the CTPR superhelix (Figure 28.B) will allow two superhelical CTPR molecules to assemble into closed protein 3D nanotubes (figure 29.C).

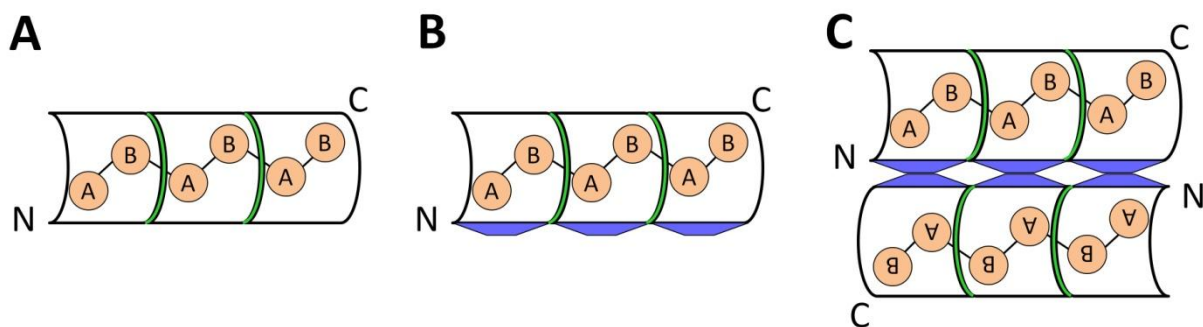


Figure 28. Schematic representation of CTPR proteins inter-repeat interactions. A) Each CTPR repeat is composed of two helices (A and B) shown as orange circles. Each repeat packs against the two nearest neighbouring repeats through a single inter-repeat interface colored in green. B) Schematic representation of the introduction of a second repeated interface (in blue) along the CTPR superhelix. C) The introduction of the second repeated interface between the building blocks leads to a novel interaction between the CTPR proteins to form a dimer.

Based on the crystal structure of the CTPR protein, it is hypothesized that mutating the amino acids in the loop of the CTPR protein (Figure 29.A) and extending this modification over all the CTPR superhelix (Figure 29.B), a novel interface will be formed. The residues involved in the generation of this interface are glycine 15 (G15) and aspartic 31 (D31). In the predicted nanotubular structure, the residue 15 between helix A and B of one CTPR protein, will interact with the residue at position 31 of other CTPR protein (Figure 29.C). This interface is designed to be established by hydrophobic interactions that will lead to a strong interaction between two monomers, [173, 174] resulting in the dimerization of the superhelices. Thus, G15 and D31 are mutated to leucines (Figure 29.A). These modifications should generate of the novel interface that will allow the interaction between two CTPR proteins to assemble into a closed protein nanotube (Figure 29.C) that is energetically more favorable than the monomeric protein units [172, 175]. The G15L and D31L mutations have been introduced at each of the 6 repeats in a CTPR protein with

6 identical repeats to form the protein named (Figure 29.B) (*Materials and Methods M.1.3.1*).

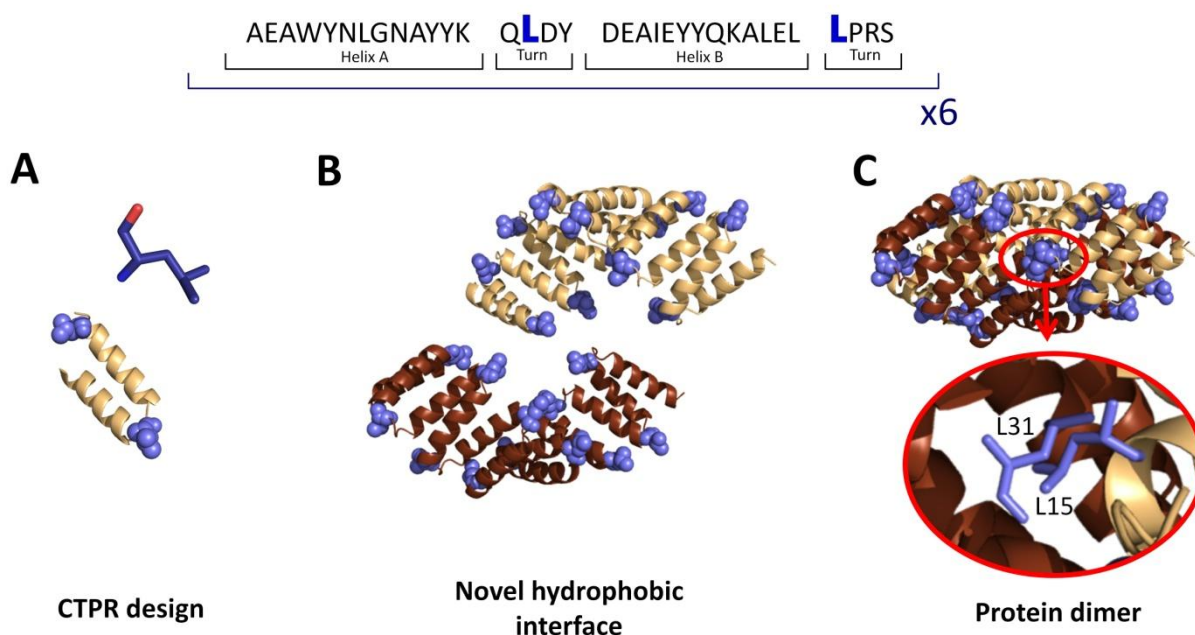


Figure 29. CTPR design for the formation of protein nanotubes. A) A representation of a CTPR modified module with leucines at positions 15 and 31 (G15L, D31L). The introduced leucines at positions 15 and 31 are highlighted with spheres in blue color. B) C6L protein, formed by 6 repeats of modified CTPR G15L D31L module. The figure shows 2 identical C6L proteins colored in light and dark brown color. The leucines that form the novel interface in the C6L protein are represented as blue spheres. C) On the top, the nanotube is formed by the interaction of two identical C6L proteins through the hydrophobic interaction between the leucines located at the novel interface. On the bottom, zoom in of the interacting region between the two CTPR proteins, where L31 of one CTPR protein is interacting with the L15 of the other.

The designed C6L CTPR variant with the novel interface is expressed and purified (*Materials and Methods M.1.3.1*). The oligomeric state is analyzed by gel electrophoresis, showing the presence of two main bands one corresponding to a monomer and another to a dimer (Figure 30.A) (*Materials and Methods M.1.3.2*). The dimerization process is driven by the hydrophobic interactions between the novel hydrophobic interface of the C6L protein. To separate the monomeric and the dimeric from size exclusion chromatography is used (Figure 30.B) (*Materials and Methods M.1.3.3*). The elution profile shows three main peaks that in the SDS-page gel correspond to a different oligomeric states: the first peak at elution volume 8 ml, corresponds to a dimer in the SDS-page gel (58.94 kDa); the second peak at elution volume 9 ml, corresponds to a mixture of dimer and monomer; and the third peak at elution volume 11 ml, corresponds to a monomer (29.47 kDa). The sample obtained at elution volume 8 ml is going to be called as C6L_{dimer} and the sample obtained at elution volume 11 ml is going to be called as C6L_{monomer} in the text. MALDI-TOF is performed in the C6L_{dimer} (*Materials and Methods M.1.3.4*). Although this fraction corresponds to a dimer in the SDS-page gel, the only

observed peak in the MALDI-TOF spectrum is 29.50 kDa, that corresponds to a C6L monomer (the mass of the monomer based in the amino acid sequence is 29.47 kDa) (Figure 30.C). This result is not unexpected since the hydrophobic interaction between the CTPR proteins that leads the dimer formation could be disrupted under the conditions used to perform the MALDI-TOF experiment, obtaining only the monomer peak.

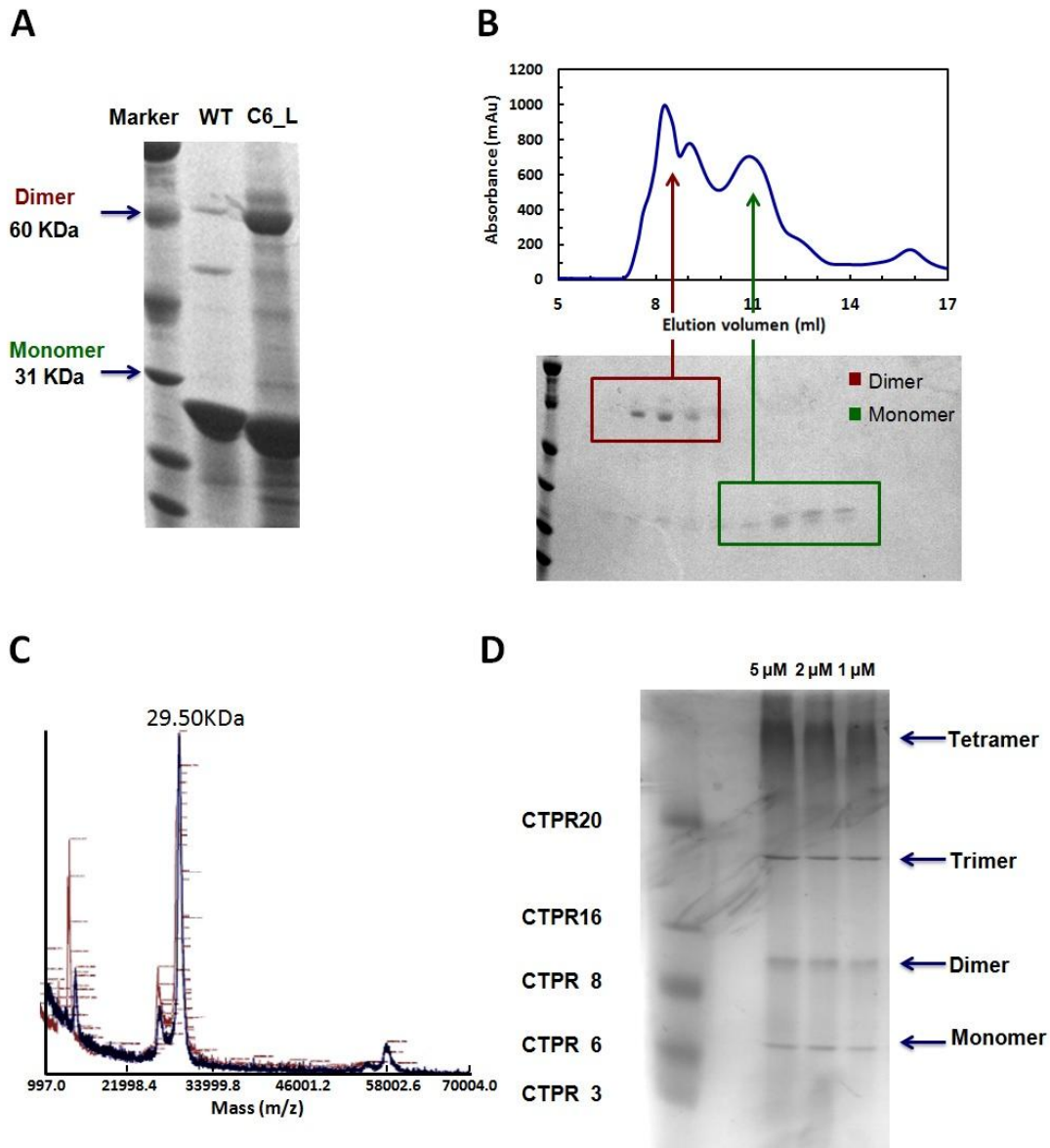


Figure 30. Characterization of the C6L oligomeric states. A) Standard SDS-page gel comparing the monomeric CTPR6 wild type protein (30 kDa) and the C6L protein, that present protein bands that correspond to the monomeric (30 kDa) and dimeric forms (60 kDa) of the C6L protein. B) On the top, size exclusion chromatogram of purified C6L protein sample. The chromatogram shows peaks at different elution volumes that correspond to populations of different sizes. The column used for the separation is Superdex75. On the bottom, the SDS-page gel of samples obtained at different elution volumes in the size exclusion chromatography. C) Mass spectrum acquired by MALDI-TOF of the C6L_{dimer} in red and CTPR6 wild type in blue. There is only one peak that corresponds to the weight of the monomer 29.50 kDa. D) Native gel of the C6L_{dimer} at different protein concentration, from left to right 5 μ M, 2 μ M and 1 μ M. In the gel monomers, dimers, trimers and tetramers are observed. In the first line a mixture of CTPR wild type proteins of different repeats as a marker.

To analyze the homogeneity of the C6L_{dimer} and C6L_{monomer} samples the hydrodynamic diameter size of the samples is measured using dynamic light scattering (DLS). The C6L_{monomer} sample resulted in a size of 7.40 ± 0.17 nm, whereas 11.54 ± 0.99 nm is determined for C6L_{dimer} sample (*Materials and Methods M.1.3.5*). The hydrodynamic diameters of monomeric and dimeric conformations are calculated from the crystal structures using the program Hydropro [128]. The obtained values based on the structure of the designed models are 4.38 nm for the monomer and 4.66 nm for the dimer (*Materials and Methods M.1.3.6*). The difference between the experimental value of the C6L_{monomer} sample and calculated value for a CTPR6 monomer could be explained because in the diameter obtained with the DLS is also considered the hydration layer formed around the proteins [176], while in the PDB of the protein this hydration layer is not taken into account. However, the difference observed between the experimental hydrodynamic diameter of the C6L_{dimer} sample and calculated value for C6L nanotube is larger than expected. A native gel electrophoresis shows that the C6L_{dimer} sample contains oligomers of different sizes (*Materials and Methods M.1.3.2*). As a marker in the native gel a mixture of different wild type proteins is used: CTPR3WT, CTPR6WT, CTPR8WT, CTPR16WT and CTPR20WT. Using this marker, in the native gel the C6L_{dimer} sample is separated in several bands: a band with a size of CTPR6WT, a band with a size between CTPR8WT-CTPR16WT, a band with a size between CTPR16WT-CTPR20WT and a band with a size bigger than CTPR20WT. These bands could correspond to a C6L monomer, dimer, trimer and tetramer (Figure 30.D). The formation of conformations bigger than the expected dimeric form suggests unspecific interactions between the protein C6L monomer.

In order to obtain structural information about the morphology of the different species that are found in the C6L_{dimer} sample, transmission electron microscopy (TEM) is used (*Materials and Methods M.1.3.7*). Different views for a C6L monomer and a C6L dimer are represented in Figure 31 based on the crystal structure of the CTPR6 protein and the designed C6L dimer structure (figure 31.A.B). These views are expected to appear in the TEM images. However, in the TEM images are not found dimers with the expected conformations, but there are oligomers with different conformations and, in most of the cases, bigger than a dimer, probably trimers and tetramers. Moreover, the different oligomers that are observed in the images have different conformation as can be observed in the Figure 31.C. This confirms the non-specificity of the dimer formation in the sample, suggesting that the residues at the designed interface are interacting in a non specific way.

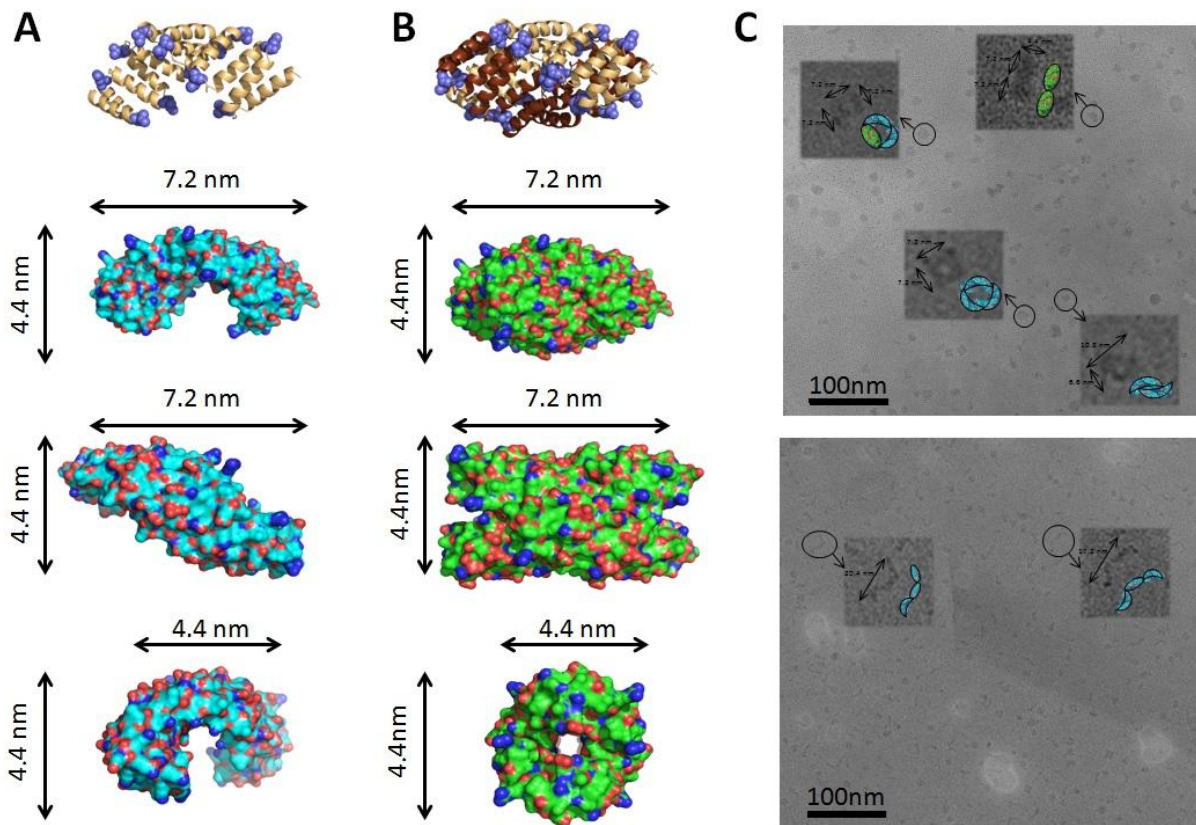


Figure 31. TEM characterization of the C6L_{dimer} sample. A) From the top to the bottom, first the structure of the C6L based on the PDB (ID: 2AVP). The protein is colored in light brown, highlighting the mutated residues G15L and D31L in blue. The next three lines correspond to three different orientations of the C6L protein showing the surface of the protein. B) From the top to the bottom, first dimeric structure of the C6L protein based on the design. The two proteins in the dimer are colored in light and dark brown, although they are the same protein, to distinguish each element in the picture. The mutated residues, G15L and D31L, and highlighted in blue. The next three lines correspond to three different orientations of the C6L dimeric structure, showing the electronic surface of the proteins. C) TEM images of the C6L_{dimer} sample. The proteins are in 10 mM NaCl, 10 mM Tris pH=7.5 and are negative stained using uranyl acetate. The sample is not homogeneous, since it is possible to observe different conformations: dimer, trimers, and tetramers. Zoom in the structures highlighted with black circles, it is possible to observe the different structures that are found in the fraction 16. In the zoom images are represented possible conformations of the molecules that are observed in the TEM image. The representations are based on the protein surfaces shown in panels A and B: the monomers are colored in blue and the dimers are colored in green.

In order to reduce the heterogeneity of the oligomeric conformations, different experimental conditions are tested changing ionic strength (the salt content in the sample) and using different detergents such as deoxycholic acid, NP-40, tween-20 and sodium dodecyl sulfate (SDS) at different percentages. For the different purified samples the hydrodynamic radius is measured using DLS, to determine the heterogeneity and the sizes of the oligomers in the sample. The secondary structure of the molecules is also analyzed using the CD to test if the α -helical structure of the CTPR protein is maintained under the different conditions.

From the tested conditions, the one that showed best results is when 0.2% of SDS is added. So, the C6L is purified using 0.2% of SDS from the first step of the purification process (*Materials and Methods M.1.3.1*). The CD analysis shows that the sample maintains the α -helical structure of the protein after SDS addition, as the CD spectrum of the sample with and without SDS has the same α -helical signal (Figure 32.C) (*Materials and Methods M.1.3.9*). An electrophoresis of the purified sample is performed using 0.2% SDS gel. Under this condition, the sample is separated in two bands that correspond to a monomer and to a dimer (Figure 32.A) (*Materials and Methods M.1.3.2*). Using size exclusion chromatography the different species in the sample are separated. The peak that corresponds to the dimer in the SDS-page gel elutes at 9 ml in the C6L sample purified with 0.2% of SDS that is later than in the sample without SDS that elutes at 8 ml meaning that the oligomers are now smaller (Figure 32.B). The sample obtained at elution volume 9 ml is going to be called C6L_{SDSdimer} and the sample obtained at elution volume 11 ml is going to be called C6L_{SDSmonomer} sample in the text. The smaller size of the population in the sample it is confirmed using DLS, where it is observed that the hydrodynamic diameter of the C6L_{SDSdimer} sample decrease from the 11.54 ± 0.99 nm obtained in the case of sample without SDS to $9.45 \pm 0,08$ nm. To characterize the stability of the sample after 0.2% SDS addition, a thermal denaturalization is performed in the C6L_{SDSdimer} and C6L_{SDSmonomer} sample. Comparing the thermal denaturalization of these samples, the dimeric form is more stable than the monomeric form, since $T_{m_monomer} = 51.04^{\circ}\text{C}$ and $T_{m_dimer} = 65.17^{\circ}\text{C}$ (Figure 32.D) (*Materials and Methods 1.3.9*).

SDS is an anionic detergent that strongly binds at hydrophobic sites of the polypeptides and proteins [177, 178]. This detergent is known to be a strong denaturant for many proteins [179-181]. However, at low concentration (below 20 mM) SDS is able to maintain the native tertiary and quaternary structure of the proteins [182, 183] and is commonly used in identification of native helix-helix interaction in transmembrane segments, because its capacity to binds to hydrophobic sites [184-188]. In our case, because of the strong interaction of the SDS with the hydrophobic sites of the protein, probably the SDS disrupt the unspecific interactions that leads the formation of big conformations, while favored the correct formation of the more stable correct dimeric structure formation.

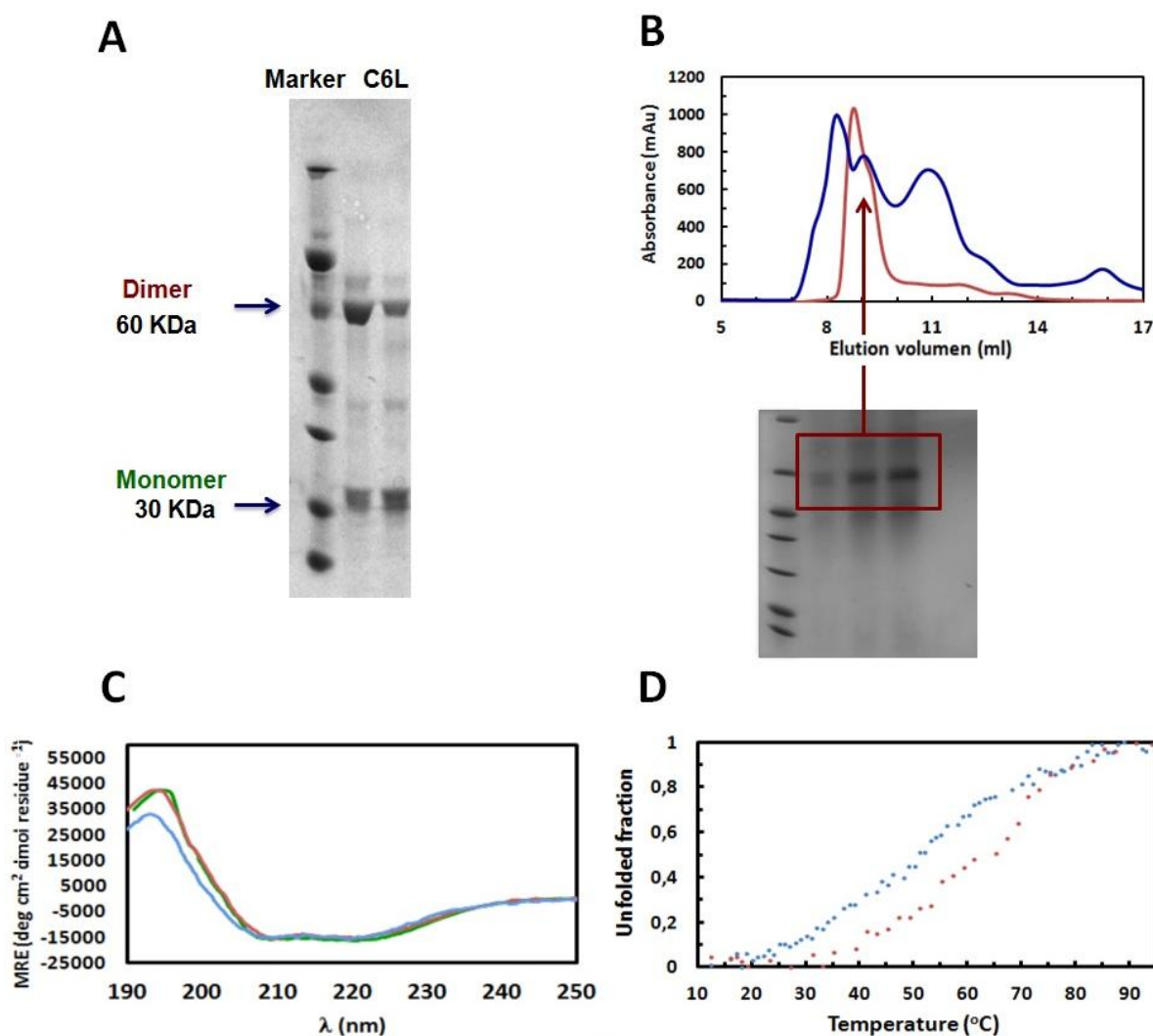


Figure 32. Characterization of the C6L sample purified with 0.2% SDS. A) 0.2% SDS gel. In the first line a wide range protein marker, in the second and third line C6L purified sample, with a band that correspond to the dimer (60 kDa) and a band that corresponds to the monomer (30 kDa). B) The chromatogram that corresponds to the elution of C6L_{SDSdimer} sample in red, and the chromatogram that corresponds to the elution of the C6L sample purified without SDS, in blue. Below, a 0.2% SDS electrophoresis gel of samples obtained from the size exclusion chromatography of C6L_{SDSdimer} sample run with 0.2% SDS at elution volume 8.5 ml, 9 ml and 9.5 ml. C) Circular dichroism spectra of the CTPR6 wild type in green, C6L_{SDSmonomer} in blue and C6L_{SDSdimer} in red. D) Thermal denaturation of the C6L_{SDSmonomer} in blue and C6L_{SDSdimer} in red.

Taking into account the results obtained, the C6L dimer with 0.2% of SDS seems more stable and more homogeneous than the sample without adding SDS. The homogeneity of the sample purified with 0.2% SDS is confirmed using TEM. Most of the molecules in the C6L_{SDSdimer} sample have the expected size for the C6L nanotube (Figure 31.B) and there are not big aggregates as in the TEM image of the sample purified without SDS (Figure 31.C). To obtain structural characterization of the formed dimer, a 2D classification is performed using the TEM images (*Materials and Methods M.1.3.8*). To do this classification, TEM images have to show individual particles (Figure 33.A). A sample protein at 0.8 μM is subjected to TEM analysis. 63 TEM images are taken. From these images, 2893 particles

are selected and classified into 20 classes using CL2D classification (Figure 33.A/B) [189]. Also, from the theoretical PDB model of the dimer (Figure 29.C), are performed theoretical projections of the dimer using EMAN software [190], taking into account the C2 symmetry (Figure 33.C) (*Materials and Methods M.1.3.8*). The obtained projections are compared with the 20 classes that are obtained from the classification of the particles in the TEM images (Figure 33.D). It is observed is that the classes obtained from the images doesn't match so well with the theoretical projections: 5/20 classes matches well with the theoretical projections (Figure 33.D/lined in green); 8/20 matches partially with the theoretical projections (Figure 33.D/lined in orange); 7/20 doesn't much with any theoretical projection (Figure 33.D/lined in red). Taking into account these results, it is not possible to conclude if the dimer has the expected structure from the original design.

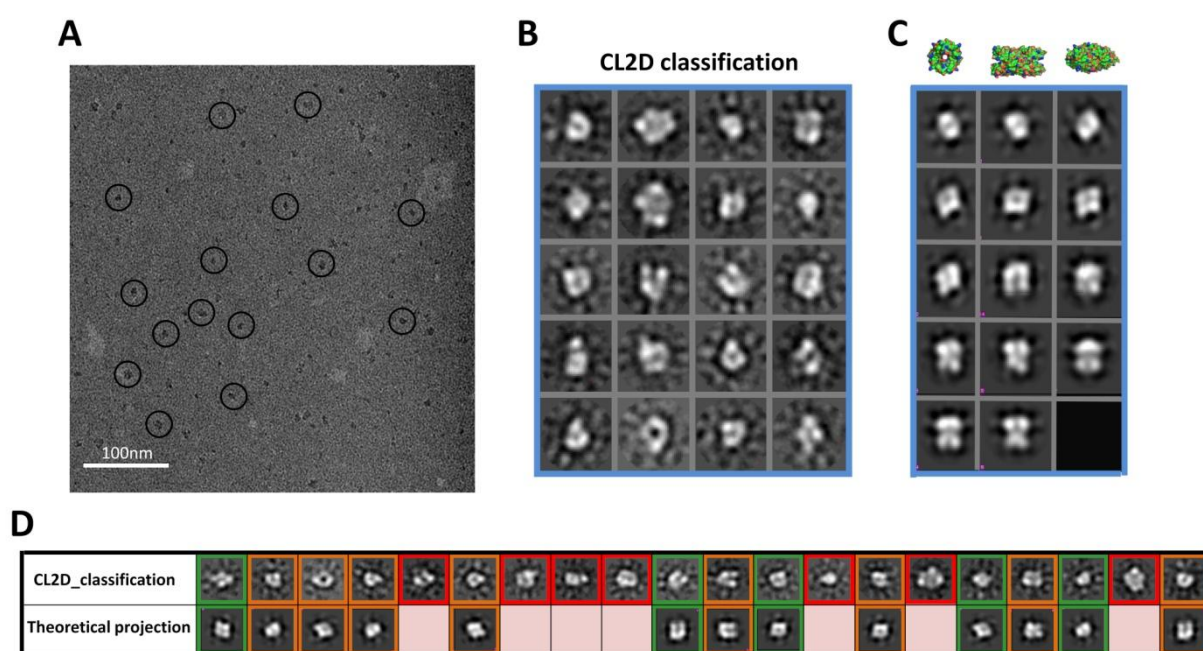


Figure 33. TEM characterization of the C6L_{SDSdimer} sample. A) General TEM image of the C6L protein adding 0.2% SDS. The analyzed sample is the C6L_{SDSdimer} sample where it is expected to have a dimeric conformation. The protein is negatively stained with uranyl acetate. The particles highlighted with black circles are some of the molecules in the TEM images which their sizes correspond to the size of a C6L designed dimer. B) 20 classes obtained by CL2D classification obtained from 2893 particles selected from different TEM images. C) Theoretical 2D projections based on the PDB of the designed dimer. On the top, theoretical electron density of the dimer structure at different views. D) Comparison between the theoretical 2D projections and real classification obtained from the TEM image. In green are highlighted the structures that are obtained in the classification of the particles of the TEM images that match with a theoretical projection based on the C6L designed dimer; in orange are highlighted the structures that are obtained from the classification of the particles in the TEM images that have something in common with a theoretical projection based on the C6L designed dimer; in red are highlighted the structures obtained from the classification of the particles in the TEM images that do not match with any of the theoretical projections.

After the TEM characterization, it is clear that there is a problem in the specificity of the nanotube formation. In order to unravel the causes of heterogeneity in the C6L sample, docking is performed using Cluspro program [191-194] (*Materials and Methods M.1.3.10*). The docking results show a large number of possible stable conformations that can be obtained from the designed monomer (Figure 34). The unspecific interactions are derived from some leucines of novel hydrophobic interface that are solvent exposed even when the tube is formed (highlighted with red circle in Figure 34.A). Because of these leucines, the interaction is not specific and different stable nanotubes could be formed (Figure 34.A). Having these stable nanotubes as starting point, it is possible the formation of trimers, tetramers and even larger oligomeric conformations (Figure 34. B/C) which are energetically more favorable than the designed nanotube (Figure 29). This agrees with the bigger structures shown by TEM (Figure 31). Once the 0.2% of SDS is added, the weak hydrophobic interactions are disrupted, because the SDS competes with these unspecific interactions, and only the dimeric conformation is obtained. However, the dimers obtained can be shifted respect the expected dimer (Figure 34.A). This would explain why the theoretical projections of the model do not agree with the classes obtained from the TEM images analysis (Figure 33.D).

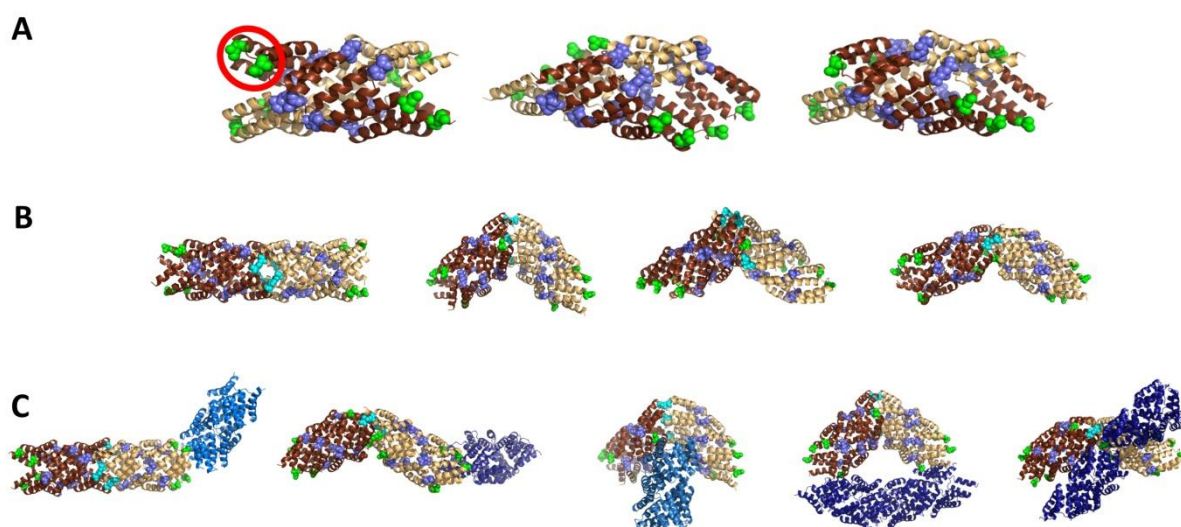


Figure 34. Different possible conformations that could be obtained from the designed C6L model using Cluspro program. A) 3 different stable possible dimers. With blue spheres the leucines from the novel interface that are interacting are highlighted and in green the leucines from the novel interface that are exposed are highlighted. The proteins are colored in dark and light brown to distinguish one from the other, but they are identical. B) 4 different stable tetramers that could be formed by the interaction of the dimers in A. In blue light are the leucines from the novel interface that are interacting highlighted and in green the leucines that are exposed highlighted. C) Different examples of stable conformations that could be get from the C6L proteins. In brown it is colored the tetramer that comes from B and in different blue colors is colored the dimer (in the three first pictures starting from the right) and tetramer (in the last two pictures starting from the right) from A and B that are docked to obtain the different possible structures.

Considering the potential problem described above in the previous design, a new model is redesigned (Figure 35.B/C) in which the leucines that remain exposed when the tube is formed in the C6L model are removed. These leucines are L31 in repeats 1 and 2 and L15 in repeats 5 and 6 and are highlighted in red in Figure 35.A. Taking this issue into account, the C6L_2 protein is designed modifying the residues L15G in repeats 1 and 2 and L31D in repeats 5 and 6 of the C6L protein. This new model will codify the formation of only one dimeric structure and avoid the competition between different conformational states.

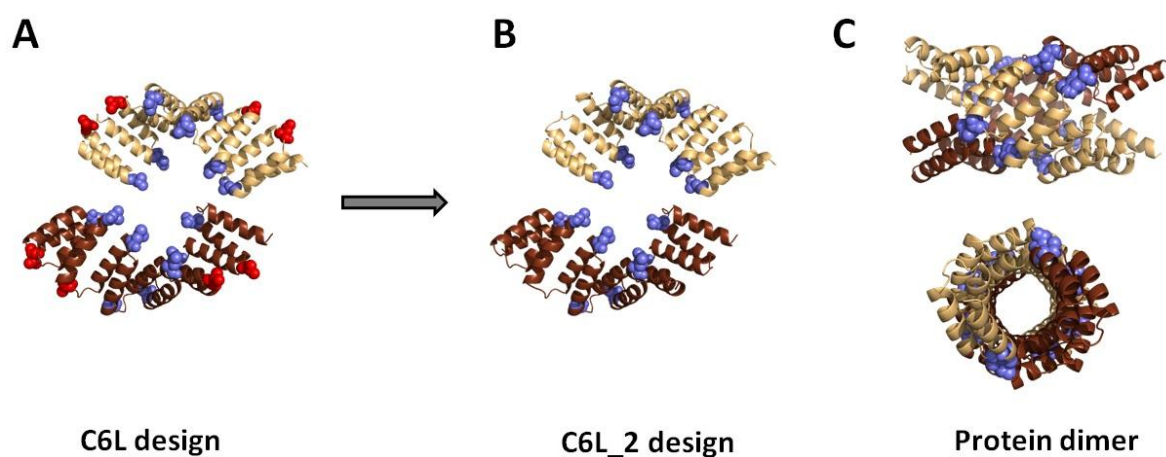


Figure 35. C6L_2 model. A) C6L protein design highlighting in red the leucines of the generated novel interface that are exposed when the tube is formed L31 in the repeats 1 and 2 and L15 in the repeats 5 and 6. B) C6L_2 protein design. There are two identical proteins colored light and dark brown color. The leucines that form the novel interface in the C6L_2 design are represented as blue spheres. C) The nanotube is formed by two identical proteins lead by the hydrophobic interaction between the leucines located in the novel interface.

After protein purification (*Materials and Methods M.1.3.1*), the C6L_2 that is obtained contains equal amounts of the monomeric and dimeric forms when analyzed by SDS-page gel electrophoresis (Figure 36.A), as in the case of the C6L protein. To separate the different species in the sample size exclusion chromatography is used (Figure 36.B) (*Materials and Methods M.1.3.2*). In this case, two peaks are obtained, the first one corresponds to a dimer in the SDS-page gel and elutes in the same elution volume as in the case of the C6L protein with 0.2% of SDS (elution volume 9 ml), suggesting that there are not large oligomeric states in the sample. The second peak corresponds to a monomer in the SDS-page gel (elution volume 11 ml). The sample obtained at elution volume 9 ml in the size exclusion chromatography is going to be called C6L_2_{dimer} and the sample obtained at 11 ml in the size exclusion chromatography is going to be called C6L_2_{monomer} in the text. The hydrodynamic diameter based on the structure of C6L_2 dimer conformation is 4.66 nm, calculated using Hydropro (*Materials and Methods M.1.3.6*) and measured value using DLS for the C6L_2_{dimer} sample is 8.53 ± 0.09 nm

(*Materials and Methods M.1.3.5*). This difference in the hydrodynamic size is around the same difference observed in the monomeric conformation and can be explained, as previously mentioned, taking into account the hydration layer formed around the protein. Moreover, electrophoresis of a C6L₂_{dimer} sample is performed using a native gel and it is confirmed that there are not big structures in the sample, since there are no protein bands larger than the dimeric form (Figure 36.C). MALDI-TOF is performed in the C6L₂_{dimer} sample (*Materials and Methods M.1.3.4*) and the only peak that is observed in the MALDI-TOF spectrum is 29.39 kDa, that corresponds to a C6L monomer (mass of the monomer is 29.36 kDa based on the amino acid sequence) (Figure 35.D). These results suggest that the hydrophobic interactions between the C6L₂ proteins in the dimer are broken, as in the case of the C6L protein.

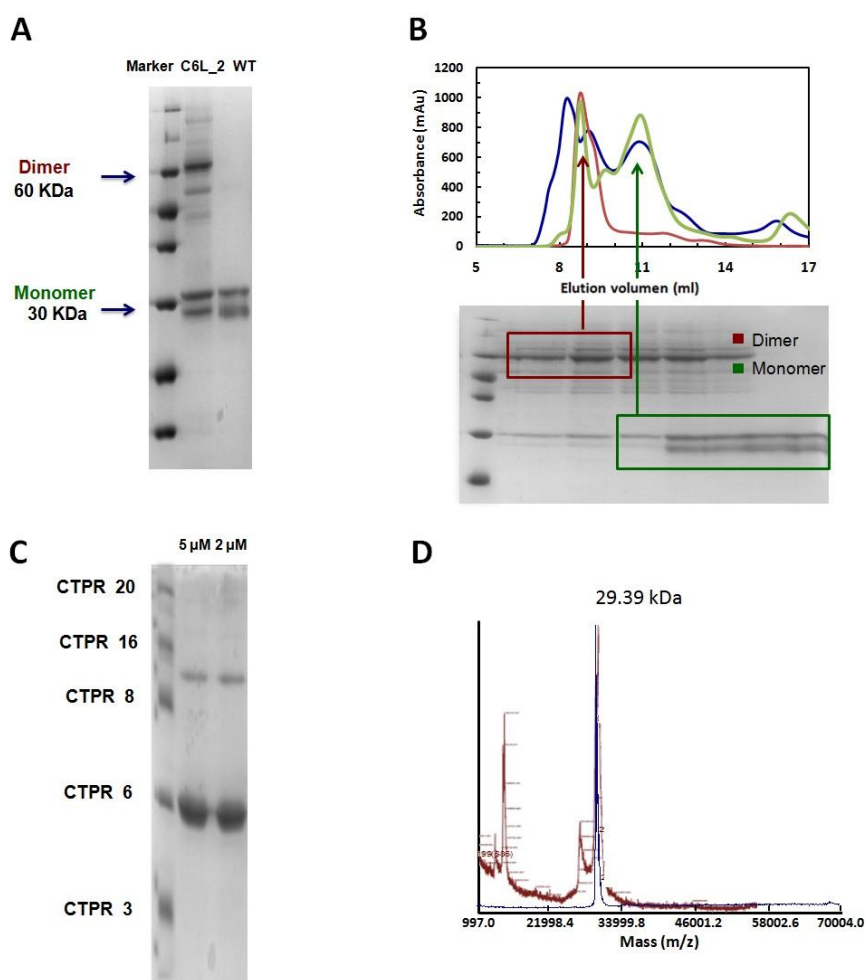


Figure 36. Characterization of the C6L₂ protein sample. A) Standard SDS-page gel comparing the CTPR6 wild type protein (30 kDa), where there is no dimer structure and the C6L protein, where there is a protein band that corresponds to the dimeric form (60 kDa) of the C6L₂ protein. B) The chromatogram that corresponds to the elution of the C6L₂ sample, in green, C6L sample with 0.2% of SDS, in red and C6L sample in blue. Below, electrophoresis gel, where there are different bands that correspond to the elution fractions obtained from the size exclusion chromatography. C) Native gel of the purified C6L₂ sample at two different concentrations (5 μ M and 2 μ M) where bands that corresponds to the monomer and dimer appears. In the first line a mixture of CTPR wild type proteins of different repeats as a marker D) MALDI-TOF mass spectrum of the C6L₂_{dimer} sample in red and CTPR6 wild type in blue. There is only one peak that corresponds to the weight of the monomer 29.39 kDa.

To determine the stability of the C6L_2 protein, circular dichroism (CD) is used (*Materials and Methods M.1.3.9*). The CD spectra show that the C6L_2_{dimer} and C6L_2_{monomer} samples maintain the α -helical structure of the protein comparing with the CTPR6 wild type protein (Figure 37.A). In the thermal denaturalization (*Materials and Methods M.1.3.9*) it is observed that the stability is not very different in the case of the C6L_2_{dimer} and C6L_2_{monomer} samples, since $T_{m_monomer}=55.08^{\circ}\text{C}$ and $T_{m_dimer}=58.09^{\circ}\text{C}$, but the dimeric conformation shows higher cooperativity than the monomeric conformation (Figure 37.B).

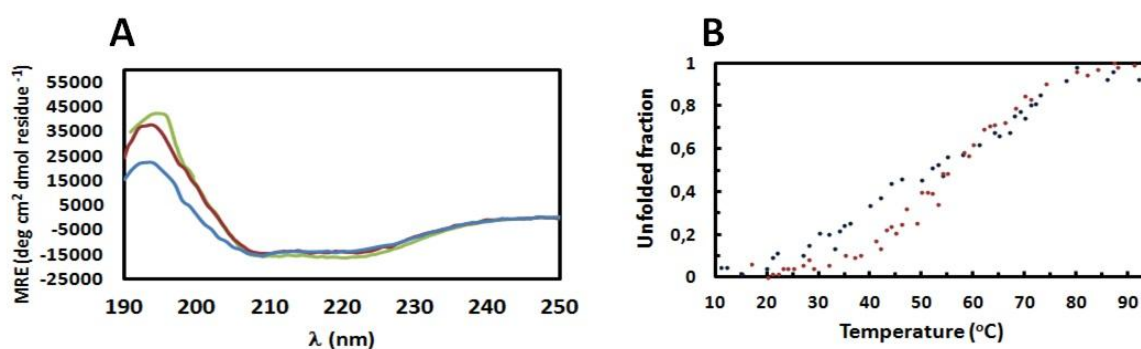


Figure 37. Characterization of the stability of the C6L_2 protein. A) Circular dichroism spectra of the CTPR6 wild type in green, C6L_2_{monomer} in blue and C6L_2_{dimer} in red. B) Thermal denaturalization of the C6L_2_{monomer} in blue and C6L_2_{dimer} in red.

The dimeric form of the C6L_2 protein seems more stable and homogeneous than the sample obtained with the previous C6L protein. In order to obtain structural characterization of the sample, TEM images of the C6L_2_{dimer} are acquired (Figure 38). In the TEM images are not observed the big aggregates present in the previous C6L protein (Figure 31.C) and most of the molecules have the expected size for the C6L nanotube (Figure 38.A). From the TEM images some particles are selected and compared with theoretical projections (Figure 33.C) expected for a C6L nanotube (Figure 38.B). Most of the selected particles match with the theoretical projections, suggesting that the nanotube is forming as expected from the model.

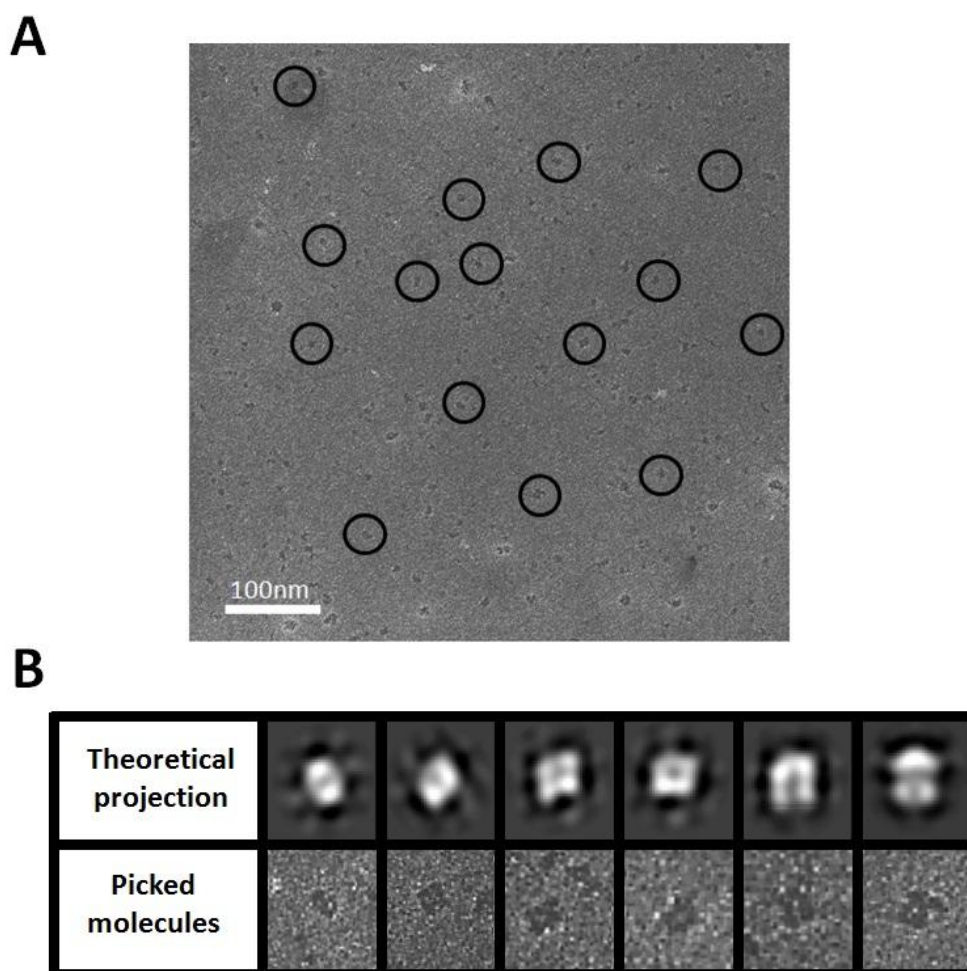


Figure 38. TEM characterization of the C6L₂_{dimer}. A) General TEM image of the C6L₂_{dimer} sample. The proteins are negatively stained with uranyl acetate. The particles highlighted with a black circle in the images are some of the particles in the image with sizes that correspond to a C6L designed nanotube conformation. B) Matching between the theoretical 2D projection based on the PDB of the designed dimer, on the top, and structure of some picked molecules in the TEM image, on the bottom.

In this section we show the structure-based design of a model to obtain CTPR protein nanotubes. The next step will be to perform a 2D classification of the different molecules of the TEM images in order to obtain the 3D reconstruction of the structure of the nanotubes, which is still a work in progress.

1.4.3.3. Conclusions

In conclusion, in this section we design a CTPR protein with a novel hydrophobic interface based on the knowledge of the CTPR protein structure. The interface is designed to lead the interaction between two CTPR proteins to form dimers that are more stable and energetically favorable than the monomeric conformation. The modification is introduced in a single CTPR module and extended over all the CTPR repeat protein.

Because of the superhelical structure of the CTPR proteins, the generated dimer should form a closed protein nanotube. However, the first hydrophobic interface designed does not lead to the formation of a unique nanotube structure due to the unspecificity of the interaction. In this first design, we induce novel interactions between the CTPR proteins, but we obtain several stable structures that are not the desired nanotube, that can be predicted using automatic protein docking. Nevertheless, we are able to change the properties of the interaction using additives such as SDS to reduce the non-specific interactions. Using the information acquired from the analysis of the obtained results, we redesign the hydrophobic interface in the CTPR protein. Using several techniques, we probe that the second redesigned model give a more homogeneous conformation, showing a more stable dimer than the first model and with a structure closer (or even the correct) to the desired nanotube.

In this section, we show how, by modifying the protein building blocks, we can introduce novel interfaces in the scaffold to encode defined higher order structures. We need to confirm that, using the second redesigned model, we obtain a unique nanotube structure. These nanotubes could be used for several applications such as filtration, encapsulation and release of small molecules or drugs, catalysis, optics and electronics, as in the case of the small peptides, but giving the robustness and additional functionality that the use of proteins provides.

1.5. Chapter 1: Conclusions

1.5. Chapter 1. Conclusions

In this chapter, we use the molecular understanding of the CTPR proteins to generate different protein assemblies. The deep knowledge of the CTPR protein structure and stability makes possible to control the self-assembly modifying the proteins in a rational way. Thus, we are able to form structures such as protein nanofibers, tightly packed monolayers, and protein nanotubes. In the different sections, we characterize the assembly process of the modified CTPR proteins and the biophysical properties of the generated structures.

The full characterization of the assembly process and the generated structures opens the door to use them as templates for different nanotechnological applications. For example, the generated nanostructures could be functionalized by introducing unique groups at which conjugate different active elements with atomic resolution. Thus, it could be possible to generate functional structures where the active elements are organized using the proteins as templates and the properties of the final hybrid structures would depend on the organization of the active elements along different length scales.

Therefore, in the next chapter, we explore the use of CTPR proteins as platforms to precisely organize at the nanoscale different active elements for applications in fields such as nanoelectronics, photovoltaics and energetics.

Chapter 2.

CTPR protein based bio-hybrid functional structures

2.1. Chapter 2: Motivation

2.1. Chapter 2. Motivation

In the last years, the development of new hybrid-materials for different nanotechnological applications, such as optoelectronics, cell signaling, plasmonics, and catalysis has attracted many research efforts. In this sense, the performance of the final hybrid-materials highly relies on the properties and organization of the molecules within the materials. Therefore many approaches are being used in order to have reliable control over the arrangement of the active components in the materials to organize them at different length scales. A key issue to achieve this organization of the components in the material is the control in the self-organization of the elements of the system and the control in the chemical recognition of the active elements by the scaffolding molecules. The major obstacle in the development of such templates is a scant understanding of atomic structure of biopolymeric materials that hinders the selective introduction of reactive functionalities.

In order to obtain novel nanomaterials where the disposition of the active components in the material is controlled, bottom-up design is the most attractive approach and has been acknowledged as a powerful strategy for the fabrication of this kind of materials [85]. For the generation of bio-hybrid materials, this strategy relies on the use of different building blocks as the basic units to introduce specific functionalities for the arrangement of the molecules and engineer defined supramolecular structures by highly specific biomolecular interactions between the components (Figure 39). To successfully apply this approach, rational understanding of the fundamental principles that govern the structure and stability of the building blocks is required. Many bottom-up approaches have been reported, for example using inorganic building blocks and nanoparticles [195, 196]. In addition, small organic templates have been also explored to control the formation of supramolecular architectures based on the organization of different molecules, at the nanometer scale, for improve their properties [197, 198]. However, these methods often do not achieve the selective orientation and positioning of the different functional components, and the control of the monodispersity at different scales is still missing.

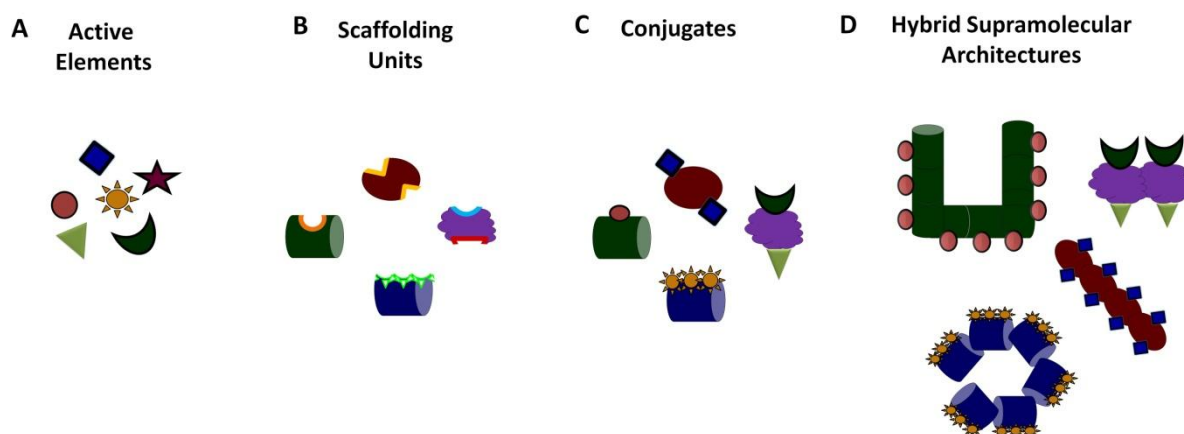


Figure 39. Schematic representation of bottom-up approach for the fabrication of functional supramolecular architectures. A) Schematic representation of different molecules with potential applications in different nanotechnological fields. Different objects represent different types of active components such as nanoparticles, metal ions, small organic molecules, and functional peptides or proteins. B) Schematic representation of different scaffolding units such as small organic templates, DNA, peptides and proteins. In this chapter we will focus on protein units as scaffolds that can be modified with different reactivities. The colored surfaces represent orthogonal reactivities that will be used to conjugate the different components. C) Conjugates in which different components are linked to the scaffolding units by selected reactivities. D) Supramolecular hybrid architectures formed by the arrangement different functionalized scaffolds. The scaffolding units are used to arrange active components into defined patterns as required to achieve optimal properties in the final structure.

However, the design and development of hybrid nanostructures using bio-inspired molecular templates has tremendous potential in nanopatterning and in the design of novel materials and functional devices [47, 71, 80, 199-204]. In fact, in the last decades it has been explored for applications in nanoelectronic devices, memory devices, and non-linear optical sensing devices [205, 206].

These methodologies enable the precise control over the structure but also over the function that can be encoded in the biomolecules. Thus, the use of bio-inspired molecular templates permits to have a high control at the nanometer scale, that is the major limitation of the conventional lithographic top-down processes used actually. Additionally, there is a growing interest to find green, easy and innovative strategies to achieve precise control at the atomic scale for the development of scaffolds and patterned structures.

In this way, the most used biomolecule for molecular templating is the DNA, because of its simplicity. For example, it is possible to arrangement gold nanoparticles (AuNPs) using DNA by introducing functionalities in the sequence of the DNA in a specific manner (Figure 40.A) [207]. Besides, DNA provides a good control over the assembly as has been reflected by the variety of two and three dimensional shapes generated by DNA origami [23, 208], where the structure is designed based on the simple rules of Watson-Crick base complementarity. Combining these characteristics, it is possible to form structures where the molecules are precisely distributed in the DNA (Figure 40.A). However, DNA can not

provide the functional and structural diversity of proteins, providing a limitation in the molecules that can be organized with these systems. Another major obstacle in the development of DNA-based templates is the fact that DNA assemblies are non-covalent and the post-assembly functionalization may destroy the structure of the system. In addition, there is a lack of understanding of the atomic structure of the final materials.

Another interesting possibility to organize active components is the use of peptides. Indeed, there are some examples where small peptides have been successfully used to organize molecules. Specially, those approaches took advantage of peptide-based fibrous systems, that are easier to control than other kind of peptide-based systems (Figure 40.B) [80]. Despite of the progress in the control over the arrangement of peptide based structures and the possibility of introducing functionalities in small peptides, it is difficult to control the order of the molecules at different length scales because of the small size of the peptides [55]. Therefore, is not easy to form biohybrid structures where the molecules are organized at nano, micro and macroscale using peptides.

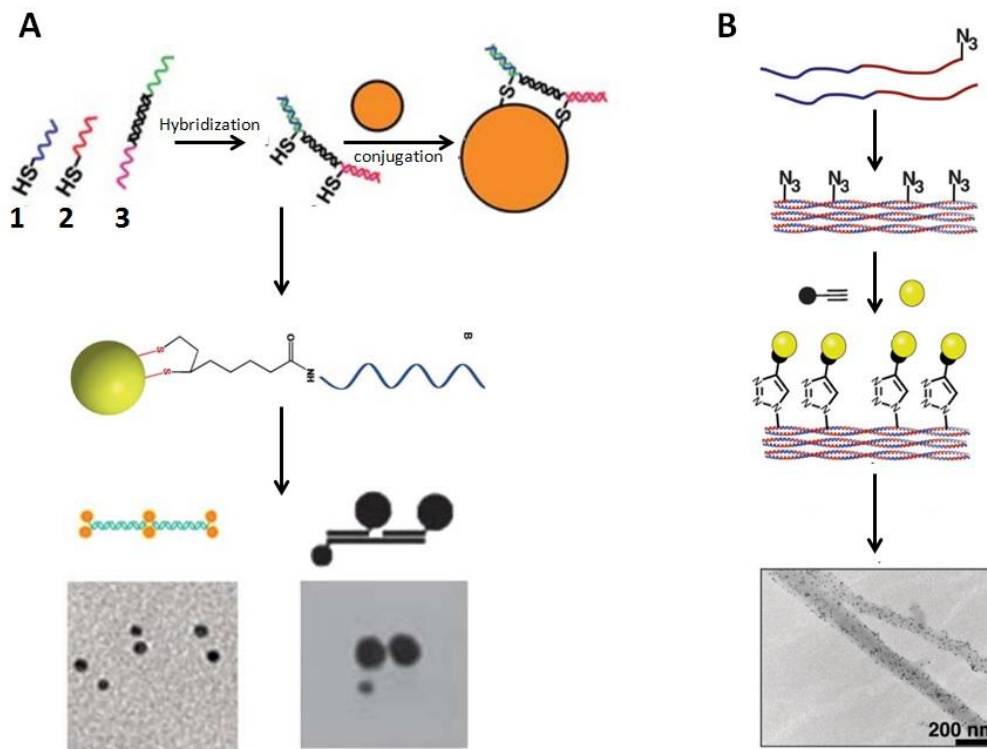


Figure 40. Examples of DNA and peptide based templating of nanoparticles. A) Gold nanoparticles templating by DNA strands. On the top, strategy for modification of the DNA with controlled spacing between the modifications. The number 1 and 2 represent small DNA fragments with a thiol group at the end of the strand. The number 3 represent the DNA strand used as a template to attach the gold nanoparticles. After hybridization process between 1, 2 and 3 the DNA is functionalized with two thiol groups that react with the gold nanoparticles. Below, DNA strand attached to a gold nanoparticles by disulfide group. In the bottom of the image, some examples where it is used DNA as a template to organize gold nanoparticles and TEM images of the conjugates. Figure from [207]. B) A click-chemistry approach for post-assembly decoration of peptide fibers incorporating azide side chains. On the top, the functionalization of peptides by azide. Below, the assembly of the peptides to form fibers with azide groups exposed and conjugation between fibers and gold nanoparticles modified with biotin-alkyne. On the bottom of the image, TEM image of the conjugates where the fibers are decorated by gold nanoparticles. Figure from [80].

Looking at the necessity of methods to control the organization of active components in the material, proteins open a new field for bottom-up approaches in nanotechnology [209]. Proteins can form organized structures with very interesting properties (Figure 4) and, in Nature, they form biohybrid materials, where different active elements such as metals, fluorophores, peptides or even DNA are conjugated to the protein to form functional structures with specific properties (Figure 5). Thus, proteins are good candidates for the development of functional bio-hybrid systems. In material science, the precise arrangement of the molecules with atomic resolution and the control of the structure of protein assemblies can help to generate bio-hybrid materials of predefined dimensions and favorable electronic, photonic or energetic properties [210]. Ideally, proteins could be used as nanoscaffolds that can be decorated with useful functionalities. As general strategy, first of all, an ideal immobilization strategy should be based on specific interactions between the protein and the molecule of interest. Once the conjugation is controlled, protein based assemblies can be used as a basis for the development of more complex nanoscale systems (nanodevices, bioelectronics, nanobiocatalysis). For this, a key requirement is to control the functionalization of proteins while retaining their activity and structure. So, the use of proteins in nanotechnology requires a control over structure and stability of proteins, protein assembly, and protein functionalization with control on the position of the element in the protein with atomic resolution.

Moreover, hybrid materials based on chiral molecules and metallic nanostructures are also of interest in chiral plasmonics applications [211], non-linear optics [212], or negative refraction index matter [213]. One example is the use of supramolecular patterning to create novel chiral superstructures of gold nanorods [214]. Also the preparation of metamaterials based on plasmonic mesophases with switchable polarization-sensitive plasmon resonances show several potential applications in liquid-crystal technology and sensing [215]. Proteins are chiral molecules because of the chirality of their individual components (L-amino acids), in addition, protein structures adopt higher levels of chirality such as the protein α -helices.

However, as explained in the section *Introduction: 1.4*, protein versatility comes with the cost of complexity and difficulty to understand the system. Thus, in the emerging field of the design of protein based nanomaterials there are only few promising works from the protein design perspective [55, 56]. In that sense, the use of designed repeat proteins as the building blocks for fabrication and patterning provides several advantages [22]. Repeat proteins present a modular structure defined by local repeated interactions and are composed of tandem arrays of the same small structural motif [23,24]. Their simple architecture makes easier to understand the basic rules that relate sequence to structure for these repeated modules and make them ideal molecular building blocks [25,26].

The use of repeat proteins as scaffolds for molecule patterning has advantages associated to their modularity. Indeed, the repetition of protein motifs allows the introduction of close and periodical reactive sites to coordinated different kind of molecules. The easy and well-controlled genetic modification and production are the key characteristics of repeat protein in this field of applications. Overall, reactive sites can be well-designed regarding type, number, and disposition among the protein, for specific functionalization with active components by different interactions. Additionally, through advanced molecular biology and biochemistry techniques un-natural amino acids are incorporated into proteins, expanding the potential reactivities of protein for bioconjugation. These strategies allow the introduction of specific binding properties to conjugate different kind of molecule in a controlled way.

2.2. Chapter 2: Background

2.2. Chapter 2. Background

To achieve the control over the stability and structure at atomic resolution that it is needed for using proteins in nanotechnology, we will take advantage of the comprehensive study of the structure and stability that has been done of the CTPR proteins in the last years.

In addition, CTPR designed protein scaffolds have been successfully used to introduce novel binding specificities by rational design in solution [216] and in solid state [103]. In particular, the Hsp90-binding residues from a natural TPR are grafted onto a CTPR3 scaffold to create a module CTPR390 with Hsp90 binding capabilities. The CTPR390 is incubated with Hsp90 and the formed complex was crystallized (Figure 41.A). The peptide appeared bounded in the concave face of the TPR, in contact with grafted Hsp90-binding residues, as in the case of natural Hsp90 binding TPRs. In order to modulate the binding affinity of the designed CTPR, the charge in the back face of the TPR was re-engineered to obtain CTPR390-, CTPR390, CTPR390+ mutants that have the recognition residues for the Hsp90 peptide but negative, neutral or positively charged back faces (Figure 41.B). The charge on the back face has significant effect in the binding affinity (Figure 41.C), the binding affinity increase dramatically, comparing with the binding affinity of the neutral back face, as the charge on the back face becomes more positive; however, with negative back face no binding was observed.

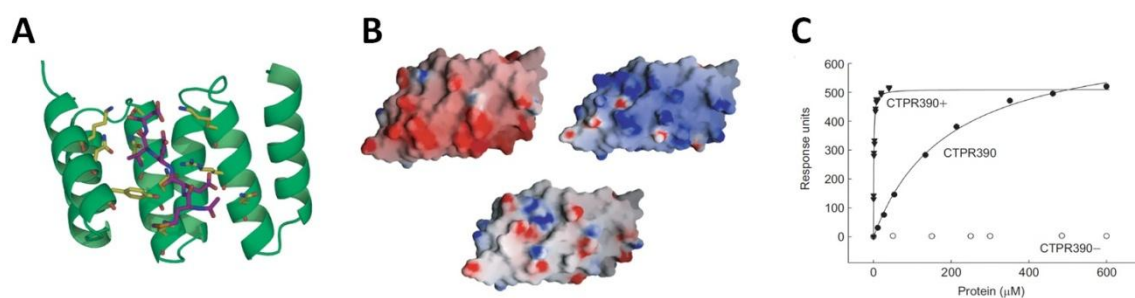


Figure 41. CTPR-peptide binding in solution. A) X-ray crystal structure of CTPR390 in complex with the C-terminal peptide of Hsp90. The backbone of CTPR390 is shown as ribbon, and side chains of residues of the TPR that directly interact with the peptide are displayed as sticks in yellow. The C-terminal Hsp90 peptide is shown as stick in purple. B) Surface representation of the electrostatic potential of the back face of CTPR390-(negative back face), CTPR390 (neutral back face) and CTPR390+(positive back face). The color range, from deep red to deep blue, corresponds to values of the electrostatic potential. C) Plot of equilibrium response levels (response units) versus protein concentration for CTPR390- (negative), CTPR390 (neutral), CTPR390+ (positive). All proteins were tested for binding to the C-terminal peptide of Hsp90. Figure from [216].

Additionally, the binding recognition properties of the CTPRs can be modulated by applying combinatorial approaches. The generation of protein libraries in which the

TPR binding interface was completely randomized allowed for the selection of CTPR modules that recognize ligands that are not target of natural TPR domains. These results demonstrate the versatility of TPRs and their potential as universal binding modules [73, 217].

In solid state, the binding capacities of the CTPR proteins are maintained (Figure 42.A) [103]. A solid film is formed using a CTPR18 that contains three designed binding sites for the C-terminal of Hsp90, CTPR390 modules alternating with CTPR3 scaffolding units (Figure 41.A). The solid film is generated in the presence of 1:10 peptide: protein molar ratio, to ensure the absence of free peptide ligand in the film. The N-terminus of the peptide is labeled with fluorescein to monitor the anisotropy of the peptide in the film by fluorescence. The film shows a strong anisotropic signal (Figure 41.B) which means that CTPR helices, and consequently the peptide-binding sites are macroscopically aligned in the film.

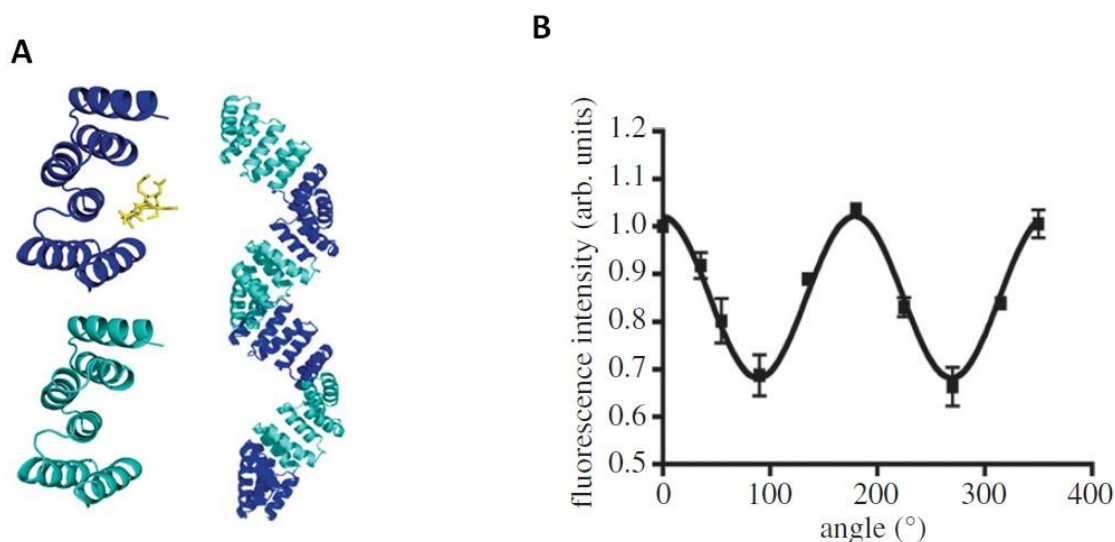


Figure 42. CTPR-peptide binding at solid state. A) Modularity of functional CTPR proteins. Ribbon representation of 3-TPR modules: peptide-binding module is shown in dark blue and the ligand peptide bound is shown in yellow sticks and the spacer 3-TPR module is shown in cyan. Ribbon representation of CTPR18 protein, made out of six CTPR3 modules, alternating peptide-binding modules (CTPR390) with spacer modules (CTPR3). B) Fluorescence spectroscopy of a CTPR18 protein film cast in the presence of 1:10 fluorescein-Hsp90 peptide:protein molar ratio. The fluorescence intensity of the peptide is plotted as function of the angle of the emission polarizer. The intensity data are shown as filled squares, the best fit of the data to a sine wave function is shown as a solid line. Figure from [105].

In these studies, it is shown that it is possible to introduce novel functionalities in the CTPR proteins to bind different kind of peptides and module the binding affinity by rational design of the protein. Also, the proteins retain their biological activity, i.e ligand binding, in the solid state and the biomolecular recognition in such materials can be used to impose order to otherwise isotropic fluorescent peptides.

The previous studies suggest that it is possible to use CTPR proteins as a particular type of biomolecular scaffolds which encompass the structural simplicity of DNA and short peptides and the functional versatility of proteins. They suggest that the use of CTPR proteins as scaffold allows the introduction of specific binding properties to conjugate different kind of molecules. Moreover, their modularity allows for the patterning of mono and multicomponent systems by having a set of scaffolding modules that carry different active components. In conclusion, by using CTPR proteins should be possible to combine the control on the assembly of the CTPR proteins with the control in functionalization of the protein modules that is required to use the proteins for the generation of bio-hybrid functional structures.

2.3. Chapter 2: Objectives

2.3. Chapter 2. Objectives

In this chapter, we will show how repeat proteins present advantages not only for the generation of protein-based assemblies as we show in the chapter 1, but also to develop the next level of complexity toward the generation of hybrid functional materials. Specifically, this chapter shows the potential of CTPR proteins as scaffolds to create bio-hybrid functional materials.

We exploited the atomic resolution knowledge of the CTPR protein, to specifically introduce functionalities (Figure 43.A) to use the proteins as a template to organize at nanoscale different active elements with atomic resolution including gold nanoparticles, donor-acceptor molecules or redox-active clusters (Figure 43.B).

Once we have the control in the conjugation process, we take advantage of the control in the self-assembly of CTPR proteins that we described in the Chapter 1 to form more complex structures by the assembly of the conjugates. In the Chapter 1 we present that under certain conditions CTPR proteins assemble into ordered structures including linear nanofibers [218], tightly packed monolayer [219] and ordered protein films [220], mimicking the packing observed in the crystal forms of CTPRs [76, 77, 102]. These results illustrate the potential of these protein modules as self-assembling building blocks. In this chapter, we take advantage of the knowledge and the control of the self-assembly of the CTPR proteins to get bio-hybrid structures where the molecules are organized in different range of scales: nano, meso and macroscale. As examples, we form gold nanoparticles monolayer, gold nanoparticles film and donor-acceptor molecule solid films. Also, we design a multifunctional system to mimic natural photosystems (Figure 43.C). Furthermore, we characterize the functional properties of the generated structures. These studies will be the basis for future functional materials and structures with many potential applications in nanotechnology.

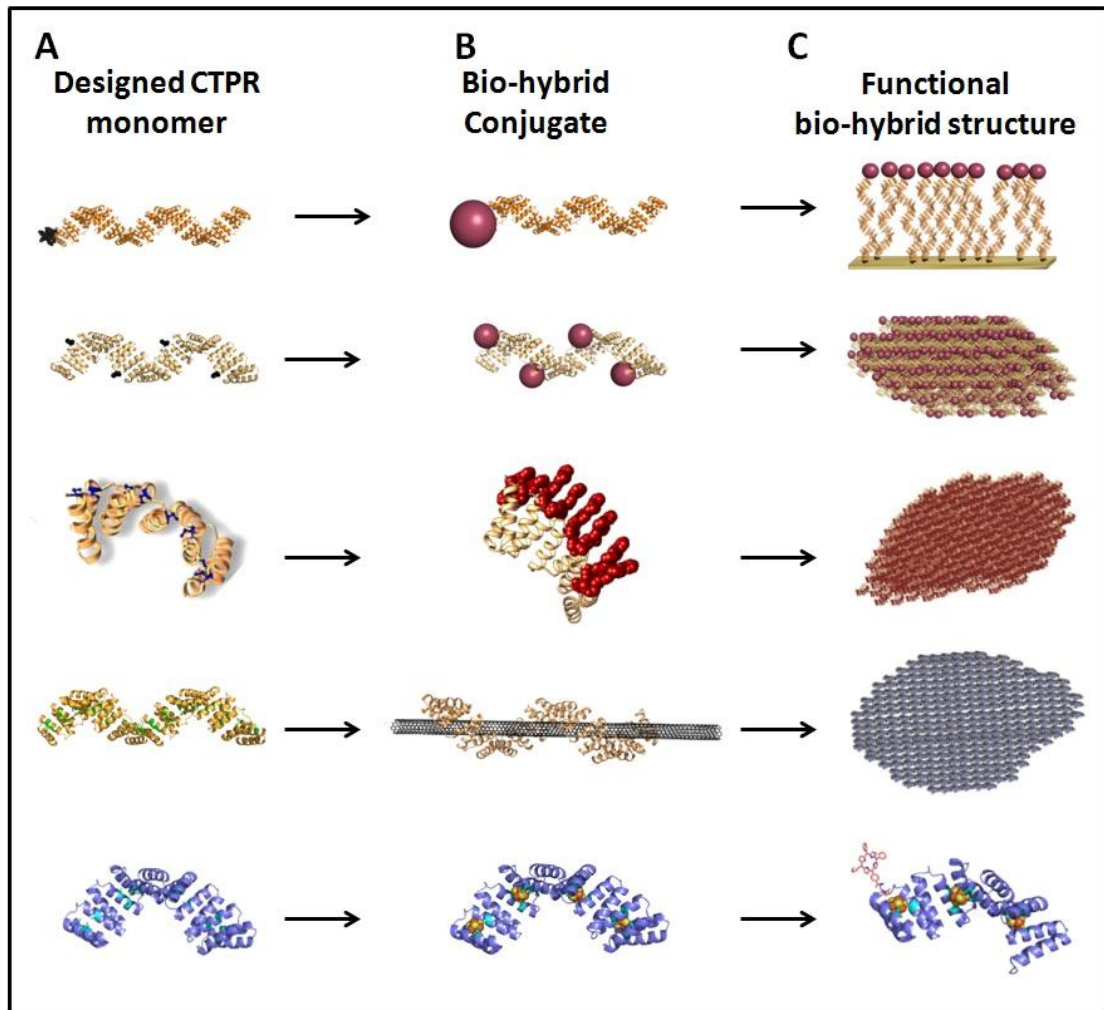


Figure 43. CTPR as template to generate functional bio-hybrid structures. A) Modified CTPR protein structure highlighting the modifications introduced in the CTPR protein. B) Conjugation between the protein and different active components. From top to the bottom: gold nanoparticles, porphyrins, carbon nanotubes, and [4Fe4S] clusters. C) Assembly of the conjugates to form different functional structures based on CTPR proteins.

2.4. Chapter 2: Results and discussion

2.4. Chapter 2. Results and discussion

2.4.1. Templating of gold nanoparticles by CTPR proteins

2.4.1.1. Motivation and objectives

In this section, we explored the use of designed repeat protein scaffolds for templating gold nanoparticles (AuNPs). The structural understanding of the protein scaffold allowed us to accurately design at atomic-level the positions in which the gold nanoparticles are coordinated. Furthermore, the control self-assembly of the molecular scaffolds provides the opportunity to form higher ordered nanostructured fibers, monolayers and thin films with the hybrid CTPR-AuNPs conjugates. This is very interesting in order to create conductive structures where the gold nanoparticles are organized in a controlled way which is necessary to use them for the fabrication of nanoelectronic devices, nanowires or plasmonic sensors.

Among the numerous research directions to explore this scientific field, the use of self-assembling biomolecules for the generation of ordered hybrid nanostructures is an innovative and challenging strategy. The main challenge is to develop synthesis and conjugation strategies that allow a stable link between metallic nanoparticles and scaffolds while keeping the stability and the properties of the scaffolding materials. In this sense, there are some works that reported how metallic nanoparticles can be adsorbed or covalently bond onto macromolecules including DNA templates [208] protein or polypeptides [221-224], polysaccharides [225], microtubules [226], enzymes [227] or virus [228-232]. Moreover, some pioneer works have opened the way studying the use of various types of biomolecules, such as, DNA [233], polypeptides [221, 234] or virus [235] as templates for ordered AuNPs for further applications in nanoelectronics. However, these works only show preliminary results, in which it is not possible to control the final structures of the material and the patterning of the particles at different length scales.

Developing nano-patterned systems with perfect control over the position, size and disposition of the elements among the material is elementary to create new small and sharp devices. In particular, designed CTPR proteins can be applied as templates to organize AuNPs and obtain different nanoparticle arrangements along the protein. In this section, we present the design and synthesis of a CTPR protein scaffold to attach functionalized AuNPs in which the exact position of the nanoparticle is controlled with atomic resolution. We show a detailed characterization of the protein-AuNPs conjugates showing that we are able to organize AuNPs using as a template CTPR proteins in two different conformations: in a monolayer and in a solid film. Finally, to test the

functionality of the nanostructures, we measure the conductivity of the films where the AuNPs are homogeneously distributed. The main goal of this work is to demonstrate that repeat proteins, and particularly CTPR proteins, can be used as scaffolds for the patterning of inorganic nanostructures, such as gold nanoparticles.

2.4.1.2. Results and discussion

2.4.1.2.a Performing AuNPs monolayers templated by CTPR proteins

In this section, we show how CTPR proteins are used to organize AuNPs in an ordered monolayer (Figure 44). In particular, here we describe how the CTPR protein tightly packed monolayer is used for the patterning of AuNPs.

Nowadays, interfacial bottom-up 2D assembly of nanoparticles has become a novel and widely used strategy to organize particles in a surface. However it presents some limitations, including the difficulty to assemble nanoparticles with different coatings and mostly the low adhesion of the film when transferred to the substrate [236]. Thus, we propose the use of CTPR proteins as a simplified system where the exact position of the conjugation in the protein can be controlled, as a new strategy to organize particles in a surface using bottom-up approaches.

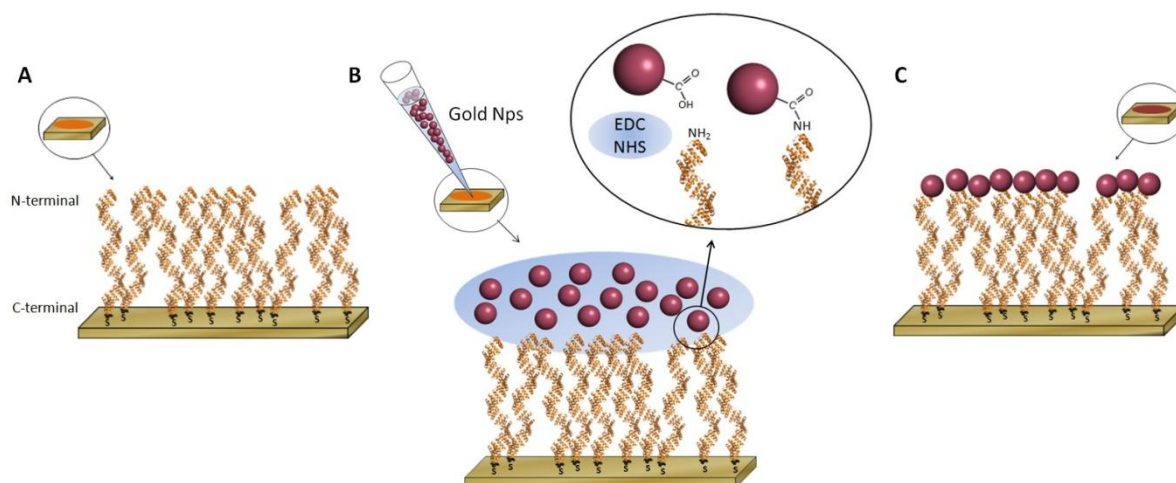


Figure 44. Strategy of the monolayer formation. A) Schematic representation of CTPR20-Cys adsorption onto gold surface through thiol–gold interactions. B) Overview of the functionalization process of gold-adsorbed CTPR20-Cys with gold nanoparticles (AuNPs) through amide bond formation in the presence of EDC and NHS. C) Schematic representation of the AuNPs monolayer patterned by the previously adsorbed monolayer of CTPR20-Cys onto gold surface.

In the section 1.4. Chapter 1: 1.4.2 it is explained how it is possible to form protein monolayer taking advantage of the side-to-side lateral interactions between CTPR superhelices (Figure 23). CTPR20-Cys is immobilized on a gold surface through gold–sulfur bond and, driven by the lateral inter-molecular interactions, can generate ordered and oriented protein self-assembled monolayers (SAMs) (Figure 44.A). The AFM results show evidence of a tightly packed CTPR20-Cys monolayer (Figure 26) onto gold surface with a homogenous coverage of protein over surfaces in the order of μm^2 . In this chapter, it is explored the potential of the CTPR protein scaffolds to act as templates for patterning AuNPs into monolayer (Figure 44.C). As the CTPR20-Cys is selectively immobilized onto the gold surface through its single C-terminal cysteine, the N-terminal moiety of the protein is selected as the anchoring point. Each CTPR20-Cys molecule should be able to interact only with one AuNP. Thus, taking into account that the area of the protein section is about 11 nm^2 , the AuNPs that are used for the experiments have 12 nm diameter, in the same range of the protein section (*Materials and Methods M.2.1.a.1*). AuNPs stabilized with thiocetic acid (TA-AuNPs) displaying free carboxylic acid functions, required for the immobilization step, are used. The anchoring reaction is achieved through the formation of an amide bond between amine of the N-terminal of the CTPR20-Cys (previously immobilized on gold surface) and the carboxylic acid functions of TA-AuNPs, which are activated in situ by EDC/NHS (Figure 44.B) (*Materials and Methods M.2.1.a.2*).

After the reaction of the TA-AuNPs with the gold surface coated with CTPR20-Cys, scanning electron microscope (SEM) is used to characterize the functional surfaces obtained (*Materials and Methods M.2.1.a.3*). Figure 45.A shows a significant AuNPs surface coverage yield (clearer areas) in a highly packed distribution. The entire surface is not fully covered by AuNPs but CTPR20-Cys can be observed on almost the whole surface (as black/grey background). An important observation is that when the functionalization with AuNPs occurred, it forms preferentially a monolayer of AuNPs with a high level of compaction, leading to islands of close-packed AuNPs monolayers (Figure 45.A/left). At a higher magnification, the organization of the AuNPs functionalized onto CTPR20 immobilized monolayer can be clearly confirmed (Figure 45.A/right). The controls of AuNPs onto gold surface and AuNPs onto CTPR20 without coupling agents give very different results in SEM, with a majority of isolated AuNPs aggregates and/or randomly dispersed (Figure 45.B.C).

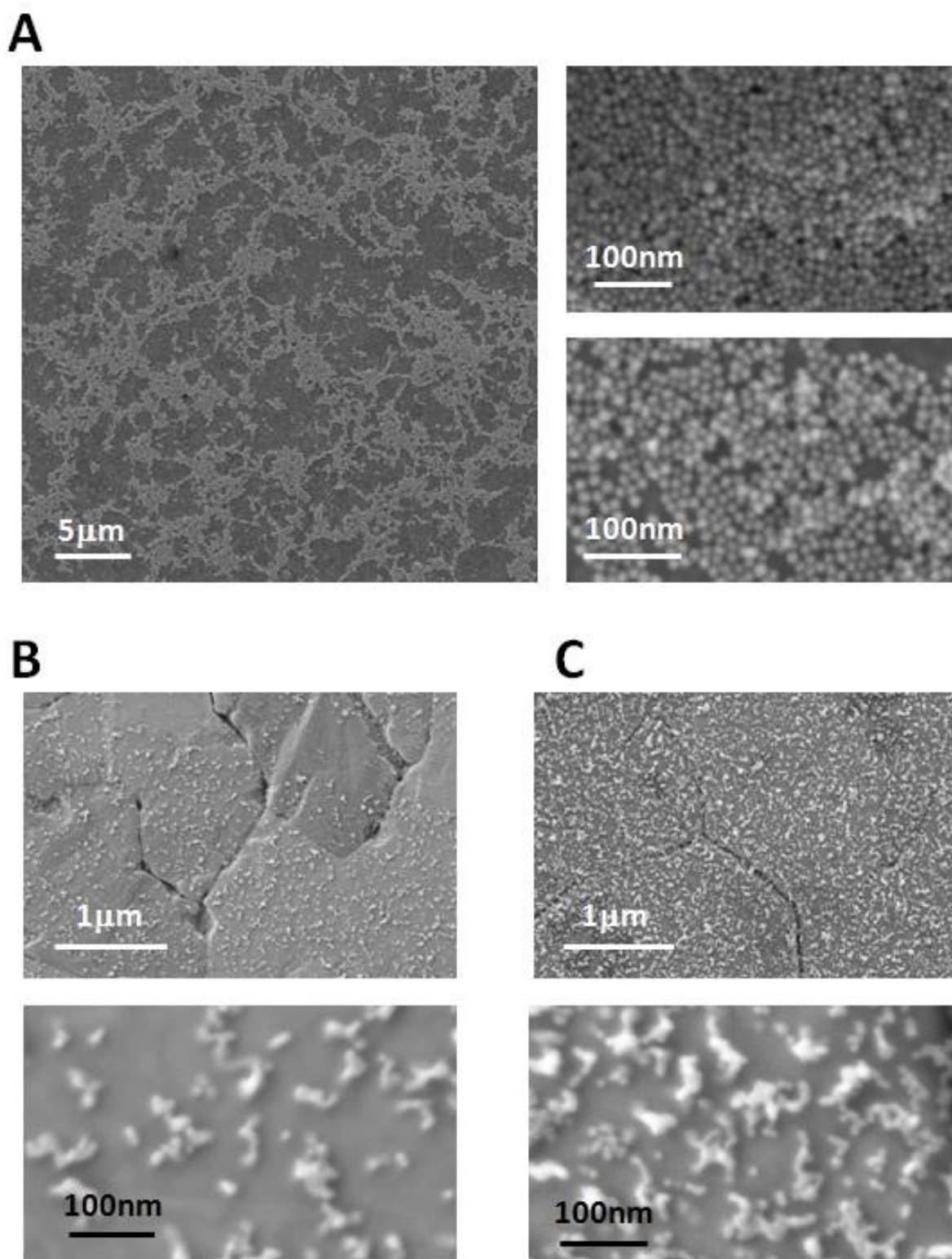


Figure 45. AuNPs deposition. A) On the left, SEM image of a large area ($30\ \mu\text{m} \times 30\ \mu\text{m}$) of gold surface covered by CTPR20 and functionalized with AuNPs. On the right, representative SEM images of AuNPs immobilized onto CTPR20 covering gold surface. Two areas in which is clear the close packed AuNPs monolayer are displayed. B) SEM images of AuNPs deposited directly onto gold surface, at different magnifications comparable with magnifications of the images in A. C) AuNPs deposited onto CTPR20-Cys without EDC/NHS coupling agents at different magnification comparable with magnification of the images in A.

The comparison of both AFM (Figure 46.A) and SEM (Figure 46.B) at the same scale confirms the monolayer organization of the AuNPs with the same kind of organization observed for the protein onto gold surface, presenting in both cases a high compaction. This compaction is not observed when the AuNPs are deposited on gold surfaces without prior CTPR20-Cys protein layer formation (Figure 46.B.C), which corroborates that the CTPR20-Cys monolayer is templating the AuNPs.

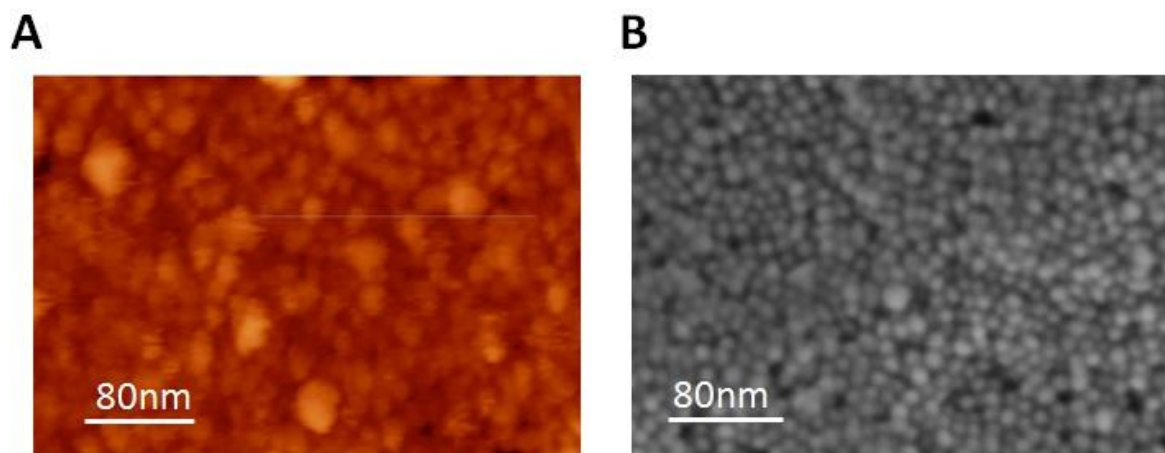


Figure 46. AuNPs template by tightly packed protein monolayer. A) AFM image of the CTPR protein monolayer. B) SEM images of the AuNPs monolayer in the same scale as the AFM images.

In conclusion, we use the control in the CTPR oriented monolayer formation (N-terminal free) to use the monolayer as a template to organize AuNPs. This work is an example where the molecular understanding of the biomolecular blocks and the control of the immobilization and assembly process on surfaces can lead to the generation of materials with defined macroscopic properties. Additionally, these ordered and structured biocompatible surfaces can be functionalized by introducing unique groups with atomic resolution at defined positions in the protein. The results of this work show the potential of these designed scaffolds as platforms for patterning AuNPs. The assembly of AuNPs into closely packed monolayer has great interest because of their potential applications in different fields such as electronics, nanophotonics, data storage, plasmonics etc. as we present before. The templating strategy described here is versatile and can be easily applied to a variety of nanoparticles and other molecules for the generation of closely packed arrays. The application of these biomolecular templates can result in novel types of hybrid materials within the bottom-up approaches.

2.4.1.2.b Performing AuNPs conductive films templated by CTPR proteins

In this section, we show how CTPR proteins are used to form a conductive solid film in which the AuNPs are homogeneously distributed along whole solid film and are close enough to obtain efficient electron transfer between them (Figure 47.D). We take advantage of the deep understanding of the CTPR protein structure and stability, explained in the section *Introduction: 1.7*, to control the conjugation between CTPR protein and AuNPs and of the self-assembly properties of the CTPR proteins to form a structured solid film, explained in the section *1.2.Chapter 1: background*, to form the conductive film in which protein-AuNPs conjugates are organized at solid state.

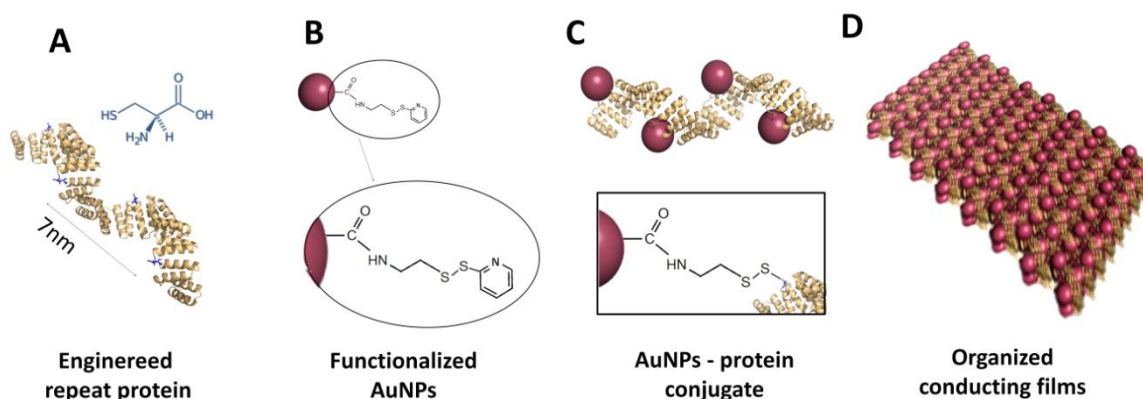


Figure 47. Schematic representation of the strategy developed to obtain ordered conducting film. A) Ribbon representation of CTPR₁₆ protein structural model based on the structure of CTPR8 (PDB ID: 2AVP). The mutated cysteines for the AuNP attachment are highlighted in blue. B) Scheme representation of the AuNP functionalize with pyridyldisulfide linker for the conjugation with the cysteines of the protein. C) Schematic representation of the structure of CTPR₁₆-AuNP conjugate. In the black box the interaction between the 2-(Pyridyldithio)-ethylaminehydrochloride linker of the functionalized AuNP and cysteine in the CTPR₁₆ protein. D) Schematic representation of the structure of a solid film generated upon the assembly of the CTPR₁₆-AuNP conjugate.

In order to generate the conductive films, first of all, it is necessary to arrange the AuNPs in the structure of the CTPR proteins with atomic resolution (Figure 47.C). The interaction between the AuNPs and the protein is controlled by specific interactions between cysteine residues introduced in the protein and modified AuNPs. For that, a CTPR₁₆ protein with 4 cysteine residues in the loop of the repeats 2, 6, 10 and 14 is designed, by the mutation R33C, leading to CTPR₁₆-4Cys protein which provides unique reactivity for the immobilization of the AuNPs (*Materials and methods M.2.1.b.1*). The distance between the cysteine in repeats 2 and 10 and the distance between the repeats 6 and 14 is approximately 7 nm calculated from the CTPR₁₆ crystal structure (Figure 47.A). These distances determine the size of the maximum diameter of AuNPs that could be used for the conjugation. These positions are not conserved and, as expected, the mutations do not significantly affect the structure or the stability of the protein scaffold

(Figure 48) [59]. A well-expressed CTPR₁₆-4Cys, stable protein sample with the same α -helical structure as the parent protein is obtained (Figure 48.A).

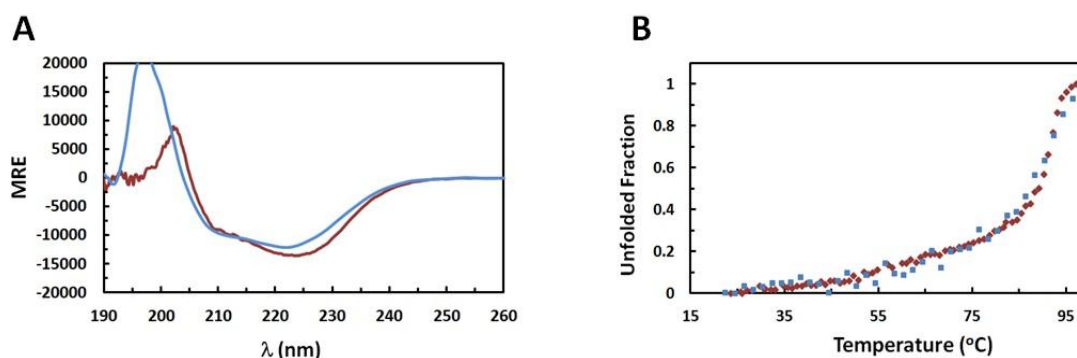


Figure 48. Comparison of the α -helical content and the stability between CTPR₁₆WT and CTPR₁₆-4Cys protein. A) CD spectra of mutated CTPR₁₆-4Cys, in blue, comparing with the original CTPR₁₆WT, in red. B) Thermal denaturalization of the CTPR₁₆-4Cys, in blue, where a $T_m = 87^\circ\text{C}$ is obtained, comparing with the thermal denaturalization of the CTPR₁₆WT, in red, where a $T_m = 89^\circ\text{C}$ is obtained.

For this study, gold nanoparticles of around 1.7 nm diameter (AuNPs) have been acquired from *NITparticles (Nanoimmunotech) (Materials and Methods M.2.1.b.2)*. This size should favor the conjugation of 4 AuNPs per protein, as is smaller than the maximum size that fits between cysteine in the CTPR₁₆-4Cys protein. As it has been deeply investigated, the photophysical properties of gold nanocrystals depend on their size (from fluorescent gold nanoclusters ($d < 1$ nm) to plasmonic properties of gold nanoparticles with sizes higher than 3 nm). The UV spectrum of 1.7 nm diameter gold nanoparticles has only a very weak plasmonic signal between 500 and 550 nm (Figure 49.B). Moreover, these AuNPs do not emit any fluorescence because the size of the nanocrystals is too big to observe the typical emission of gold nanoclusters. Unfortunately, the lack of specific photophysical characteristics does not permit the exact quantification of the AuNPs among the modification and functionalization steps.

The AuNPs are modified with 2-(Pyridyldithio)-ethylaminehydrochloride (PDA*HCl) linker, (Figure 47.B), (*Materials and Methods M.2.1.b.3 and M.2.1.b.4*), for the conjugation with the free thiol of the cysteine residues of the CTPR₁₆-4Cys protein (Figures 47.A and Figure 47.C). To quantify the attached linker in the AuNPs, DTT is added in excess to the filtrates and analyzed by UV-Visible spectroscopy. The presence of free unreacted linker in the filtrates is then monitored by the by-product formed during the reaction of PDA*HCl and DTT, pyridine-2-thione, (*Materials and Methods M.2.1.b.5*). After a minimum of 5 washes, free linker is not observed anymore in the filtrates and the typical yield observed is around 70%.

The conjugation process between CTPR₁₆-4Cys and modified AuNPs relies on the reactivity of the PDA*HCl moiety of the AuNPs linker and the free thiol function of the cysteine residues of the design protein to form a disulfide bond. The conjugation step between modified AuNPs and CTPR₁₆-4Cys is done in PBS buffer (300 mM NaCl, 150 mM Tris pH=7.4) at room temperature for 48 hrs.

After the reaction, a purification step is needed to eliminate the free AuNPs from the rest of the reaction medium (i.e. free protein, if still is in the mixture and protein-AuNPs conjugate) in order to be able to characterize the presence of templated AuNPs by CTPR₁₆-4Cys protein using high resolution techniques as transmission electron microscopy (TEM) and scanning transmission microscopy (STM). To eliminate most of the free AuNPs in the sample, the strong interaction between His-Tag of the protein with Cobalt in the Co-NTA column is used. The conjugation mixture is incubated with the cobalt resin as explained (*Materials and Methods M.2.1.b.7*). The free AuNPs appears in the flow through and in the water wash of the column. The elution with 300 mM imidazole leads to the release of the possible free CTPR₁₆-4Cys protein, if still is in the mixture, and the CTPR₁₆-AuNPs conjugate. The confirmation that the conjugation occurred can be observed directly by the colour of the elution fraction and characterized by UV-Visible spectroscopy (Figure 49.B).

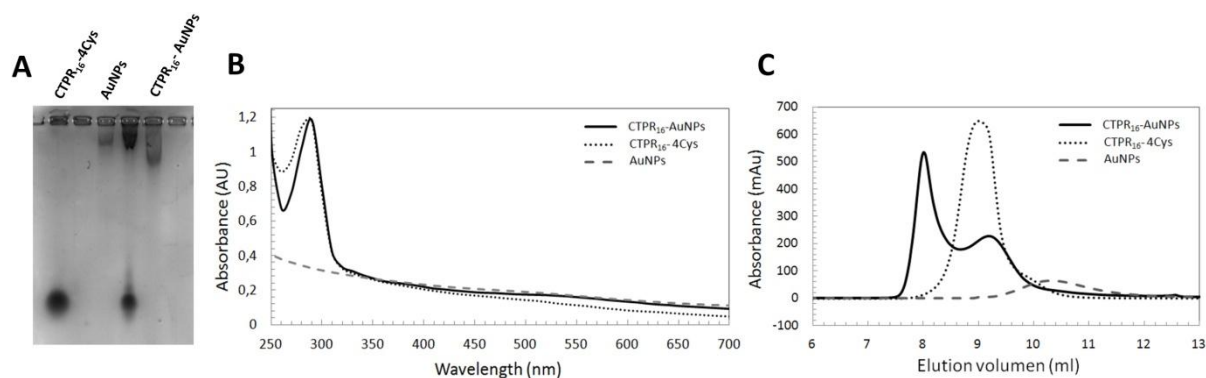


Figure 49. CTPR₁₆-AuNPs conjugation process. A) 0.5% agarose electrophoresis gel of the CTPR₁₆ protein alone (lane 1), AuNPs (lane 2), CTPR₁₆ and AuNPs mixed together but not conjugated (lane3) and CTPR₁₆-AuNPs conjugate (lane 4). The gel is stained using Coomassie blue staining solution. B) Absorption spectra of the CTPR₁₆ protein (dotted line), AuNPs (dashed line) and the CTPR₁₆-AuNPs conjugate (solid line). C) Size exclusion chromatogram of the CTPR₁₆ protein (dotted line), AuNPs (dashed line) and the CTPR₁₆-AuNPs conjugate (solid line).

After Co-NTA column, the dialyzed elution fractions are analyzed by Fast Protein Liquid chromatography (FPLC) (*Materials and Methods M.2.1.b.8*). As FPLC elutes the entities by their size, the differences observed between the Co-NTA elution fractions and the reference controls (CTPR₁₆-4Cys and AuNPs alone) are consistent with what is expected if the conjugation is performed (Figure 49.C). For the Co-NTA elution fractions, a first peak

is observed at 8 mL before the elution of the free CTPR₁₆ at 9 mL. The reaction medium also contains some free CTPR₁₆-4Cys protein by its characteristic peak at 9 mL. The control of free AuNPs alone have a retention time between 10 and 11 mL but it is not observed that peak in the reaction mixture because the sample is purified before with Co-NTA column and remove all the free AuNPs from the sample. The characterization experiments are performed using the purified CTPR₁₆-AuNPs conjugate using both the Co-NTA column and FPLC separation.

Agarose gel electrophoresis has also been performed in order to characterize CTPR₁₆-AuNPs through its electrical mobility in a 0.5% agarose gel in TBE 0.5X buffer (Tris 50mM pH=8.3, Boric Acid 50mM and 1mM EDTA,) as explained (*Materials and Methods M.2.1.b.9*). The coomassie stained electrophoresis gel shows clear difference between AuNPs alone, the purified CTPR₁₆-AuNPs conjugate and the control mixture of unmodified AuNPs and protein CTPR₁₆-4Cys (Figure 49.A). The electrophoretic gel reveals that the CTPR₁₆-AuNPs conjugate runs faster than the AuNPs alone marking the influence of the presence of CTPR₁₆-4Cys with the AuNPs.

In order to go deeper in the characterization of the CTPR₁₆-AuNPs conjugate tending to the nanoscale resolution, two complementary microscopy techniques are employed, high resolution Transmission Electron Microscopy (HR-TEM) and room temperature Scanning Tunnelling Microscopy (STM) (*Materials and Methods M.2.1.10*). Figure 50.A shows the comparison of a HR-TEM image of a 1 nM sample of CTPR₁₆-AuNPs with the corresponding image for AuNPs. A statistical study based on the design of the CTPR₁₆-AuNPs conjugate is performed and it shows a higher occurrence of 3 and 4 AuNPs in-lined disposition compared to free AuNPs, for which mostly individual AuNPs are observed (Figure 50.A/bottom). Using similar concentrations, STM images of CTPR₁₆-AuNPs on gold in air and at room temperature are obtained. Using bias voltages between 0.1 V and 1.5 V, only the conjugated AuNPs are visible, indicating that the body of the protein has a very low conductance. While in most of the cases a linear disposition of the nanoparticles, equivalent to that of the HR-TEM images, is observed, in some occasions with very sparse CTPR₁₆-AuNPs, four AuNPs groups disposed in good agreement with the protein geometry of Figure 50.B are observed. Using both TEM and STM it is confirmed that the structure of the CTPR₁₆-AuNPs conjugate is in agreement with the designed model (Figure 47.C).

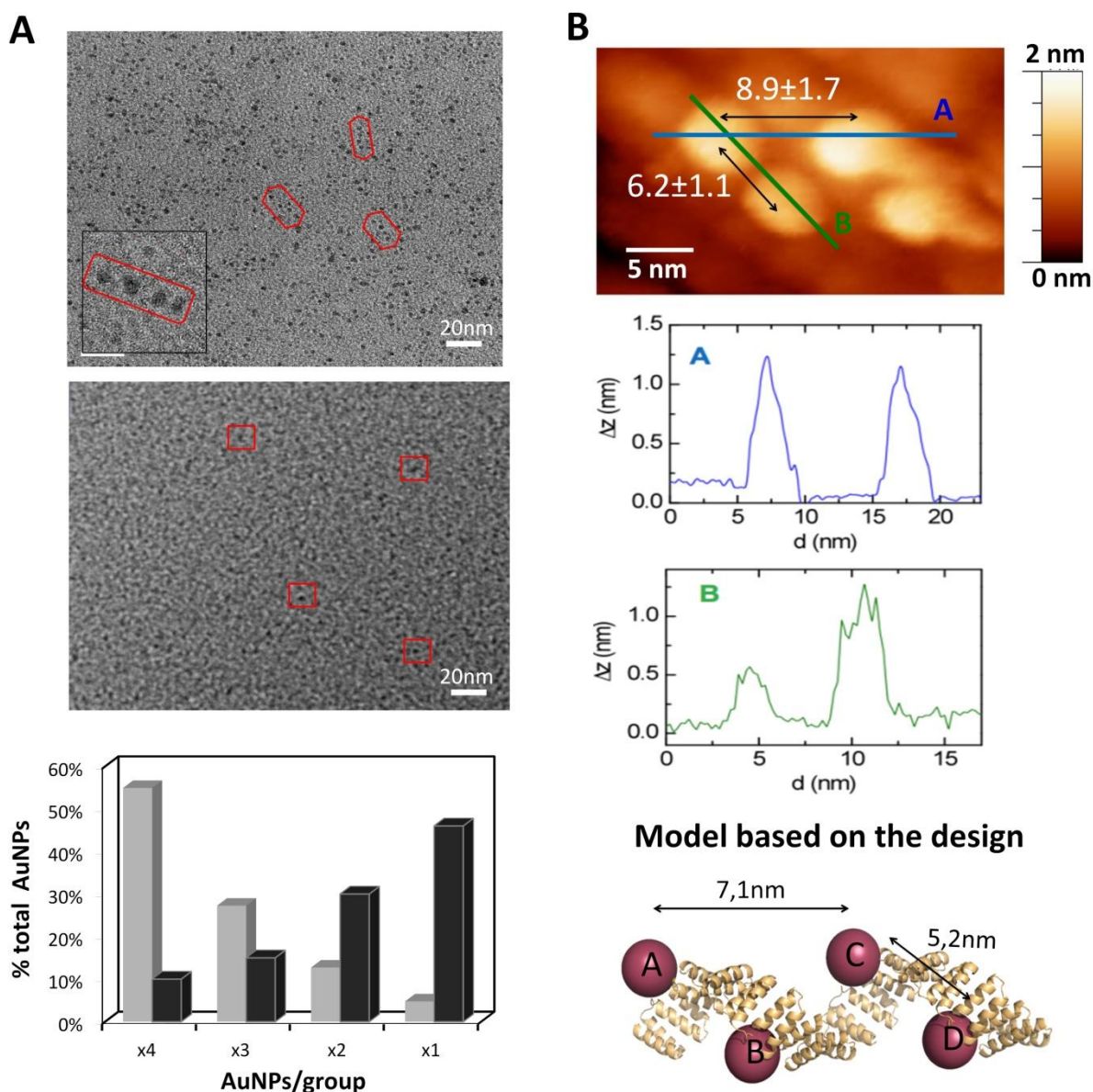


Figure 50. CTPR₁₆-AuNPs conjugate characterization. A) Top panel TEM images of the CTPR₁₆-AuNPs conjugates using JEOL JEM 2100F TEM, with red squares have been marked the group of 4 particles that corresponds to the of CTPR₁₆-AuNPs conjugates. Bottom panel TEM images of particles without protein, as control. Lower panel statistics of the both TEM images: the statistics of the conjugation process counting the AuNPs that are at distances in agreement with the distances between each cysteine residues in the CTPR₁₆-4Cys protein in grey and the statistics of the image of the control in black. B) STM image of CTPR₁₆-AuNPs conjugate with measurements of 2 profiles A (blue) and B (green). The average distances between the particles in different STM images with the error are represented in the image. Lower part shows a ribbon representation of the CTPR₁₆-AuNPs conjugates and distances between the particles based on the designed model.

Next, self-assembly properties of the CTPR₁₆-4Cys protein is analyzed using STM by depositing a 1 μ M protein concentration sample of the CTPR₁₆-AuNPs conjugate. Figure 51.A displays STM images of the CTPR₁₆-AuNPs layer surface, where the AuNPs are clearly arranged in an ordered disposition. From these images, and based on the model structure of the CTPR₁₆-AuNPs conjugates, a probable assembly of the conjugate is proposed as

shown in Figure 51.B, and superposed to the STM image of Figure 51.A. In this configuration, the gold nanoparticles are close together, suggesting that electron transfer between them can occur when applying a potential difference [237]. For this reason, the next step is to prepare a device based on an ordered film of the CTPR₁₆-AuNPs conjugate and study its electrical transport properties.

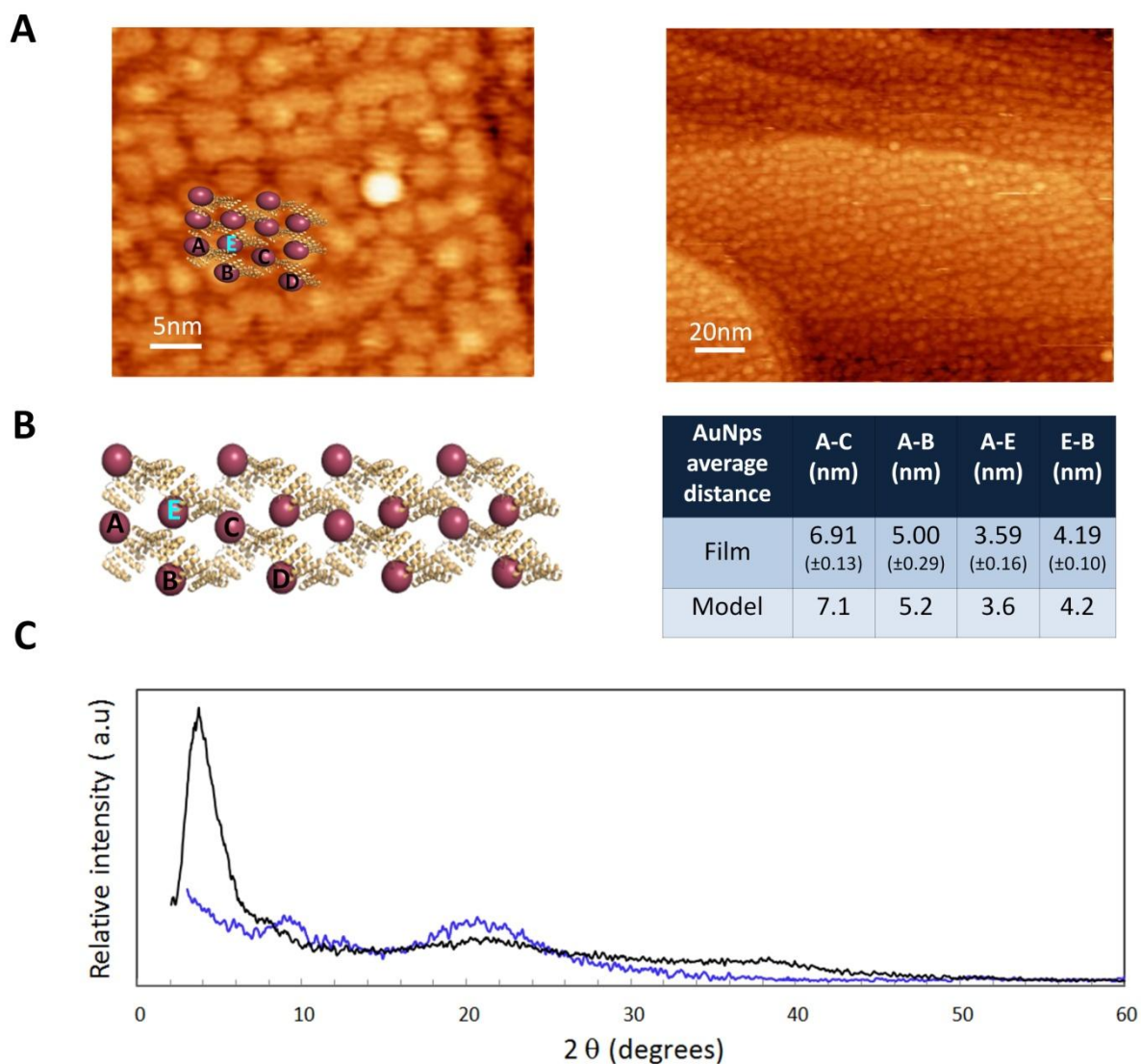


Figure 51. Ordered film formation using CTPR₁₆-AuNps conjugate. A) STM images at different magnifications of the CTPR₁₆-AuNPs conjugates forming a monolayer at 1 μ M CTPR₁₆-AuNPs concentration. B) A model of the arrangement of CTPR₁₆-AuNPs conjugates in the monolayer. In the table, the comparison between the AuNPs distances in the model and in the STM images. C) X-ray Diffraction (XRD) spectrum of CTPR₁₆-AuNPs conjugate film in black and XRD spectrum of the CTPR₁₆ film in blue.

Toward the generation of a conductive device, a solid film is formed with CTPR₁₆-4Cys protein and CTPR₁₆-AuNPs conjugate (*Materials and Methods M.2.1.b.12*). In film

composed of CTPR₁₆-4Cys protein the expected packing for the proteins is the one described by Grove et al. [103]. To explore the packing in the films X-ray diffraction (XRD) is performed (*Materials and Methods M.1.b.13*). XRD pattern of CTPR₁₆-4Cys (Figure 51.C) shows a peak at $2\theta = 9.63^\circ$ which is accompanied by a number of reflections satisfying a reciprocal spacing ratio of another peak a $2\theta = 20.01^\circ$. This value with a d -spacing of around 0.43 nm has been previously associated to the α -helical pitch of CTPR proteins [238].

In the film composed of the CTPR₁₆-AuNPs conjugate the expected packing is the one observed in the AuNPs layer on gold using the STM (Figure 51.A.B) since the interactions that leads the film formation is expected to be the same. The XRD pattern for the CTPR₁₆-AuNPs film shows a peak at $2\theta = 37.68^\circ$ is observed, that can be attributed to the crystal structure of the AuNPs [239]; another peak at $2\theta = 20.14^\circ$, that corresponds to the organized CTPR proteins in the film, since also appears in the film formed by CTPR₁₆-4Cys protein. Finally, there is another peak at $2\theta = 3.663^\circ$. This peak is more intense and can be attributed to the organization of the gold nanoparticles in the film, since the repetition pattern of the gold nanoparticles is more frequent than the repetition pattern of the proteins, as observed in the disposition of AuNPs in the CTPR₁₆-AuNPs monolayer (Figure 51.A). These results confirms that the particles in the film are close enough to result in an efficient electron transfer between them in the solid film [237].

In order to study the electrical transport properties of the films, CTPR₁₆ protein and CTPR₁₆-AuNPs conjugate is deposited over Si/SiO₂ wafers with interdigitated gold electrodes on top (*Materials and Methods M.1.3.b.14*). These electrodes describe a channel with a ratio width W to length L (W/L) of approximately 600, which minimizes the effect of edge currents between the electrodes. The current versus voltage (IV) curves is recorded in the interval (-1 V, +1V) for more than four different films of both the CTPR₁₆ protein and the CTPR₁₆-AuNPs conjugate (*Materials and Methods M.1.3.b.15*). Figure 52 shows the high reproducibility of the films transport properties, when scaling the results by the film thickness (t). To calculate the thickness of the films, a scratch is done in the film to the substrate and the difference between the surface and the top of the film (thickness of the film) it is measured using AFM as explained *Materials and Methods M.1.3.b.16*. From the IV curves, the conductance of the film is obtained using Equation (10):

$$I = \frac{GV}{t} \quad (10)$$

where V is the applied voltage in volts (V), G is the conductance of the film is Siemens (Ω^{-1}) and I is the measured intensity in amps (A), and t is the thickness of the film (m). The measured IV curves, corrected with the film thickness, are shown in Figure 52. From these

curves averaged conductivity values of the CTPR₁₆ protein and CTPR₁₆-AuNPs conjugate films are obtained using Equation (11):

$$\sigma = G \frac{L}{W \times t} \quad (11)$$

Where σ is the conductivity of the film ($\Omega^{-1}\text{m}^{-1}$), G is the conductance of the film in Siemens (Ω^{-1}), L is the length of the film between the electrodes (m), W is the width of the films between the electrodes (m) and t is the thickness of the film (m).

The average conductivity of CTPR₁₆-4Cys protein film is $\sigma = 1.38 \pm 0.19 \times 10^{-7} \Omega^{-1}\text{m}^{-1}$ and of CTPR₁₆-AuNPs film is $\sigma = 1.37 \pm 0.35 \times 10^{-3} \Omega^{-1}\text{m}^{-1}$, when the conductivity is calculated from the IV curve slope in the interval (-0.3 V, 0.3 V) from the linear region of the IV curves. The comparison between the conductivity values obtained for CTPR₁₆-4Cys protein film and CTPR₁₆-AuNPs films shows that the conjugate has a conductivity approximately 4 orders of magnitude larger than the protein. These results confirm that there is indeed a good electron transfer among the AuNPs in the films, showing the improved properties of the CTPR₁₆-AuNPs conjugate.

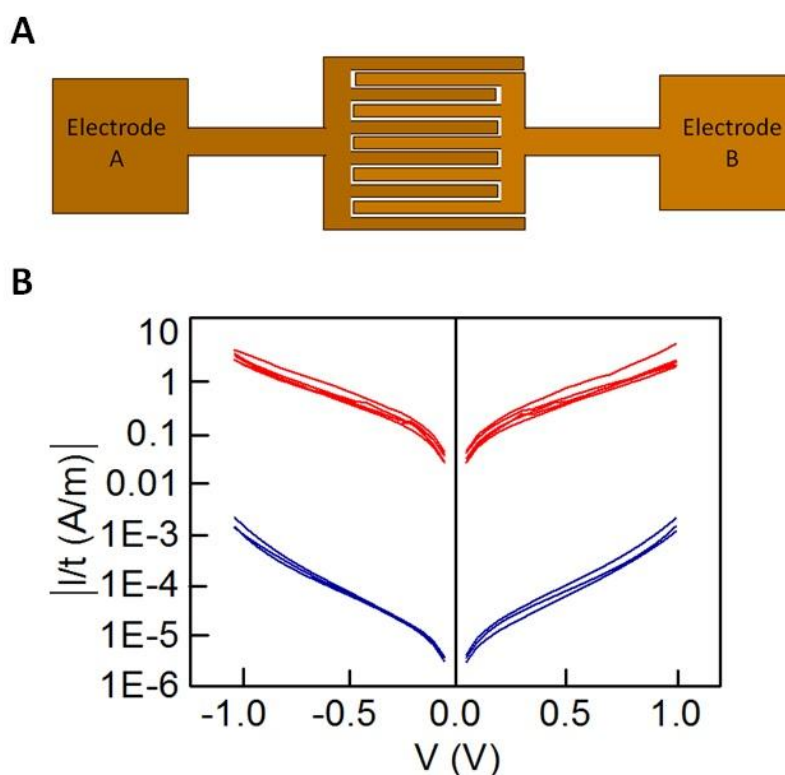


Figure 52. Electron transfer measurements in the films. A) Schematic representation of the electrodes on which the electron transfer measurement are performed. Electrode A and B are made by gold deposited by lithography in a silicon (covered by SiO₂) surface. The width of the electrodes is 20 μm and the length is 11800 μm . B) I/V curves obtained by electron transfer measurement of the CTPR₁₆-AuNPs film in red and CTPR₁₆ films in blue. In X axis it is represented the voltage variation and in Y axis the intensity corrected to the film thickness and represented at logarithmic scale.

In conclusion, in this section we show that we can control the conjugation of the AuNPs into the CTPR protein with atomic resolution. We characterize the conjugation process using high resolution techniques, first to determine the efficiency of the conjugation by TEM and then to define the agreement between the CTPR₁₆-AuNPs conjugate that we obtain experimentally and the designed model by STM. Taking advantage of the self-assembly properties of CTPR₁₆-4Cys protein we form a monolayer in which the AuNPs are organized by the proteins. We use the same assembly properties to generate a CTPR₁₆-AuNPs order film, in which the disposition of the AuNPs in the film is controlled.

An application of the biomolecular templating of gold particles is the generation of efficient electron-transfer devices. Thus, we explore the use of the CTPR proteins to form solid conductive films. We perform current measurement in the solid films and obtain efficient electron transfer between the AuNPs.

The results presented in this section show the potential use CTPRs as designed scaffolds for patterning AuNPs forming a solid film in which the AuNPs are homogeneously distributed. The control in the distribution of the elements in the films opens the door to use the CTPR proteins as templates to organize not only AuNPs, but other elements with different properties that are useful in nanotechnology.

2.4.1.3. Conclusions

In this section, we explore the use of CTPR proteins as a template to organized AuNPs. We demonstrate that we can control the conjugation between AuNPs in the CTPR protein with atomic resolution using different strategies. We rationally introduce modifications in the CTPR proteins to conjugate, in a specific way, the AuNPs. Indeed, we use the control in the CTPR self-assembly to use them as templates to form structures in which the AuNPs are organized at two different conformations: in a monolayer and in a solid film. Moreover, the generated functional structures have properties that depend on the organization of the AuNPs. These studies establish the basis to use CTPR protein to form different structures with AuNPs are organized at different length scales: nano, micro and macroscale. Furthermore, it is possible to expand this knowledge to use the CTPR proteins to organize different metallic particles using protocols similar to the ones developed in this section.

2.4.2. Generation of electron donor-acceptor systems

2.4.2.1. Motivation and objectives

In this section, we explored the use CTPR scaffolds for organized electron-donor and acceptor molecules. One of the main challenges in current science is the construction of highly ordered materials bearing photonic and/or electronic active units [240, 241]. Nowadays, many bottom-up approaches have been undertaken to control the organization and morphology of electro- and photoactive components at the nanometer scale, enabling the formation of supramolecular architectures that typically result in a significant improvement of their properties [242-244]. Nonetheless, the majority of these methodologies lack the selective positioning and orientation of the different functional components, and the control at different length scales is still challenging. In addition, poor monodispersity of the aggregates is another drawback. For instance, average size of the assemblies can be only partially controlled by concentration, temperature or nature of the solvent using the basic principles of supramolecular self-assembly, from an entirely chemical point of view.

However, templating these components with bioinspired scaffolds can provide the control over the location, monodispersity and chirality needed for technological purposes. Promising examples show the use of biomolecules for ordering organic materials. DNA is an especially appealing template because of its molecular recognition abilities enabling the formation of programmed nanostructures with geometrical, size and positional control and have been use to organize photo ad electroactive molecules successfully [245, 246].

Peptides [247-249], peptoids [250] and virus capsids [251, 252] have also been used for selectively controlling the position of a great variety of chromophores. For example, self-assembling peptide nanotubes [253] have been used to arrange photo- and electroactive molecules such as fullerenes. Similarly, helical peptides and polymer-peptide have been applied to pattern semiconductors and organic chromophores into ordered structures [254, 255].

These works represent the proof of concept that bioinspired assembly can be successfully used to generate more efficient organo-electronic devices. However, an important limitation of these approaches is how to transfer hierarchically this organization to the different length scales: nanoscale, microscale and macroscale. The combination of both the generation of bio-organic conjugates with photonic and/or electronic properties and the macroscopic organization of these building blocks could lead to the construction of functional bulk biomaterials with applications in fields such as organic electronics and photovoltaics.

For this end, we take advantages of the structural understanding of the CTPR proteins to accurately design the atomic-level the positions in which the molecules are coordinated, organizing the molecules of interest at nanoscale, that is essential for generate ordered materials. Moreover, we exploit the capability of CTPR scaffolds to form nanostructures such as ordered thin films, to ultimately employ them as a framework to fabricate well-ordered hybrid bio-organic materials. Thus, we are able to form a bio-hybrid systems, where the electron-donor or acceptor molecules are organized at different length scales.

2.4.2.2 Results and discussion

2.4.2.2.a Organizing porphyrins in solution and solid state

In the first section, we evaluate a mutated CTPR protein as scaffold of photoactive porphyrin arrays both in solution and in the solid state through the conjugation of mutated residues of the protein with designed porphyrins.

Porphyrins represent an important class of photo- and electroactive molecules. However, in order to get electron transfer between them they have to be interacting bi- π interactions. These interactions are distance dependent. Thus, to get π - π interactions between the porphyrins the distance between them have to be 0.45 and 0.7 nm [256]. A big research effort have been done to form systems where the porphyrins are supramolecularly organized using a variety of biomolecules. Thus, as representative examples, nucleic acids [257, 258], peptides [259-261] and the tobacco mosaic virus [262] have been used as scaffolds for precisely controlling the position and orientation of porphyrins.

Despite of the advances in the field, an important limitation of these approaches is how to transfer hierarchically this organization to the macroscopic level in the solid state, which is essential to the construction of efficient devices with applications organic electronics and photovoltaics. In this section, we use a designed CTPR protein as scaffold for assembling photoactive porphyrin rings in solution and solid state. The characterization in solution of the bio-organic conjugates obtained is fully detailed. In the next stage, solid thin films of these conjugates are prepared and studied in terms of organization and orientation. Finally, the anisotropic photoconductive properties of these ordered solid materials are elucidated.

It is designed a modified CTPR protein as scaffold for assembling a specific number of porphyrin chromophores in a defined distance and orientation. Based on the crystal structure of the CTPR protein, different amino acid positions are selected to arrange the chromophores along the CTPR superhelix that allow for efficient π - π interactions

between the molecules. The selected positions are not conserved between CTPR modules, thus it is expected that their modification will not affect the structure of the protein template.

Specially, two cysteine residues are introduced in each repeated unit to provide unique reactivity for the immobilization of the porphyrin derivatives. A CTPR protein with four identical repeats is generated, CTPR4_CP, resulting in a protein with eight quasi-equally spaced cysteine residues in four loops of the protein (Figure 53.A.B) (*Materials and Methods M.2.2.a.1*). The distance between the Cys side chains is between 7 and 8 Å, in the order of the distance required to establish π - π interactions between the porphyrin rings. As expected, the mutations don't significantly affect the helical structure of the protein scaffold (Figure 53.B).

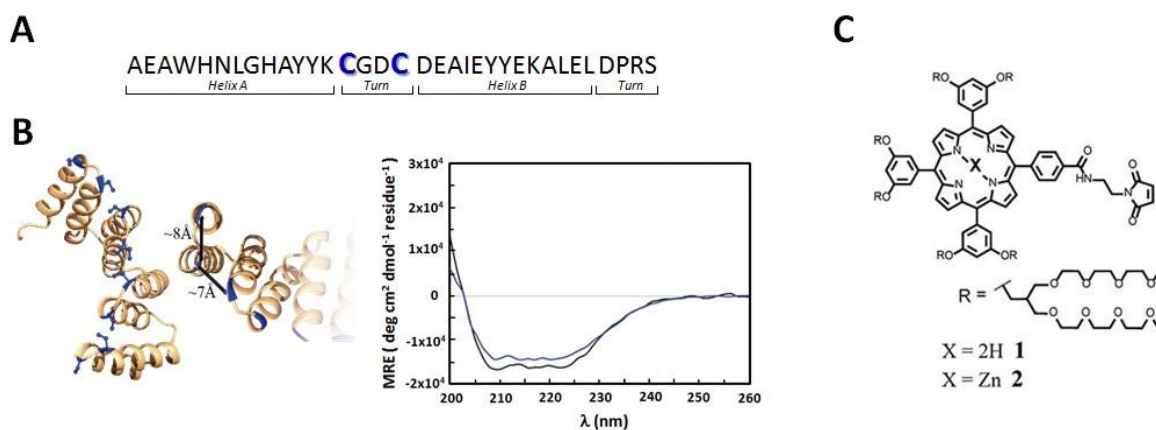


Figure 53. Mutated CTPR4 protein and designed porphyrin. A) Sequence of amino acids of the repeat module, highlighting the positions 14 and 17 that have been mutated to Cysteine. B) Left, ribbon representation of mutated CTPR4_CP protein structural model based on the structure of CTPR8 (PDB ID: 2HYZ). The mutated cysteine residues for the porphyrin attachment are highlighted in blue. Right, CD spectra of mutated CTPR4_CP comparing with the original CTPR4. C) Molecular structure of porphyrin derivatives **1** and **2**.

Besides, in the design of the porphyrin moiety two important requirements should be considered: (i) porphyrin molecules must exhibit high solubility in water but avoiding charged functional groups, since undesired interactions with charged side-chains of the amino acids in the protein should be eliminated and, (ii) an efficient cross-linker must be placed in the porphyrin structure in order to promote quantitative or quasiquantitative conjugation with the eight cysteine residues of the designed protein. With this in mind, two porphyrin derivatives are synthesized, free-base and zinc-metalloporphyrin (**1** and **2**, respectively), decorated with twelve triethylene glycol watersoluble tails and a maleimide reactive group as efficient crosslinker for the conjugation reaction with the cysteine residues in CTPR protein (Figure 53.C). The synthesis of the porphyrin is performed by Dr. Nazario Martín's group (Universidad Complutense de Madrid).

The designed CTPR4-CP and porphyrin moieties (**1** or **2**) are conjugated using the maleimide-cysteine chemistry, as follows: 300 μL of 50 μM of freshly reduced protein is mixed with 40 equivalents of **1** or **2** giving a ratio of 1:5 cysteine:maleimide and shaken gently. After one hour, an extra of 20 equivalents of **1** or **2** are added to the mixture giving a final ratio 1:7.5 cysteine:maleimide (*Materials and Methods M.2.2.a.2*). The conjugation reaction is displayed in (Figure 54).

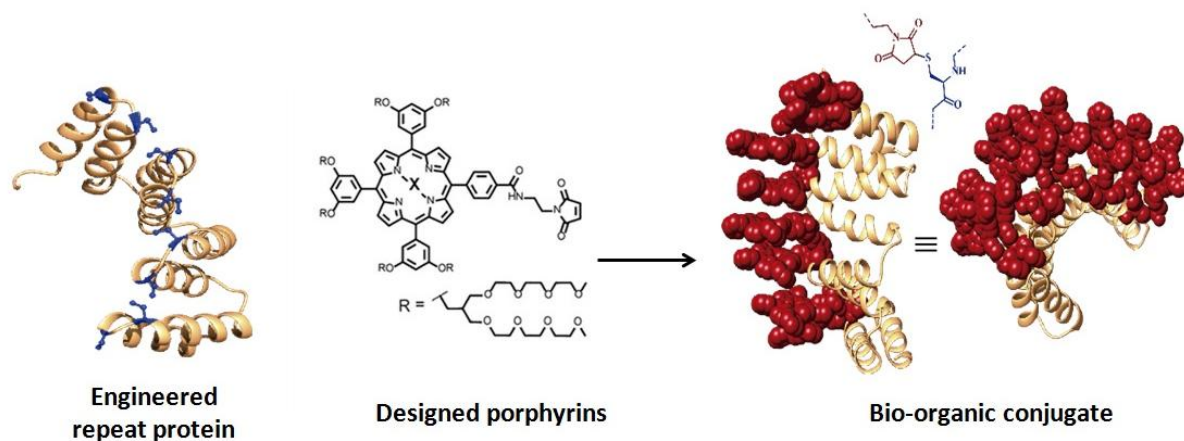


Figure 55. Conjugation reaction to obtain conjugates CTPR4_CP-1 and CTPR4_CP-2. Schematic representation of the maleimide-cysteine reaction between mutated CTPR4_CP and porphyrin **1** or **2**. The conjugate is shown from different perspectives.

At different times during the conjugation reaction, the analysis of the samples by gel electrophoresis shows that a conjugate of higher molecular weight than CTPR4_CP protein (MW = 19163 Da) and porphyrins (MW₁ = 3160 Da, MW₂ = 3223 Da) is obtained even at the shortest reaction time (Figure 55.A). When the gel is imaged without staining, the fluorescence signal of the porphyrin can be detected showing a new band between 31 KDa and 45 KDa, corresponding to the molecular weight expected for CTPR4_CP-1 or CTPR4_CP-2 conjugates, containing from four to eight porphyrin moieties. Moreover, the staining of the protein with Coomassie Blue confirms that the higher molecular weight band is composed of both protein and porphyrin (*Materials and Methods M.2.2.a.2*).

To quantify the number of porphyrin molecules bound per protein in the conjugation reaction, mass spectrometry is used (Figure 55.B) (*Materials and Methods M.2.2.a.3*). After three-hour reaction time, there is not signal corresponding to the CTPR4_CP protein and the most abundant peak appeared at 41 kDa in the case of conjugated CTPR4_CP-1, which corresponds to the CTPR4_CP protein covalently linked to six to eight porphyrin units.

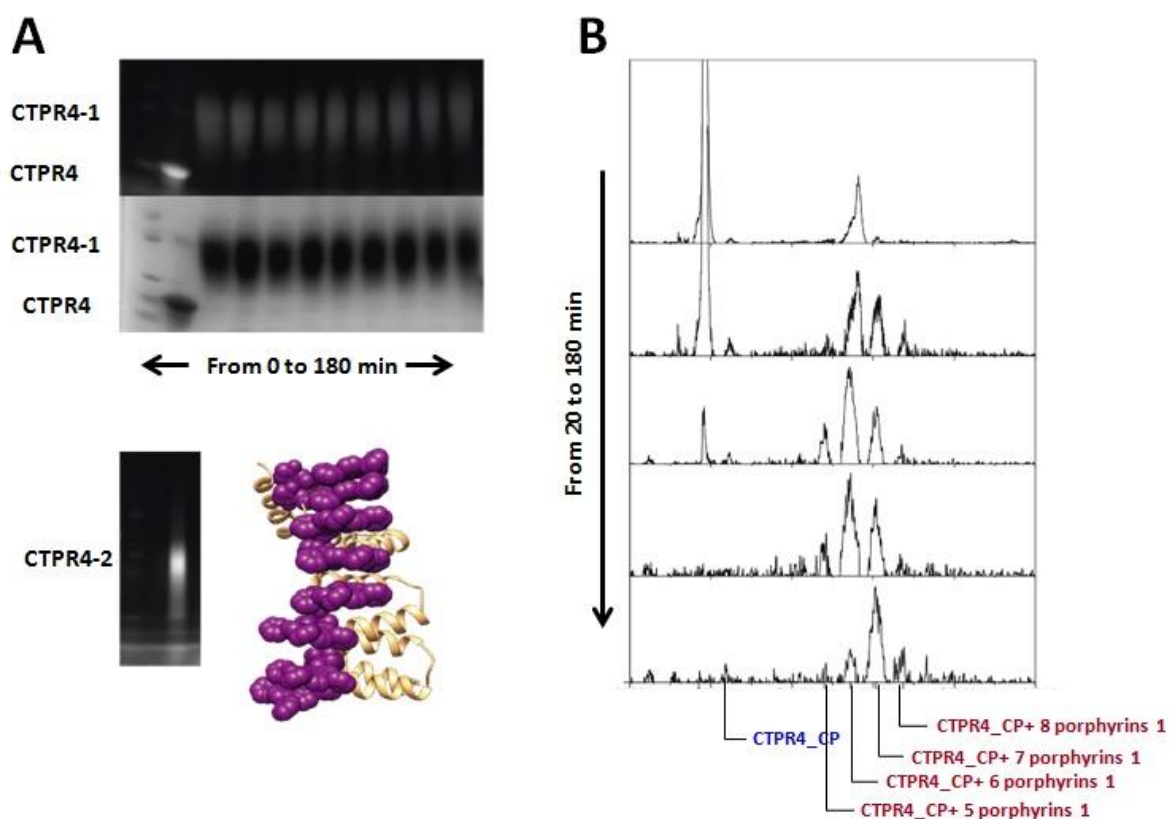


Figure 55. Monitoring experiments for the conjugation reaction. A) Top: SDS-Page electrophoresis gel of the CTPR4_CP-1 conjugation at different reaction times (20, 40, 60, 80, 100, 120, 140, 160 and 180 minutes, lanes 3 to 11). Molecular weight marker (lane 1) and CTPR4 control protein with Amresco EZ-vision fluorescent loading buffer (lane 2). The gel is shown prior staining imaged using UV-light to monitor the fluorescence of the porphyrins (upper panel) and after Coomassie Blue staining (lower panel). Bottom: SDS-Page electrophoresis gel of the CTPR4_CP-2 conjugate. Molecular weight marker with Amresco EZ-vision fluorescent loading buffer (lane 1) and CTPR4_CP-2 conjugate (lane 2). B) MALDI-TOF spectra of the conjugation reaction in CTPR4_CP-1 at different reaction times.

The purification of the protein-porphyrin conjugates from the excess of free porphyrin is an essential step for further use and characterize homogeneous hybrid structures. CTPR4_CP-1 and CTPR4_CP-2 purification is successfully carried out using size exclusion chromatography (Figure 56) (*Materials and Methods M.2.2.a.4*). All the characterization experiments are performed using the purified conjugates.

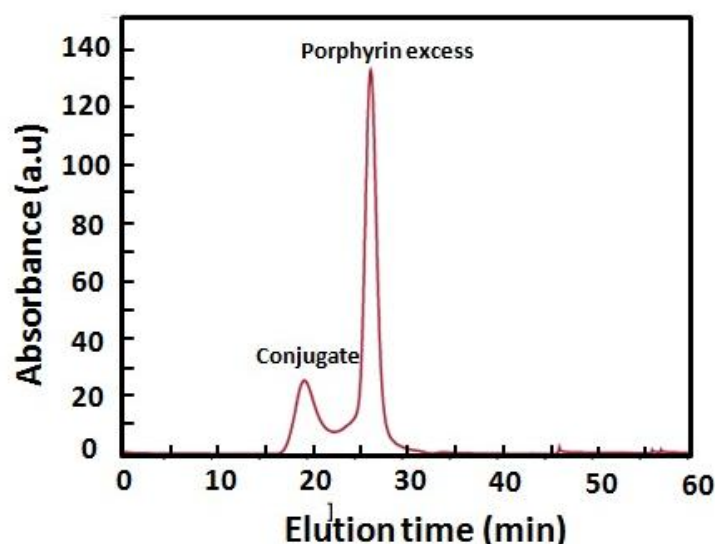


Figure 56. Size exclusion chromatogram of the purification process of the conjugate for removing the excess of porphyrin.

As has previously been mentioned, the distinctive properties of the CTPR scaffold allowed for the design of a CTPR4_CP protein with an appropriate distance between the reactive moieties that, considering the rotational freedom of both the protein side chains and the linked porphyrins, will enable an efficient π - π intermolecular interaction between the porphyrin rings.

It is well established that porphyrins have the ability to self-aggregate noncovalently to form H-aggregates (face-to-face) or J-aggregates (side-to-side), these states are characterized by a shift on the Soret absorption band towards the blue or red, respectively, compared to the monomeric absorption band [263-267].

Thus, the UV-vis spectra of both porphyrins, **1** and **2**, as references and CTPR4_CP-1 and CTPR4_CP-2 conjugates are recorded in order to obtain more information about the conformation of the porphyrins in the conjugates (*Materials and Methods M.2.2.a.5*).

The absorption spectrum of porphyrin **1** in a PBS buffer solution presents an intense Soret band at 406 nm and four Q-bands in the region between 500 and 700 nm. Meanwhile, porphyrin **2** presents the Soret band at 425 nm and two Q-bands at 555 and 595 nm. In contrast, the UV-vis spectra for CTPR4_CP-2 and CTPR4_CP-1 conjugates show a red shift of 5 and 17 nm in the Soret band of the porphyrin, respectively, compared with the corresponding free porphyrins (Figure 57.A.B). These shifts suggest the formation of a J-type aggregate of porphyrins in both conjugates, certainly induced by the geometry of the protein and the position of the cysteine residues in the framework.

To further confirm the π - π interaction between porphyrin moieties within the conjugates, UV-vis spectra are measured at different H₂O : MeOH ratio leading to an increased intensity and a blue-shift of the Soret band in while the percentage of methanol increases (Figure 57.C.D) (*Materials and Methods M.2.2.a.5*). This fact is in agreement with the disruption of the π - π interaction when protein denaturation is achieved with an organic polar solvent.

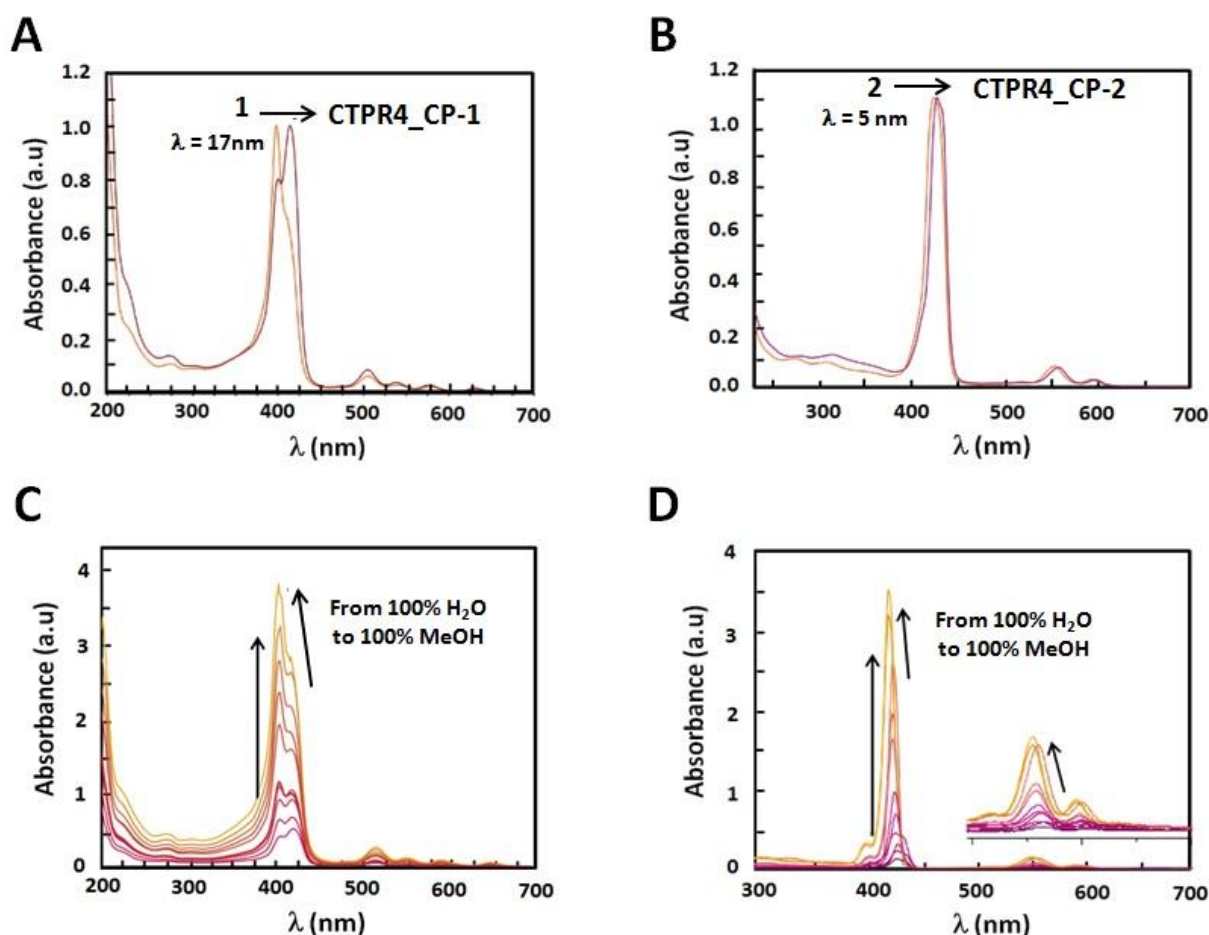


Figure 57. UV-vis spectra of conjugates. A) Normalized UV-vis spectrum of CTPR4_CP-1 conjugate, compared to UV-vis spectra of **1** in PBS buffer. B) Normalized UV-vis spectrum of CTPR4_CP-2 conjugate compared to UV-vis spectra of **2** in PBS buffer C) UV-vis spectra of CTPR4_CP-1 at different methanol concentrations. Arrows show the increase of the methanol concentration from 0 to 100%. D) UV-vis spectra of CTPR4_CP-2 at different methanol concentrations. Arrows show the increase of the methanol concentration from 0 to 100%.

Circular dichroism (CD) studies in solution are performed on the CTPR4_CP-1 and CTPR4_CP-2 conjugates (Figure 58) (*Materials and Methods M.2.2.a.6*). In the first case, the CTPR4_CP-1 conjugate shows a bisignate signal with a negative Cotton effect at 417 nm, a positive Cotton effect at 428 nm and a zero cross point at 423 nm. The presence of these Cotton effects, that is, exciton-coupled circular dichroism, is indicative of a close

proximity between the chromophores, as previously reported (Figure 58.A) [268-270]. In the second case, the CTPR4_CP-2 conjugate shows a strong trisignate signal with a negative Cotton effect at 425 nm, positive Cotton effect at 434 nm and another negative Cotton effect at 446 nm with zero cross points at 429 and 442 nm is shown for the CTPR4_CP-2 conjugate (Figure 58.A).

Multisignates in the Soret band have previously been described in the literature to strong π - π stacking interactions between close porphyrins [268, 271]. Thus, these dichroic signals in the porphyrin absorption region demonstrate the transfer of chirality from the CTPR_CP protein scaffold to the porphyrin molecules, since free porphyrins **1** and **2** lack optical activity in their CD spectra (Figure 58.A). Furthermore, the feature CD signal for alpha-helical secondary structure of the CTPR4_CP protein is totally retained in both CTPR4_CP-1 and CTPR4_CP-2 conjugates in PBS buffer (Figure 58.B). This result confirms the successful ability of CTPR_CP proteins to be used as a robust scaffold for ordering organic chromophores, preserving its secondary and tertiary structure even when six to eight porphyrin molecules per protein are incorporated. It is important to note that the signal molecular weight of both conjugates represent more than twice the initial one corresponding to the protein alone; however, the structural integrity of the biological framework is remarkably conserved. Moreover, a denaturalization is also carried out for both conjugates through CD measurements. No appreciable dichroic signal in the range of Soret band of the porphyrins is observed after adding methanol, which confirms the rupture of π - π interactions when the protein scaffold is unstructured (Figure 58.C.D).

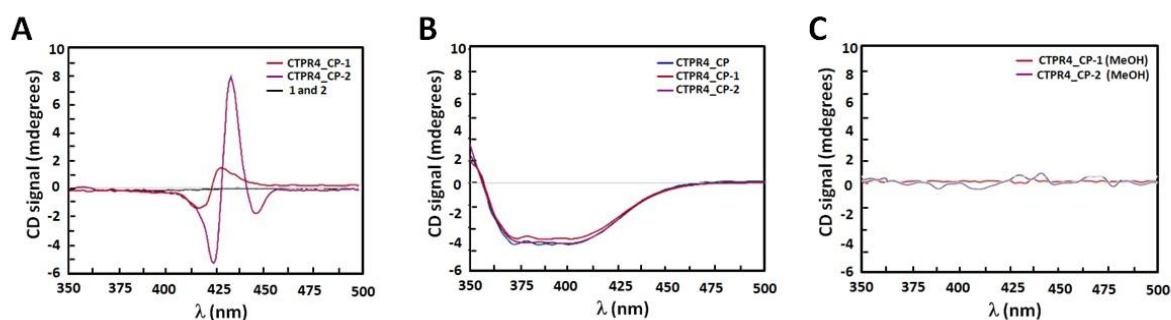


Figure 58. Circular dichroism of the conjugates. A) CD spectra of porphyrin **1** and **2**, in black, CTPR4-1, in red, and CTPR4-2, purple, conjugates (1.70 μ M) in PBS buffer in the spectral region of the porphyrin absorption. B) CD spectra of CTPR4_CP, in blue, CTPR4_CP-1, in red, and CTPR4_CP-2, in purple, conjugates in PBS buffer in the spectral region of the protein's secondary structure absorption. C) CD spectra of CTPR4_CP-1, in red, and CTPR4_CP-2, in purple, conjugates in methanol in the spectral region of the porphyrin absorption.

The intrinsic capability of CTPR proteins to assemble into highly ordered thin films makes them good candidates to organize porphyrin arrays in the solid state [105]. However, the formation of protein-based solid ordered materials has been achieved in

only a very few protein systems. Furthermore, the combination with the building-block tunability is unique of the CTPR scaffold. In this sense, ordered thin films with both conjugates are effectively generated under similar experimental conditions to those described for CTPR proteins alone (Figure 59.A.B) (*Materials and Methods M.2.2.a.7*).

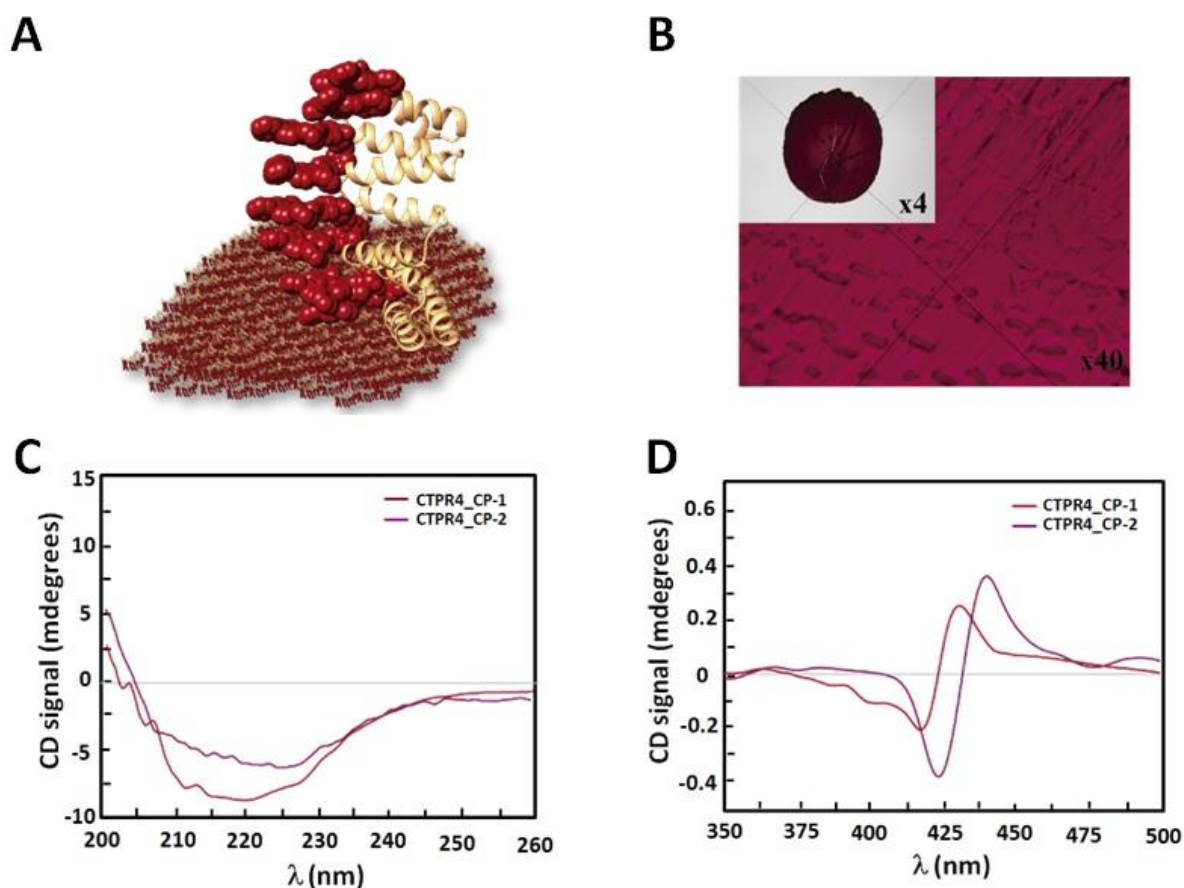


Figure 59. Characterization of CTPR4_CP-1 and CTPR4_CP-2 conjugates in the solid state. A) Schematic representation of the oriented film obtained from the porphyrin-protein conjugate. B) Photograph of a thin film obtained after evaporation of a 5% w/v CTPR4_CP-2 in 10 mM sodium chloride, 10 mM sodium phosphate buffer (pH = 7.0) on Teflon tape. C) CD spectra of the CTPR4_CP-1, in red, and CTPR4_CP-2, in purple, conjugates in solid thin films in the spectral region of the protein's secondary structure absorption. D) CD spectra of the CTPR4_CP-1 and CTPR4_CP-2 conjugates in solid thin films in the spectral region of the porphyrin absorption.

Importantly, the structural integrity of the protein block and the porphyrin arrangement is tested by CD experiments (Figure 59.C.D) (*Materials and Methods M.2.2.a.6*). Figure 59.C.D shows how the secondary structure of CTPR4_CP scaffold remained alpha helical with no evidence of any significant structure other than alpha helix, and how the chiral environment of the porphyrins is maintained in both conjugates. Moreover, as has previously been reported, CTPR proteins are macroscopically aligned in the solid film form [104, 105], thus it is expected that our conjugates would also be aligned. To shed light on this fact, fluorescence anisotropy measurements and X-ray

diffraction (XRD) are carried out. On the one hand, Figure 60.A shows the change in the fluorescence intensity of the porphyrin moieties when the emission polarizer is rotated from 0 to 360° (*Materials and Methods M.2.2.a.8*). The signal shows clear maximum and minimum values, which indicates the anisotropy of the sample. The data can be well fitted to a sine wave function with maximum to minimum peak distance of 107°, in agreement with the alignment phase determined for the CTPR films. This result indicates that the macroscopic ordered pattern of the film is indeed imposed on the porphyrins that are, otherwise, isotropic.

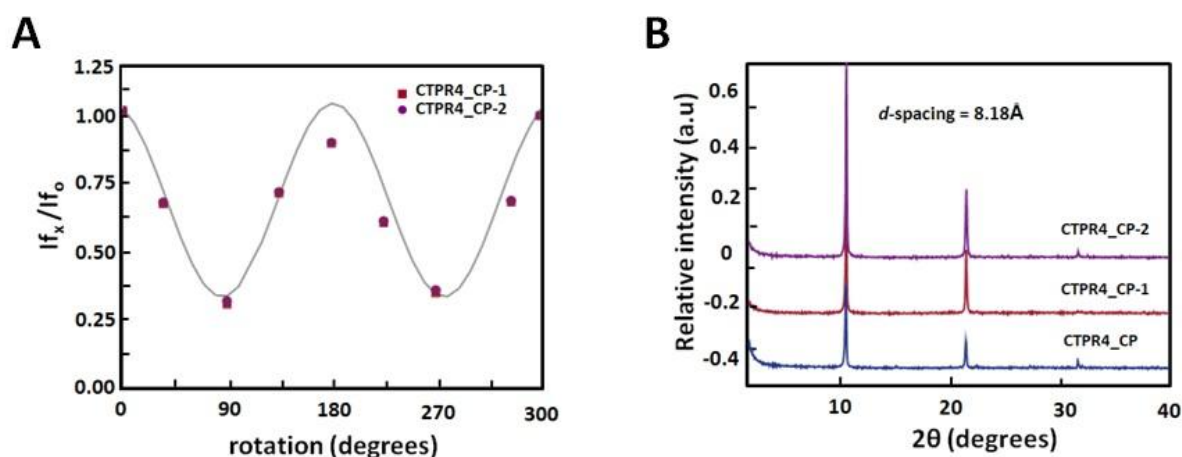


Figure 60. Characterization of CTPR-CP-1 and CTPR_CP-2 conjugates in solid state. A) Fluorescence anisotropy of the CTPR4_CP-1 and CTPR4_CP-2 films. The fluorescence intensity of the porphyrin is plotted as function of the angle of the emission polarizer when the excitation polarizer is fixed at 0°. The intensity data are shown as filled squares, the best fit of the data to a sine wave function ($y = \text{baseline} + \text{amplitude} \sin(\text{frequency} \times X + \text{phaseshift})$) is shown as a solid line. B) X-ray diffractogram of CTPR4_CP, in blue, CTPR4_CP-1, in red, and CTPR4_CP-2, in purple, thin films.

On the other hand, a deep understanding on the organization of the films is reached by XRD experiments (*Materials and Methods M.2.2.a.9*). The XRD pattern of a film obtained from CTPR4_CP protein shows a set of three clear and intense sharp peaks at $2\theta = 10.82$, 21.64 and 32.05° that can correspond to a lamellar packing with a periodical d -spacing of 8.18 \AA (figure 60.B). A similar d -spacing has been previously observed in other films based on repeat proteins and α -helical coiled-coil proteins, being related to the meridional spacing for an α -helix structure, dependent on the angle of inclination [24, 104, 272]. Overall, these data demonstrate the directional orientation of the protein on a surface when a film is formed. Moreover, it is remarkable that the diffraction patterns observed for the films based on the conjugates CTPR4_CP-1 and CTPR4_CP-2 present exactly the same reflections peaks as CTPR4_CP. These experimental findings corroborate that the long-range order of the protein is preserved even with these number of porphyrins introduced in its structure. This structural feature highlights the robustness of this

biological scaffold to be used as an efficient template for ordering organic chromophores not only in solution but in the technologically relevant solid state.

Organized porphyrin arrays on the protein scaffold are interesting as charge carrier transporting and photoconductive motifs. The photoconductivity along the arrays in CTPR4_CP-1 and CTPR4_CP-2 conjugate films is examined by using the flash-photolysis time-resolved microwave conductivity (FP-TRMC) technique [273, 274] that enables the device-less, non contact evaluation of transient conductivity upon photo-generated charge carriers with minimized damage of the scaffold structure in collaboration with Shu Seki group as explained in (*Materials and Methods M.2.2.a.10*). In this method, charge carriers are photo-generated upon exposure to 355 nm laser pulses to the films. Then, local-scale motions of the generated charge carriers can be probed by monitoring the reflected microwave (~ 9.1 GHz) power from the film sample set at the microwave cavity.

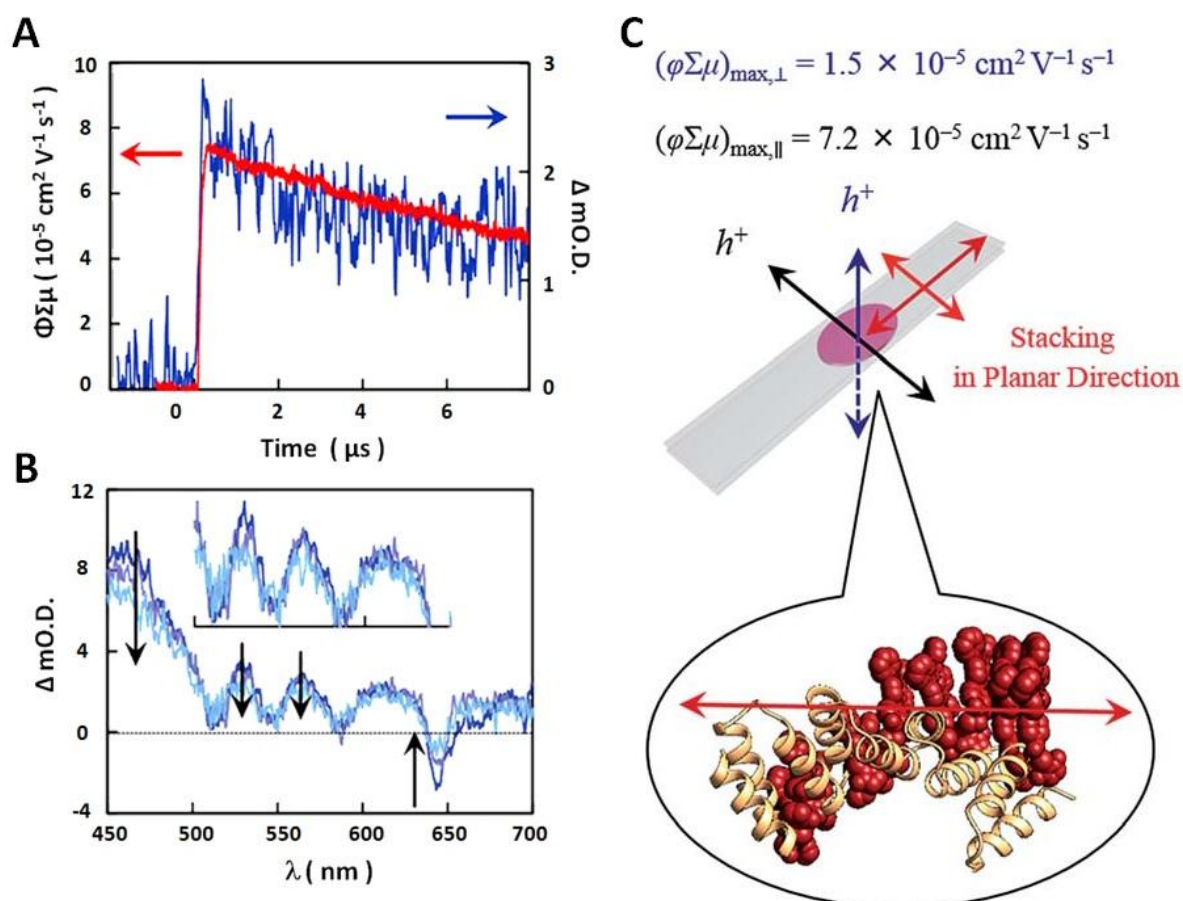


Figure 61. Photoconductivity of CTPR_CP-1 thin films. A) Kinetic traces of transient absorption spectra at 530 nm (blue) and kinetic traces of FP-TRMC (red) of a film of CTPR4_CP-1. B) Snapshot of transient absorption spectra of a film of CTPR4_CP-1 at time range 0.2 (blue), 3 (light purple), and 8 (turquoise) μs after an excitation. C) Schematic illustration of a CTPR4-1 film on quartz and stacking direction of porphyrin arrays together with observed values of conductivity maxima in perpendicular and planar direction to the substrate surface.

Upon laser flash, a drop-cast film of CTPR4_CP-1 shows a conductivity transient with prompt rise and slow decay features (Figure 61.A). The conductivity ($\Phi\Sigma\mu$) indicates the product of charge carrier generation efficiency (Φ) upon photoexcitation and sum of charge carrier mobilities ($\Sigma\mu$: $\mu_h + \mu_e$). Furthermore, the same film yields a clear transient absorption spectra (TAS), where photo-bleach of the neutral and generation of radical cations are observed at the Q-band region of the freebase porphyrins (Figure 61.B) [275]. In fact, the normalized profiles of FP-TRMC and TAS at 530 nm gives almost identical kinetic traces (Figure 61.A), indicating that hole transport is the dominant factor for the observed local-scale photoconductivity of CTPR4_CP-1 under air. By using a typical absorption coefficient of the radical cation of tetraphenyl free-base porphyrin [259], Φ is calculated to be 5.0×10^{-2} , followed by the evaluation of one-dimensional mobility (μ_h) of $1.5 \times 10^{-3} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. Then, through the same measurement processes, μ_h of CTPR4_CP-2 is determined as $1.3 \times 10^{-3} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ (Figure 62).

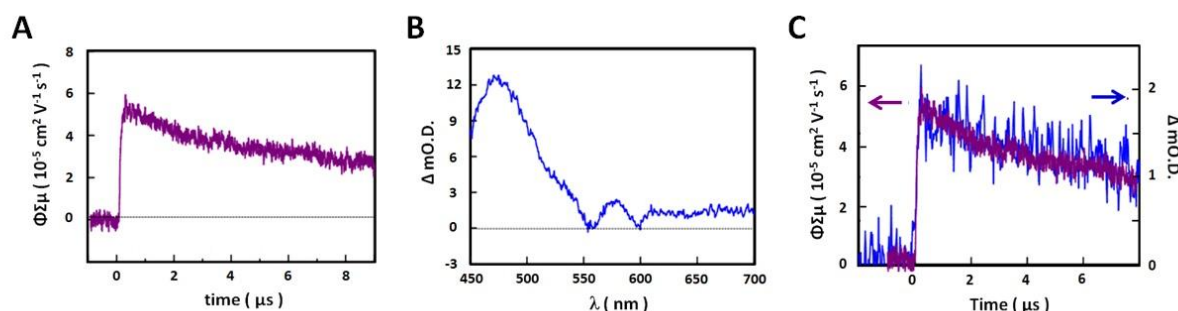


Figure 62. Photoconductivity of CTPR4_CP-2 thin films. A) FP-TRMC kinetic traces of a film of CTPR4_CP-2 at 0–8 μs B) Snapshot of transient absorption spectra of a film of CTPR4_CP-2 at time range 0.2 μs . D) Kinetic traces of transient absorption spectra at 530 nm (blue) and normalized FP-TRMC profile (purple) of a film of CTPR4_CP-2.

Of further interest, it is revealed that structural alignment of the arrays in a macroscopic scale give rise to the anisotropic electrical conductivity. For example, by changing the direction of the CTPR4_CP-1 sample in the cavity [276], $(\Phi\Sigma\mu)_{\text{max}}$ along the perpendicular and parallel directions to the substrate surface are evaluated as 1.5×10^{-5} and $7.2 \times 10^{-5} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, respectively (Figure 61.C). This increment of the electrical conductivity along the parallel direction is in agreement with the anisotropy of the film observed by fluorescence measurements (Figure 60.A).

In conclusion, in this part of the work we describe an innovative approach in which a protein building block is designed to organize porphyrin molecules. In contrast to recent works in which existing biomolecular structures have been used to arrange similar molecules, here we have developed a strategy based on a modular protein unit with tunable properties including stability, function and self-assembly. The scaffold is based on

a simple unit, which allows for a controlled engineering and introduction of reactive functionalities at defined positions for the conjugation of organic molecules. Here, we evidence this potential by precisely organizing porphyrin molecules on the CTPR scaffold at the distance required to achieve efficient π - π interactions between the rings. In the signal conjugates the protein retains its signature helical structure and imposes order and chirality into the porphyrin molecules that show efficient stacking interactions. These results show the potential of this versatile scaffold that could be applied for control grafting of a variety of functional molecules and nanostructural elements.

Another feature that makes the applied system superior to other protein templates is its assembly properties. These observations give the way to their application for the fabrication of solid functional devices. In this work, as a proof of concept, we have shown the formation of ordered films using hybrid protein-porphyrin conjugates. The films obtained display the described nanostructured directional order both in the protein and in the photoactive components. Finally, the photoconductivity of the hybrid thin films shows a remarkable anisotropy in agreement with the directional order of the photoactive molecules. The developed approach is simple and should be easily translatable to other systems that require precise order at different length scales to achieve materials and devices with enhanced properties.

2.4.2.2.b Generation of different sizes CTPR-porphyrin arrays

In this part, we take advantage of the modular structure of the CTPR proteins, to form different sizes porphyrin arrays using CTPR as template. To form this kind of system, it is employed three mutated CTPR proteins with three different lengths (CTPR4_CP, CTPR8_CP and CTPR16_CP) (*Materials and Methods M.2.2.b.1*). The mutated CTPR4_CP described in the previous section is used as the smaller protein scaffold, with a dimension of approximately 40 x 40 Å and eight cysteine residues in the outer area of the superhelix, in particular, in the four loops of the protein, for the conjugation with the maleimide-type porphyrin **2**. The distance between the Cys side chains is between 7 to 8 Å, in the order of the distance required to establish π - π interactions between the porphyrin rings. Meanwhile, two new mutated proteins are designed, that is, CTPR8_CP and CTPR16_CP. CTPR8_CP consisted on eight repeat units comprising one superhelical turn with overall molecular dimensions of approximately 80 x 40 Å and 16 cysteine residues. Finally, CTPR16_CP, with two superhelical turns, a dimension of 160 x 40 Å and 32 cysteine residues, is the larger biological framework (Figure 63).

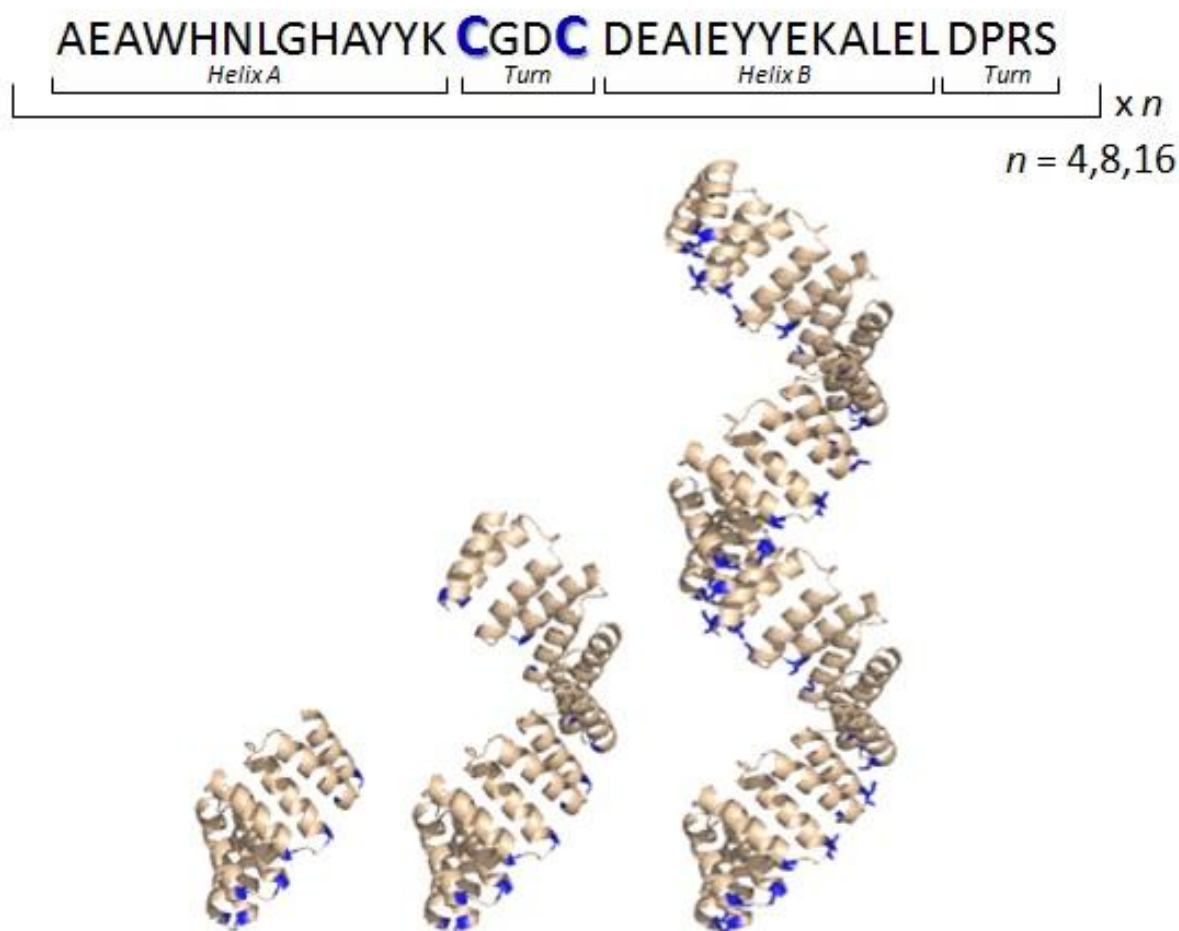


Figure 63. Mutated CTPR4, CTPR8 and CTPR16 protein. Top, sequence of amino acids of the repeat module, highlighting the positions 14 and 17 that have been mutated to cysteine. This module is used to form proteins with n repeats, $n = 4, 8$ and 16 . Bottom, ribbon representation of mutated CTPR protein. From the left to right: CTPR4_CP, CTPR8_CP and CTPR16_CP protein structural model based on the structure of CTPR8 (PDB ID: 2HYZ). The mutated cysteine residues for the porphyrin attachment are highlighted in blue in each protein.

The designed proteins and zinc metalloporphyrin **2** are conjugated using the maleimide-cysteine chemistry to obtain CTPR4_CP-2, CTPR8_CP-2 and CTPR16_CP-2 conjugates (Figure 64.A) as explained *Materials and Methods M.2.2.b.2*). The analysis of the samples by gel electrophoresis shows that conjugates of higher molecular weight than their corresponding CTPR proteins is obtained (*Materials and Methods M.2.2.b.2*). Specifically, conjugates present more than double weight comparing with their non-functionalized proteins, meaning a near quantitative yield in the conjugation reaction of the cysteine residues. When the gel is imaged without staining, the fluorescence signal can be detected only in the lanes containing conjugates holding the porphyrin moieties, as expected (Figure 64.B).

The purification of the protein-porphyrin conjugates from the excess of free porphyrin is a mandatory step to produce homogenous hybrid structures. It is successfully carried

out using size exclusion chromatography (*Materials and Methods M.2.2.b.3*). As predicted, the elution time decreases while increasing the size of the conjugate (Figure 64.C).

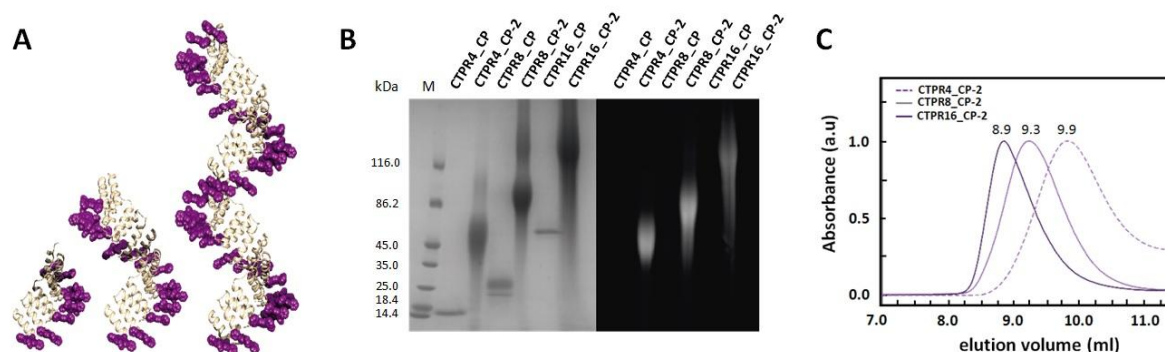


Figure 64. Analysis and purification of conjugates. A) Representation of the three conjugates, front the left to right: CTPR4_CP-2, CTPR8_CP-2 and CTPR16_CP-2. B) SDS-Page electrophoresis gel of the conjugation reactions: Molecular weight marker (lane 1), CTPR control proteins (lane 2, 4 and 6) and CTPR-porphyrin conjugates (lane 3, 5 and 7). The gel is shown prior staining imaged using UV-light to monitor the fluorescence of the porphyrins (right panel) and after Coomassie Blue staining (left panel). C) Size exclusion chromatogram showing the different elution volumes of the three conjugates.

A precise quantification of the number of porphyrins attached to each protein is not possible to elucidate by mass spectrometry due to the high content in porphyrin molecules and the high molecular weight of these new conjugates (80-87 KDa for CTPR8_CP-2 with 14-16 porphyrins and 158-171 KDa for CTPR16_CP-2 with 28-32 porphyrins), which probably difficult the volatilization of the samples. The effectiveness of the conjugation reaction is qualitatively demonstrated using absorption spectroscopy by the relative intensity of the Soret absorption band of porphyrin **2**, at 430 nm, compared to the protein absorption band, at 280 nm (*Materials and Methods M.2.2.b.4*). Figure 65.A shows, in one hand, that the Soret and the Q-bands in the CTPR8_CP-2 and CTPR16_CP-2 conjugates are not displaced to neither higher or lower wavelengths comparing to the previously reported UV-vis spectrum of CTPR4_CP-2, thus, the location of the porphyrins and their electronic communication should be comparable, that is, establishing J-type aggregates. In the other hand, the ratio of the absorbance at $\lambda = 430$ nm respecting to $\lambda = 280$ nm in the three conjugates unambiguously prove that the relative amount of porphyrins is maintained while the dimension of the protein is enlarged. In our previous section, it is demonstrated by mass spectrometry that 6 to 8 porphyrins in average are attached to the CTPR4_CP. Therefore, it can conclude that the effectiveness of the conjugation reaction is maintained in all the cases and is unaffected by the size increment in CTPR8_CP-12 to 16 porphyrins and CTPR16_CP-24 to 32 porphyrins (Figure 65.B).

Regarding secondary structure of the protein scaffold within the conjugate and the spatial organization of the porphyrin moieties, CD studies are carried out (*Materials and*

Methods M.2.2.b.5). In the absorption region of proteins, that is, from 260 to 190 nm, all conjugates reveal the feature CD signal for α -helical secondary structure presented in CTPR proteins (Figure 65.C). Furthermore, optical activity is exhibited in the Soret absorption band of achiral porphyrins, indicating a chiral arrangement of these chromophores in the conjugates (Figure 65.D).

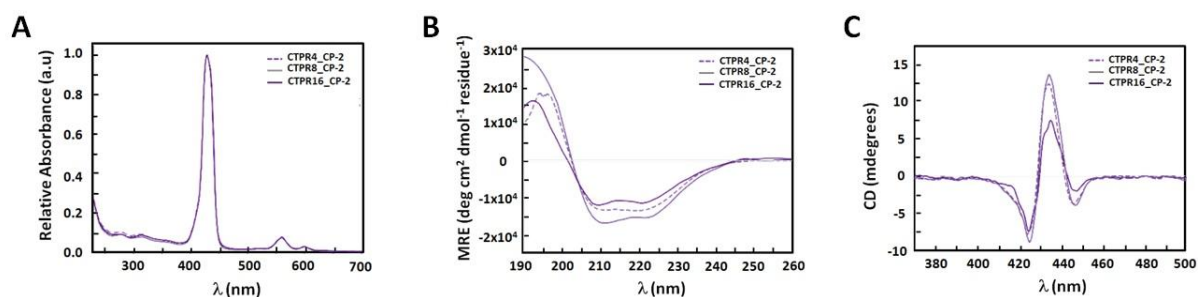


Figure 65. Spectroscopic characterization of bio-organic conjugates. A) Normalized UV-vis spectra of conjugates CTPR4_CP-2, CTPR8_CP-2 and CTPR16_CP-2 conjugates. C) CD spectra of the conjugates in the protein absorption region. D) Same than C) but in the Soret band absorption region of the zinc-metalloporphyrin.

In conclusion, in this section we generate CTPR proteins with different number of repeats, taking advantage of the structural understanding of the CTPR4_CP protein presented in the previous section 2.4.2.2.a to form different number of porphyrin arrays. We see that in all the conjugates the CTPR protein retains its signature helical structure and imposes order and chirality into the porphyrin molecules that show efficient stacking interactions. The obtained results confirm the successful ability of the mutated CTPR proteins to be used as a robust biological scaffold for ordering organic chromophores.

2.4.2.2.c Donor/acceptor nanohybrids based on CTPR and CTPR-porphyrin conjugates and SWCNTs.

In the previous section, we presented a novel bioinspired approach in which photoactive porphyrin derivatives are organized and oriented along a bio-organic conjugate using a designed CTPR protein as a precise scaffold. In this part, we have expanded our recent design to the creation of protein-single walled carbon nanotube (SWCNT) bionano donor-acceptor hybrids using the designed CTPR4_CP, CTPR8_CP and CTPR16_CP proteins together with their corresponding conjugates with a zinc metalloporphyrin derivative, in order to clarify the importance of the preorganization and steric hindrance of these systems for the achievement of our goal.

Symbiosis between carbon nanotubes (CNTs) and biological entities to create nanohybrid materials is a cutting-edge area with high impact in fields such as medicine, sensing and energy production [277, 278]. Rationally, a requirement of these biological nanohybrids for being applied in light-harvesting devices is the introduction of an appropriate electron donor entity, acting the CNT as the electron acceptor material.

Up to now, DNA-CNT bionano hybrids are leading this research area due to the well-established and accurate design and manufacture of artificial nucleic-acid structures [279-281]. In this line, a great example presented by D'Souza and co-workers can be found in literature in which a three-component hybrid, porphyrin-DNA-CNT, is utilized for photoinduced electron transfer processes [282]. In contrast, the use of proteins for the generation of protein-CNT hybrids for optoelectronic applications is not so expanded presumably in behalf of the gap between the developments of *de novo* protein design comparing with that of DNA nanotechnology. However, in the last years, both computational and experimental studies have demonstrated the great potential of these biological frameworks that are proteins and oligopeptides, for the functionalization and organization of CNTs [283-287]. In this sense, some evaluations of non-covalent interactions between complex proteins and these carbon nanostructures have been performed during the last years [283].

As example, Fang and col. demonstrated by Molecular Dynamic (MD) simulations the introduction of single-walled carbon nanotubes (SWCNT) of different diameters inside the hydrophobic pocket of a protein [288]. Trp and Tyr residues in this binding site of the protein were found to stabilize this interaction through π -stacking between the aromatic side-chain of the amino acids and the surface of CNTs. Interestingly, this contact resulted in the alteration of the functionality of the protein, inhibiting the recognition of its ligand.

In the other example, reported by Bonifazi and col., the interactions between the monoclonal Ab Cetuximab (Ctx) and carbon nanotubes (CNT) of different diameters were studied by experimental and theoretical approaches [289]. Investigations by MD simulations and Density Functional Theory (DFT) calculations revealed that the most interacting residues with CNTs were the charged (Arg, Lys and Glu) and the polar ones (His and Thr). These charged or polar moieties in the amino acids pointed towards the solvent, while the long alkyl chains would act as surfactants when contacting with the CNTs. Thus, the nature of this binding was, as the previous example, hydrophobic but involving other type of amino acids and not only aromatic ones.

Despite the advances in the field and the understanding of the interaction between the proteins and CNT, no precedents can be found about donor-protein-CNT hybrids. In this part, it is tested the CTPR proteins and their corresponding bio-organic porphyrin-protein conjugates as wrapping agents for SWCNTs, for obtaining stable nanohybrids. In this case, it is taken advantage of the peculiar superhelical conformation of CTPR proteins

that have an optimal inner cavity for accommodating this carbon nanostructure (Figure 66.A).

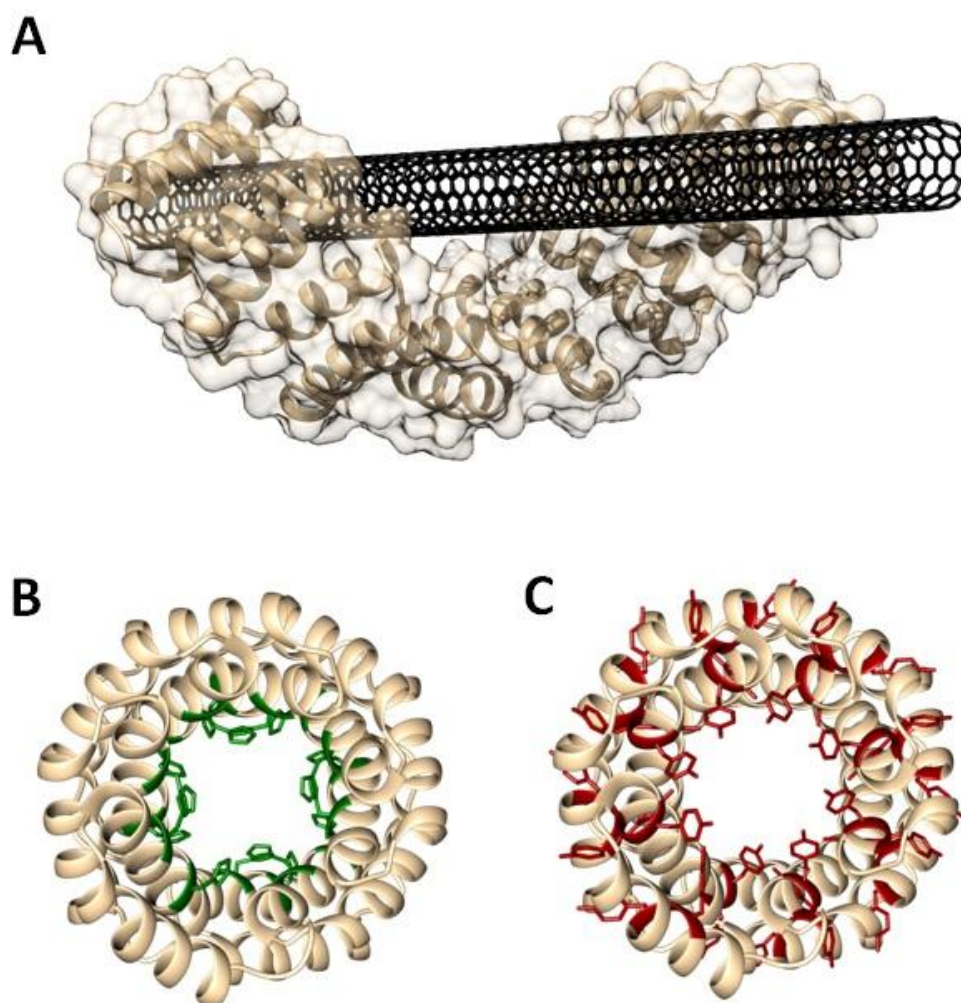


Figure 66. Suitability of morphology and composition of CTPR proteins with SWCNTs. A) Model of a CTPR8 (PDB ID: 2AVP) and a (7,6)-SWCNT showing the morphological compatibility between the concave face of the protein and the SWCNT surface. B) View along the superhelical axis of a CTPR8 highlighting the mutated His residues (Y5H and N9H) in green. C) View along the superhelical axis of a CTPR8 highlighting the mutated Tyr residues in red

Favorable bindings between protein residues and CNT surface are mandatory for an effective interaction. In this sense, a deep search in literature of the reported non-covalent forces ruling the adsorption of peptides and proteins onto CNT surfaces is made. This examination is aimed to validate the composition of the protein and, if possible, to modify those non-conserved positions of the repeat sequence by more interacting amino acids, without affecting the structure of the protein framework.

As described before, the strongest interacting residues with CNTs are those bearing aromatic units in their side chains by establishing π - π and $XH\cdots\pi$ ($X = C, N$) contacts [283, 290]. In this sense, Trp, Tyr, Phe and His are good candidates to be localized in those positions forming part of the concave face in CTPR proteins. Among these, important experimental evidences demonstrated that peptide sequences rich in Trp and His show strong adsorption to the CNT surface [285, 288, 291].

As previously described, binding residues on the concave face in TPR proteins are mainly localized in helix A, and bind with high specificity to target guests. With this knowledge and the perfect understanding of the consensus positions in TPR sequence, positions Y5H and N9H, both localized in the concave protein surface, are modified by His residues to increase the interaction to the CNT wall (Figure 66.B) (*Materials and Methods M.2.2.c.1*). Luckily, Tyr residues are found in the original sequence of CTPR unit, two of them with the phenol group pointing through the concave face (Figure 66.C). Altogether, two His and two Tyr in each repeat fragment could play a major role in the adsorption of the protein to the CNT (Figure 66.A). As expected, the two Cys and two His mutations in each repeat unit do not significantly affect the helical structure of the protein scaffold. Thus, well-expressed, stable protein samples with the same α -helical structure as the parent protein are obtained (Figure 67) (*Materials and Methods M.2.2.c.2*).

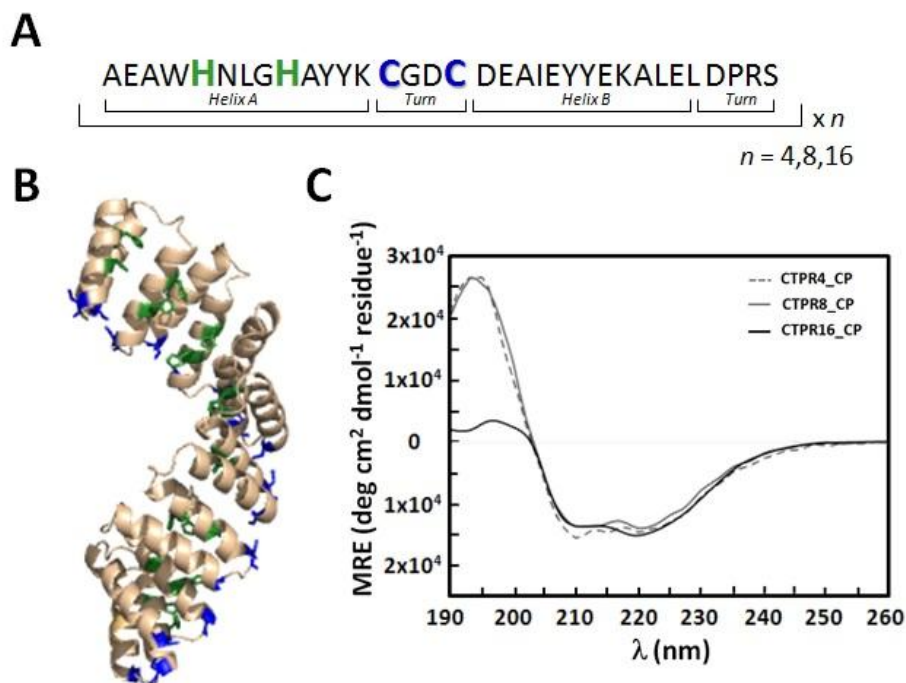


Figure 67. Design and synthesis of mutated CTPR proteins. A) Sequence of amino acids of the repeat module, highlighting the positions 5 and 9 mutated to Histidine (green) and 14 and 17 mutated to Cysteine (yellow). B) Ribbon representation of the mutated CTPR8 with Histidine residues in green and Cysteine residues in yellow. C) CD spectra of mutated CTPR4, CTPR8 and CTPR16.

After obtaining and characterizing the corresponding CTPR proteins and conjugates, they are evaluated as wrapping agents for SWCNTs (*Materials and Methods M.2.2.c.3*). Both of them form water-soluble assemblies of SWCNTs, producing aqueous solutions that are stable for months in the fridge that are stable even after centrifugation at high speed1 (more than 15000 r.p.m) during 30 min (Figure 68).

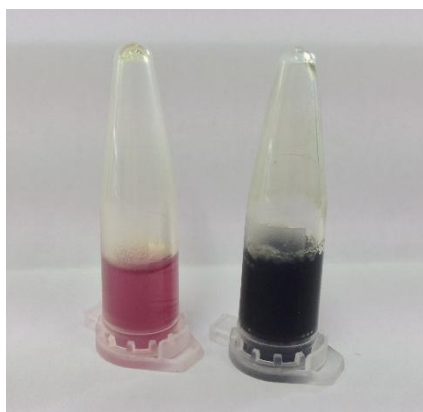


Figure 68. Solutions containing CTPR8_CP-2 conjugate (left) and CTPR8_CP-2/SWCNT nanohybrid (right), in PBS buffer.

Insights into protein-SWCNT and conjugate-SWCNT ground-state electronic interactions came from UV-*vis* spectroscopic experiments (Figure 69) (*Materials and Methods M.2.2.c.4*). The study performed on the protein-SWCNT nanohybrids is depicted in figure 69.A/B. Although only three lengths of CTPR proteins have been used for this study, some trends can be pointed out: at the same amount of protein, that is 0.005% (w/v), the largest systems are able to more effectively dissolve SWCNTs in aqueous media. In particular, CTPR16_CP, with two completed superhelical turns, are the most efficient wrapping agents followed by CTPR8_CP, with one superhelical turn. CTPR4_CP, being the shortest biological scaffold, is also the least effective in this issue. Figure 69.B shows the absorption intensity at five different wavelengths as function of the protein length in the nanohybrid. In all the cases, data can be more precisely plotted to a non-linear fitting rather than a linear one, attesting that the limit of the effectiveness of CTPR proteins as wrapping agents of SWCNTs is not going further than 16 to 20 repeat units. This experimental data can be explained by the increment in the cooperativity of the amino acids responsible of the wrapping that are in the cavity of the proteins when the system is preorganized by a larger protein (CTPR16_CP) more than in a shorter one (CTPR4_CP).

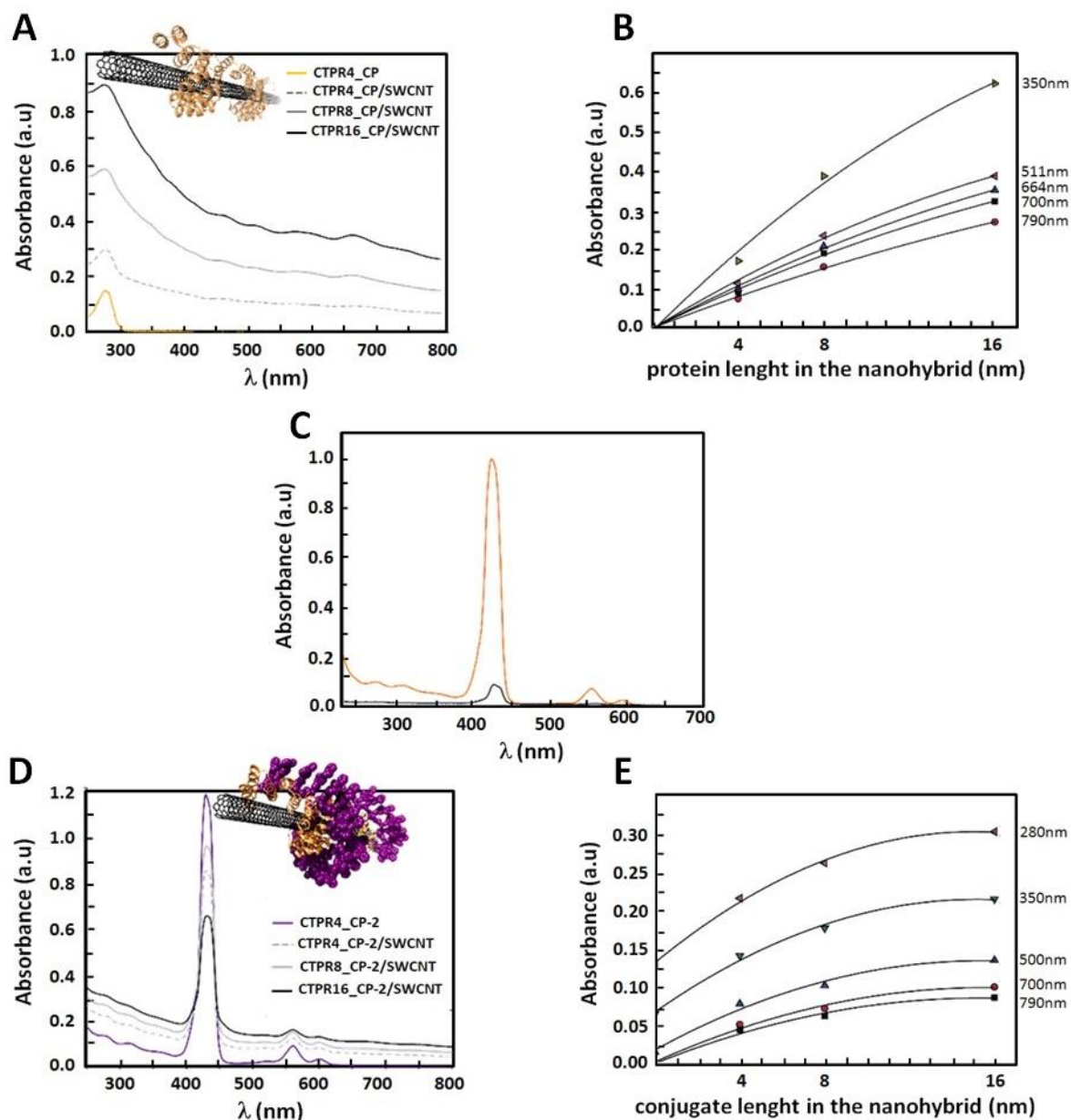


Figure 69. Spectroscopic characterization of the bio- and the bio-organic nano hybrids. A) UV-vis spectra CTPR4_CP (yellow line), CTPR4_CP/SWCNT (grey dashed line), CTPR8_CP/SWCNT (grey solid line) and CTPR16_CP/SWCNT (black solid line). B) Absorption intensities of protein-SWCNT nano hybrids as function of protein lengths, at five different wavelengths. C) UV-vis spectra of **2** (orange line) and **2**/SWCNT (black line). D) UV-vis spectra of CTPR4_CP-**2** (purple line), CTPR4_CP-**2**/SWCNT (grey dashed line), CTPR8_CP-**2**/SWCNT (grey solid line) and CTPR16_CP-**2**/SWCNT (black solid line). E) Absorption intensities of conjugate-SWCNT nano hybrids as function of conjugate lengths, at five different wavelengths.

Meanwhile, the same comparative study is performed over the three different conjugates. For initially establishing the influence on the exfoliation and supramolecular functionalization of SWCNTs of the zinc metalloporphyrin fragment **2** in the conjugates, a control experiment is carried out. To a solution containing non-conjugated porphyrin **2** (6.1 μM in PBS buffer) around 1 mg of SWCNTs is added and the same aforementioned

procedure is followed, that is, prolonged sonication and high-rating centrifugation. Figure 69.C shows the result, remarking the slight presence of SWCNTs in solution and more than 90% of porphyrin **2** precipitates out, probably because of π - π interaction with the surface of nanotube, establishing porphyrin-SWCNT nano hybrids with no stability in water. Thus, it can be anticipated that the photoactive moieties in the conjugates should play a negligible role in the photophysical characteristics of the nano hybrids.

UV-*vis* studies performed on the conjugates with SWCNTs are illustrated in Figure 69.D/E with comparable results relating to their analogue non-conjugated proteins. The trend is similar to that in the proteins, that is, at the same amount of the conjugates, longer scaffolds are more effective wrapping agents for the carbon nanostructures. Nevertheless, the maximum efficiency of these bio-organic structures seems to be attained with lengths near or even lower than 16 nm, in contrast to the corresponding proteins (Figure 69.E). This fact can be explained by the steric hindrance increment and the protein flexibility reduction, both induced by the introduction of the bulky chromophores. It is also visible that the Soret band of the zinc-metalloporphyrins decreased, in spite of the increasing intensities in the whole spectral region of protein-SWCNT. This behavior can be attributed to interactions between this chromophore and the nanotube surface of the same nano hybrid or others nanotubes in the solution.

Structural integrity of the proteins and conjugates, when they constitute their corresponding nano hybrids with SWCNTs, is proven by CD spectroscopy (*Materials and Methods M.2.2.c.2*). Figure 70 shows how the α -helix signature of CTPR framework is maintained in all the cases. The only difference to point out is the slight reduction in the dichroic signal comparing with the pristine samples without SWCNTs. This experimental finding can be attributed to a decrease of the protein/conjugate concentration during the centrifugation process. These CNTs without a dense covering of protein material can precipitate out of the aqueous solution, carrying with them some wrapped macromolecules. Nevertheless, partial unfolding of the α -helices can be another explanation to consider.

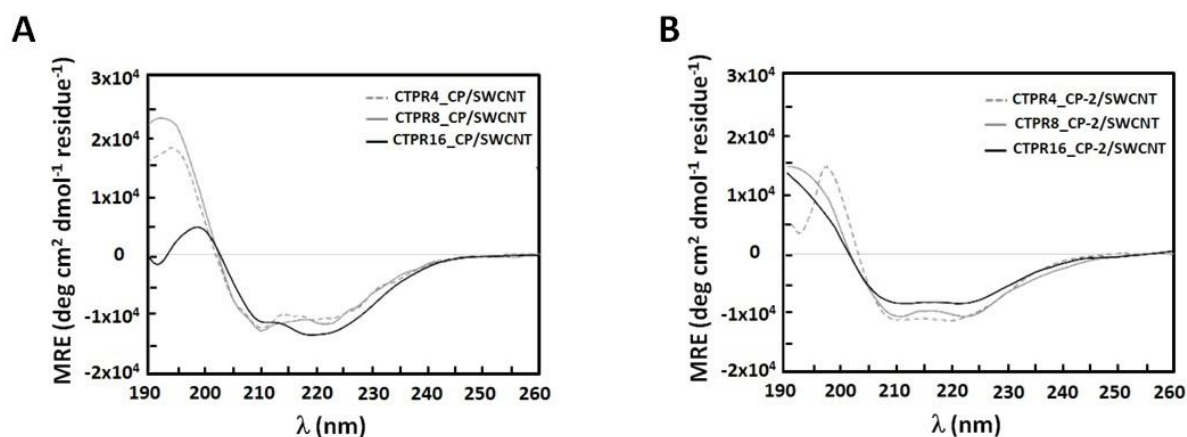


Figure 70. CD spectra of CTPR-SWCNT and CTPR-porphyrin-SWCNT nano hybrids. A) CD spectra of CTPR4_CP protein with SWCNT in dashed grey, CTPR8_CP protein with SWCNT in grey and CTPR16_CP protein with SWCNT proteins in black. B) CD spectra of CTPR4_CP-2 conjugated with SWCNT in dashed grey, CTPR8_CP-2 conjugate with SWCNT in grey and CTPR16_CP-2 conjugate with SWCNT proteins in black.

Once elucidated CTPR16_CP protein and its corresponding CTPR16_CP-2 conjugate as the best dissolving agents of SWCNTs, the following assays are described only for their resultant nano hybrids: concomitant studies came from Raman spectroscopy which provide really valuable information regarding the interaction between both entities (*Materials and Methods M.2.2.c.5*). The first evidence to appoint is that D band intensities, relating to G bands, reveal no difference comparing with pristine SWCNTs, as expected from a non-covalent interaction (Figure 71.A). In contrast, G-modes, shows in figure 71.B, downshifted from 1592 cm^{-1} , in pristine SWCNTs, to 1589 and 1585 cm^{-1} , in CTPR16_CP/SWCNT and CTPR16_CP-2/SWCNT, respectively. Previous studies have demonstrated that the G band peak in CNTs shifts to lower frequencies when doping with electron donor agents and to higher frequencies with electron acceptor ones [245, 246, 292]. However, in the particular case of using peptides as wrapping agents of SWCNTs, even with modified Phe residues by Tyr residues (more electron donor moieties), the G-mode is upshifted comparing with uncoated SWCNTs. Only when comparing a Phe-rich peptide/CNT with a Tyr-rich peptide/CNT, a slight downshift of the G-mode is appreciated (0.6 cm^{-1}) [291]. In our current case, a remarkable downshift of 3 cm^{-1} in CTPR16_CP/SWCNT comparing with pristine SWCNT is found. This experimental finding attested the strong interactions of electron-donating residues in CTPR16_CP with the surface of the nanotube. In particular, 32 His and 32 Tyr presented in the concave face of CTPR16_CP can be the responsible of such substantial shift to lower frequencies in the G band peak of SWCNTs. In the case of CTPR16_CP-2 conjugate, the effect is even more forceful, with a downshift of 7 cm^{-1} , supplying evidence of a possible charge transfer process between the porphyrins and the SWCNTs.

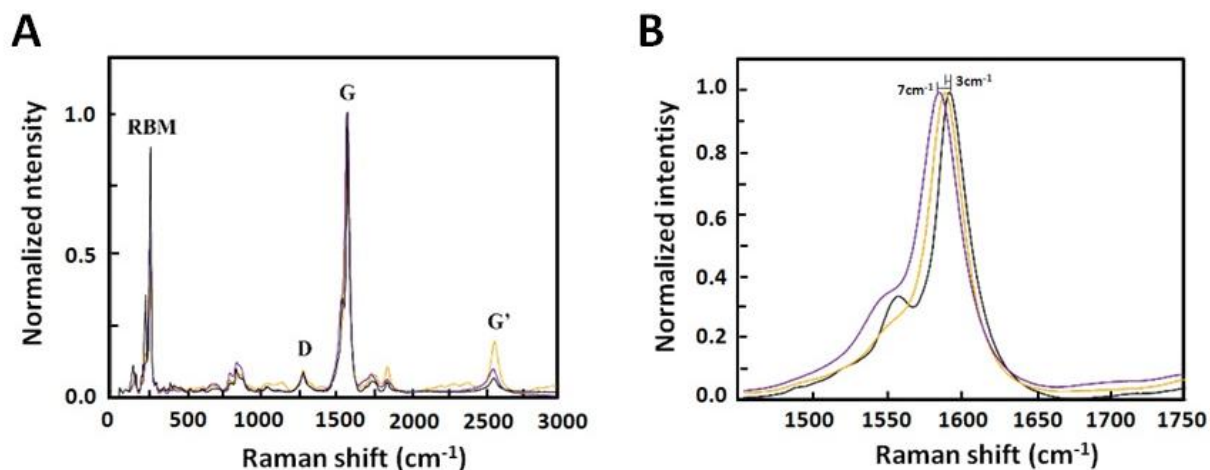


Figure 71. Raman spectroscopy of the bio-organic nanohybrids. A) Normalized Raman spectra from 50 to 3000 cm^{-1} . B) Normalized Raman spectra in the region of G-mode. All of them have been acquired at the excitation wavelength of 785 nm. Pristine SWCNTs (black line), CTPR16_CP/SWCNT (yellow line), CTPR16_CP-2/SWCNT (purple line).

Transmission electron microscopy (TEM) provides conclusive support of the existence of the protein-SWCNT and conjugate-SWCNT nanohybrids (*Materials and Methods M.2.2.c.6*). TEM images are acquired by drop-casting the corresponding aqueous solution onto a carbon grid. Figure 72 shows that the nanotubes are in an excellent exfoliated state and most of the individual SWCNTs show their walls covered or surrounded by organic material, that is, the protein or the conjugate depending on the sample.

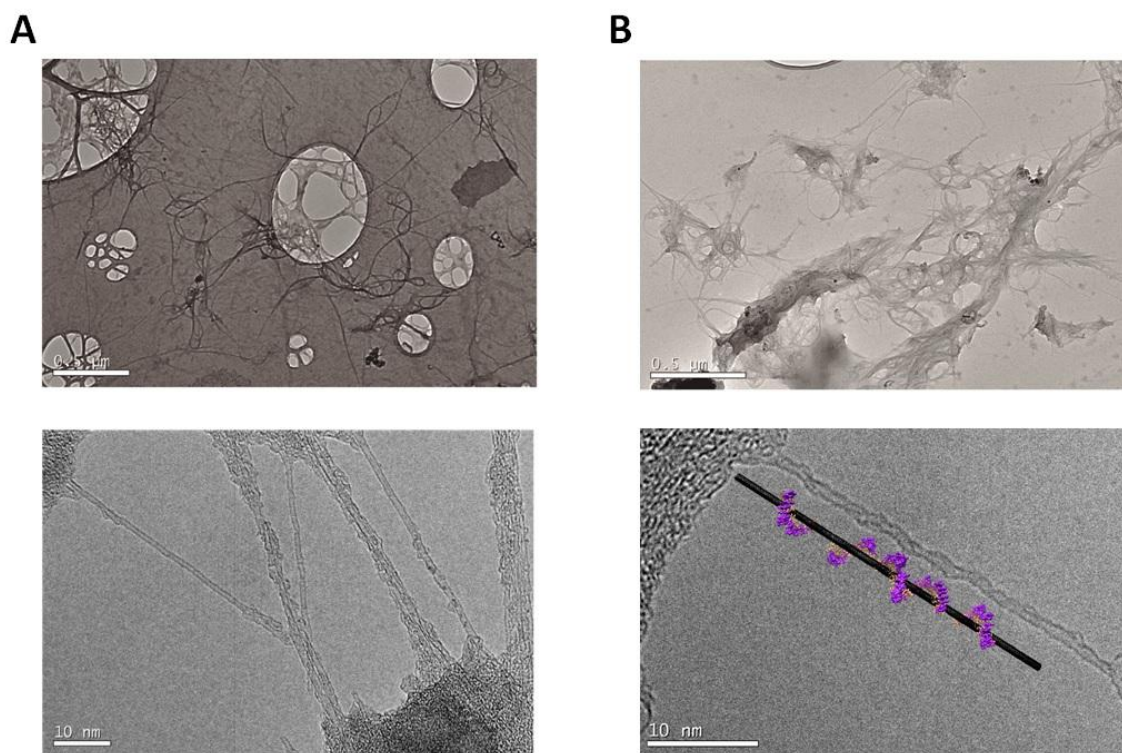


Figure 72. TEM micrographs at different magnifications A) CTPR16_CP/SWCNT B) CTPR16_CP-2/SWCNT.

In our previous section it is described that in the solid state, when forming thin films, porphyrin-protein conjugates preserve their structural integrity. Now, they are produced the aforementioned thin films with these new hybrid materials containing and it is tested their features by X-ray diffraction experiments. XRD data of the mutated CTPR16_CP, its conjugate CTPR16_CP-2 and their corresponding nanohybrids CTPR16_CP/SWCNT and CTPR16_CP-2/SWCNT all revealed a most intense broad signal at 2θ around 19 to 21 ° (Figure 73) (*Materials and Methods M.2.2.c.7*). This value with a d -spacing of around 0.43 nm has been previously associated to the α -helical pitch of tandem repeats configuring the superhelix, as observed in the high-resolution crystal structure of the protein. Therefore, CTPR16_CP scaffold retains its superhelical conformation both with the enormous amount of porphyrins covalently attached to the structure and, more challenging, the interacting SWCNTs.

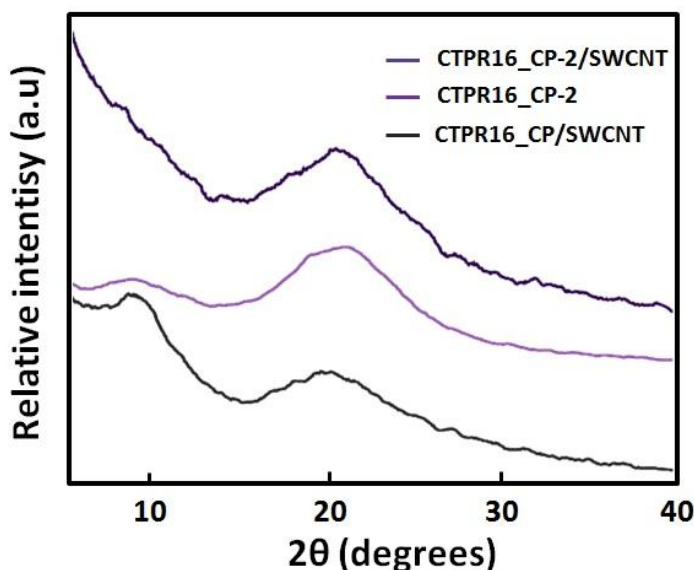


Figure 73. X-Ray (XRD) diffractograms of the films. The XRD spectrum of CTPR16_CP/SWCNT in grey, the XRD spectrum of CTPR16_CP-2 spectrum in light purple and the XRD spectrum of CTPR16_CP-2/SWCNT in dark purple.

The present soluble nanohybrids accommodating conducting SWCNT and photoactive porphyrin systems are interesting for soft electronics or bioelectronics fields. The photo-carrier injection and intrinsic charge carrier transporting properties of CTPR16_CP-1, CTPR16_CP/SWCNT, and CTPR16_CP-2/SWCNT are investigated by FP-TRMC technique as in the section 2.4.2.2.a with CTPR4_CP-1 and CTPR_CP-2 conjugates (*Materials and Methods M.2.2.c.10*). As is similar for CTPR4_CP-2, a thin film of CTPR16_CP-2 shows a clear conductivity transient upon photoexcitation with 355 nm laser pulse (Figure 74.A) with long-lived stable charge carrier species, suggesting significant contribution from the local motion of positive charge carriers on porphyrin-arrays. The obtained transient

absorption spectra (TAS) support the formation of photo-oxidated states on the porphyrin-arrays (figure 74.B) [275]. Upon laser flash, a drop-cast film of CTPR16_CP-2 shows a conductivity transient with slow decay features (Figure 74.B). In fact, the normalized profiles of FP-TRMC and TAS at 580 nm give almost identical kinetic traces (Figure 74.C) indicating that hole transport is the dominant factor for the observed local-scale photoconductivity of CTPR16_CP-2 under air, as it is observed in the case of the CTPR4_CP-2 in section 2.4.2.2.a. By using a typical absorption coefficient of the radical cation of tetraphenyl free-base porphyrin [259], the calculated yield of radical cations are $\Phi = 4.0 \times 10^{-2}$, leading the local mobility of holes on porphyrin-arrays as $\mu_h = 4.5 \times 10^{-3} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ (Figure 74.C). Thus, the electrical conductivity ($\Phi \Sigma \mu$) for the CTPR16_CP-2 film is $1.8 \times 10^{-4} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, that is almost 3 times more than the value obtained for the CTPR16_CP-2 in the section 2.4.2.2.a ($6.5 \times 10^{-5} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$). This improvement in the conductivity can be attributed to the preorganization of the system in a larger protein (CTPR16_CP) where the distance between the porphyrin is suitable for the π - π interactions between them more than in a shorter one (CTPR4_CP).

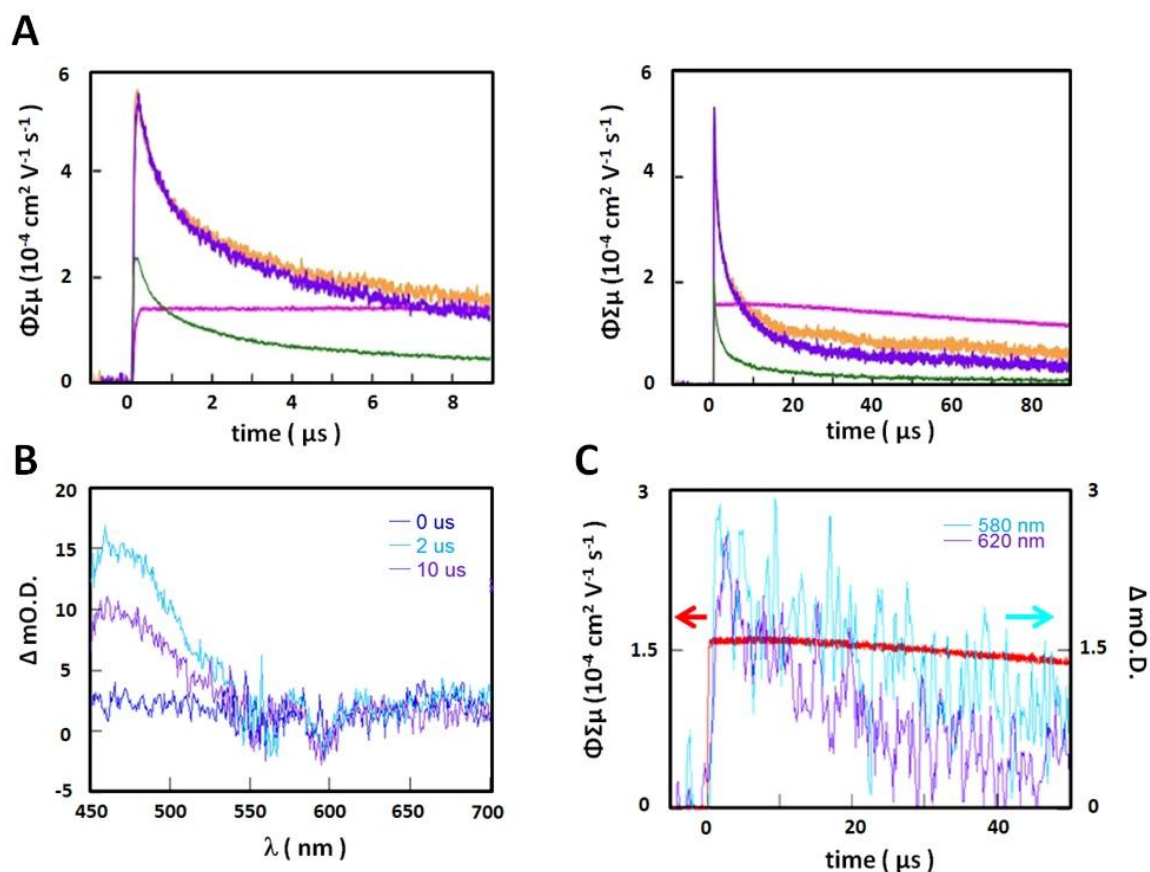


Figure 74. Photoconductivity characterization of the films. A) Kinetic traces of conductivity transients of a film of CTPR16_CP-2 (magenta), CTPR16-CP/SWCNT (yellow), and CTPR16_CP-2/SWCNT (purple) photoexcited at 355 nm, and CTPR16_CP-2/SWCNT (green) photoexcited at 420 nm at the time range of 10 μs on the left and 100 μs on the right. B) Transient absorption spectra recorded for CTPR16_CP-2 upon photoexcitation at 355 nm, $4.6 \times 10^{15} \text{ cm}^{-2} \text{ pulse}^{-1}$. C) Kinetic traces observed at 580 and 620 nm with the photoconductivity trace. The CTPR16_CP-2 exhibits clear photoconductive nature when they are excited by 580 nm.

A dropcasted film of CTPR16_CP/SWCNT provides a transient with prompt rise and moderate decay features when it is photoexcited at 355nm (Figure 74.A). Namely, CTPR16_CP/SWCNT films retain the conductivity of SWCNTs. It is disclosed that CTPR16_CP-2/SWCNT displays a conductivity transient whose profile is almost identical with that of CTPR16_CP/SWCNT, indicating that the major charge carriers of electrons are injected into SWCNTs and mobile along the 1D-structures (Figure 74.A). However, presence of Zn porphyrin chromophores result in the effective charge carrier availability of excitation light over wide wavelength ranges. For example, photoexcitation of CTPR16_CP-2/SWCNT by 420 nm laser pulses also exhibited photoconductive nature (Figure 74.A). This observation suggests that Zinc porphyrins absorb visible lights and mobile charge carriers are generated on SWCNTs.

In conclusion, in this section we present a designed protein serving as wrapping agent of SWCNTs. The optimal superhelical conformation of CTPR proteins for this purpose and the possibility to mutate some residues with more interacting amino acids have made possible such ambitious goal. We have highlighted the effectiveness of our approach by a wide range of techniques, corroborating the introduction of the SWCNT inside the cavity of the superhelix and the stability of the folding state of the protein even with the carbon nanostructure inside. Additionally, we have performed a comparative study regarding the length of the protein, using CTPR4_CP, CTPR8_CP and CTPR16_CP, concluding that preorganization of the superhelical conformation is highly beneficial for improving their wrapping capability. Moreover, we have expanded this approach to the conjugates with metalloporphyrin **2**, overlooking the great impact that these newly nanohybrids containing donor and acceptor entities could have in materials science. The ability of the conjugates for coating SWCNTs has been demonstrated to be virtually identical than their corresponding non-conjugated proteins. Only some differences derived from their higher steric hindrance and lower flexibility have been found.

Finally, we take advantages of the assembly properties of the CTPR protein to fabricate solid functional structures. We measure the photoconductivity characteristics of the CTPR16_CP-2, CTPR16_CP/SWCNT and CTPR16_CP-2/SWCNT because the CTPR16 shows the best wrapping capability. On one hand, we observe that CTPR16_CP-2 shows a clear conductivity transient with long lived stable charge carriers suggesting significant contribution from the local motion of positive charge carriers on porphyrin-arrays. On the other hand, we observe that CTPR16_CP/SWCNT and CTPR16_CP-2/SWCNT provide a transient with prompt rise and moderate decay feature, meaning that these films retain the conductivity of CNTs indicating that the major charge carriers of electrons were injected into CNTs and mobile along the 1D-structures. However, the introduction of the Zn porphyrins resulted in the effective charge carrier availability of excitation light over wide wavelength ranges because of the photoexcitation of the porphyrins.

2.4.2.3. Conclusions

In this section, we have explored, in one hand, the capacity of the CTPR proteins to form organized donor-acceptor systems. The control in the structure and stability of the proteins, makes them ideal for organizing these systems in which the distances between the elements are crucial to improve their efficiency. In this work, we use porphyrins as electron-donor molecules and we are able to form arrays with different number of porphyrins using CTPR proteins of different lengths. Moreover, we take advantage of the control in the CTPR self-assembly to form a functional solid film in which the porphyrins are organized in the solid state.

Additionally, we explore the use of CTPR proteins as a wrapping agent for carbon nanotubes, forming a biohybrid electron donor-acceptor system. From these results, we conclude that the cooperativity of the amino acids responsible of the wrapping in the protein has influence in the wrapping effectivity. Thus, a larger protein, such as CTPR16_CP, has more wrapping capability than a shorter CTPR4_CP. Using CTPR16_CP protein we form a CTPR16_CP/SWCNT conjugate and perform solid films in which the electrons are transported by the SWCNT carriers when we photoexcited the conjugate at 355 nm. In this sense, we take advantage of the control over the conjugation between CTPR proteins and porphyrins to expand the photoexcitation range, taking advantage of the properties of the porphyrins. In these studies we use as electron donor molecules porphyrins and CTPR protein and as electron acceptor molecule carbon nanotubes to form biohybrid systems, where the elements interact in a controlled way, because of the order imposed by the CTPR proteins.

2.4.3. Mimicking natural photosystems

2.4.3.1. Motivation and objectives

In this section, we explore the use of designed repeat protein scaffolds to create single integrated artificial photosystem (Figure 75). The understanding of the stability and structure of the CTPR proteins allow us to accurately design proteins in which we control, with atomic resolution, the positions of the functionalities in the system. This is essential to create complex multifunctional systems such as photosystems with different properties [293] in which the interaction between the different elements of the system can be defined (Figure 75).

Nowadays, at least 80 percent of the energy that people use to drive, heat their homes, and power gadgets comes from fossil fuels such as coal, oil, and natural gas. However, this driving force is facing a monumental crisis as the rate of consumption by today's energy-dependent technology substantially surpasses its geological formation. Therefore, there is a growing interest in finding new energy resources to replace unsustainable fossil fuels [294, 295].

Hydrogen (H₂) is expected to play a prominent role as sustainable fuel in a future economy based on renewable energy sources [296]. As H₂ energy is sustainable and environmentally-friendly, H₂ production on a catalyst is the focus of the great amount of investigations in this field [297]. At the present, a major challenge for industrial-scale H₂ production is the development of inexpensive and durable catalysts [295]. Nowadays, hydrogen production relies on steam reforming of hydrocarbons at high temperatures or on the use of precious metal catalysts, such as platinum, and can't be scaled up in a sustainable manner. Nature, in contrast, utilizes a class of enzymes called hydrogenases that reversibly catalyze the reduction of protons to molecular hydrogen in oxygenic natural photosynthetic processes [298-302]. In this sense, artificial photosynthetic H₂ production is a highly-promising strategy to convert solar energy into hydrogen energy for the relief of the global energy crisis and much of the research in this field has focused on understanding and engineering hydrogenases for future applications [303]. The main limitation of using natural enzymes to scalable hydrogen production is hampered by their difficulty of overexpression and their instability outside the cell [304, 305]. For these reasons, it is important to develop robust synthetic systems, based on these natural enzymes, as alternatives to the natural enzymes [306-309].

In this way, protein design offers a new tool to mimic these natural processes in a more robust and stable system than natural enzymes. There are some works showing that it is possible to tune the stability of the protein only changing some amino acids,

obtaining proteins that are more stable than natural proteins [310]. Thus, there is a growing interest of using these novel designed proteins to mimic natural photosystems.

In order to generate artificial system for photosynthetic hydrogen production, the first step is to generate an efficient photosystem at which the light is efficiently captured and converted to electron flow being the electrons transported through redox cofactors with small energy cost. These electrons will have specific energy depending on the redox system and can be used for further catalytic reactions (Figure 75) [311]. Each system is unique according to the chemical nature of their electron acceptors and donors as well as redox potential of the electron transfer system [312]. Thus, in an efficient photosystem the control of the position of the different elements is required in order to convert a maximum fraction of the solar photon energy to chemical energy.

Another challenging requirement is that all components of the systems operate efficiently under the same, preferably benign, environmental conditions. Harsh conditions, especially regarding pH, are likely to limit the durability of components. Thus, single integrated systems are attractive to go for this issue, which is essential for large scale use. However, until today, most of the works are done organizing the different elements in separated system that only interact with each other under certain conditions [313]. In this section, we propose the use of CTPRs to form an artificial photosystem in which all the required elements are integrated in the same protein: the antenna complex and electron transfer system (Figure 75.B).

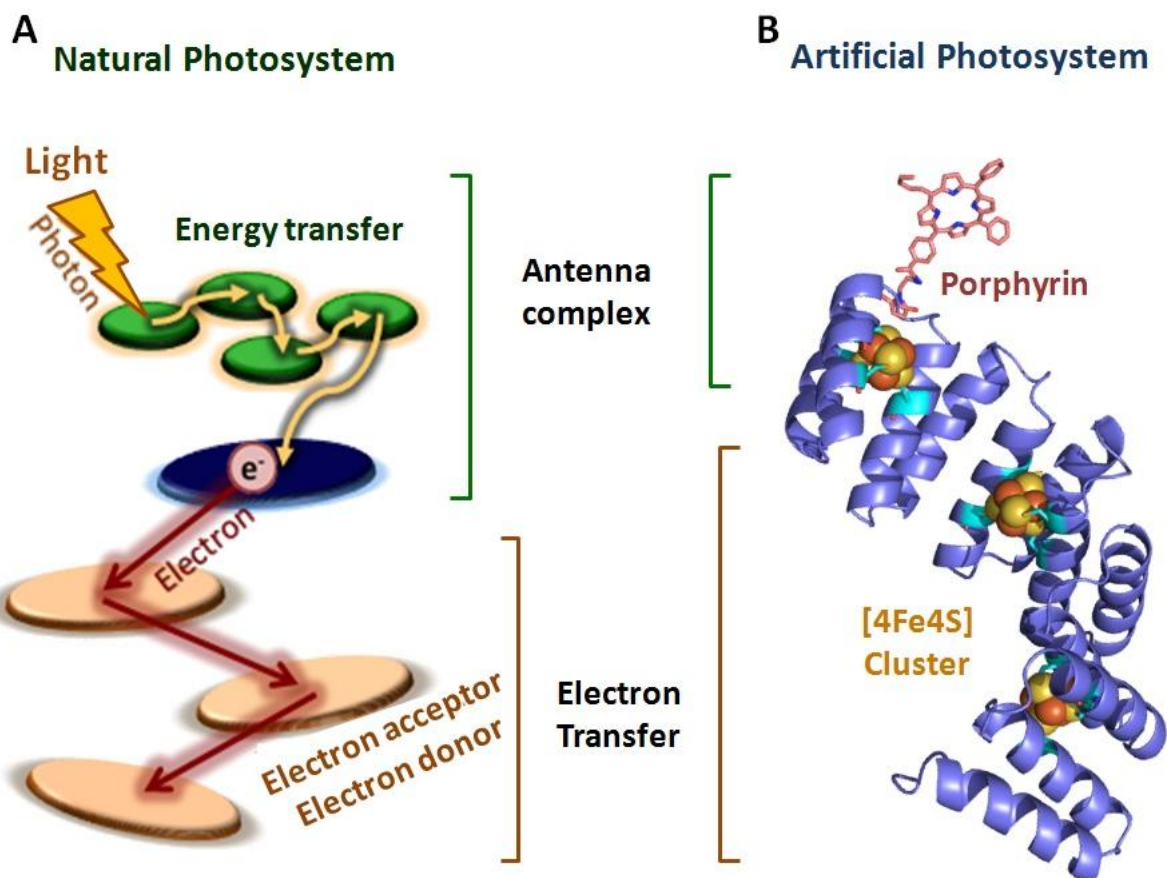


Figure 75. Different parts of a photosystem. A) A scheme of the different parts of a natural photosystem: antenna complex, where the green part is activated with light and performs the energy transfer; the blue circle that represents an electron donor molecule that is activated with the energy of the light. Following the antenna complex, the electron transfer system is represented by orange circles, that transports the electron from the antenna complex without energy cost. B) A CTPR protein designed to mimic natural photosystems: a porphyrin, colored in red, as antenna complex and [4Fe-4S] clusters arranged at defined distances on the CTPR scaffold colored in orange and yellow, as electron transfer system.

2.4.3.2. Results and discussion

First, CTPR proteins are used to allocate the redox cluster that constitutes the electron transfer system (Figure 75). The design of redox active proteins should include the engineering of a channel for the delivery of electrons to and from the active site, at which multiple redox active centers are arranged in a controlled manner. Here, we describe a designed CTPR protein to coordinate different number of iron-sulfur cluster [Fe-S] within its concave face (Figure 76). Frequently, [Fe-S] clusters act as donors or acceptors of electrons over a wide range of potentials, and they are organized in protein-embedded redox chains (of two or more clusters) to transport electrons from the antenna complex of the photosystem to catalytic centers [305, 314-317]. The most common type of [Fe-S] cluster is [4Fe-4S] cubic like cluster which regulates long-range electron transfer to and

from the active site of the complex [305, 318-320]. This cluster is generally coordinated by four cysteines in a variety of protein folds, often in a combination of loops and secondary structure elements [321].

In this section, three CTPR proteins have been engineered to allocate one, two and four [4Fe-4S] clusters (in CTPR2_4cys, CTPR4_8cys and CTPR8_16cys) embedded in the concave face (Figure 76.B). To form the cluster a four-cysteines cubic cavity of about 58 Å located between the A helices of 2 CTPR repeats is designed based on the 4-Cys cavity used by Prof. Ghirlanda's group [322]. Indeed, this 4-Cys cubic cavity is used several times to insert iron-sulfur clusters into natural and designed proteins [323-325]. Manual docking has been used using Pymol program [326] to position an idealized cluster into the 4-Cys cavity formed in the CTPR protein. The designed cavity is filled when the cluster is formed taking into account the dimensions of the cluster (Figure 76.A). The distances between the clusters in the different proteins are designed to be in the range of the electron transfer distances between the redox clusters (between 12-15Å).

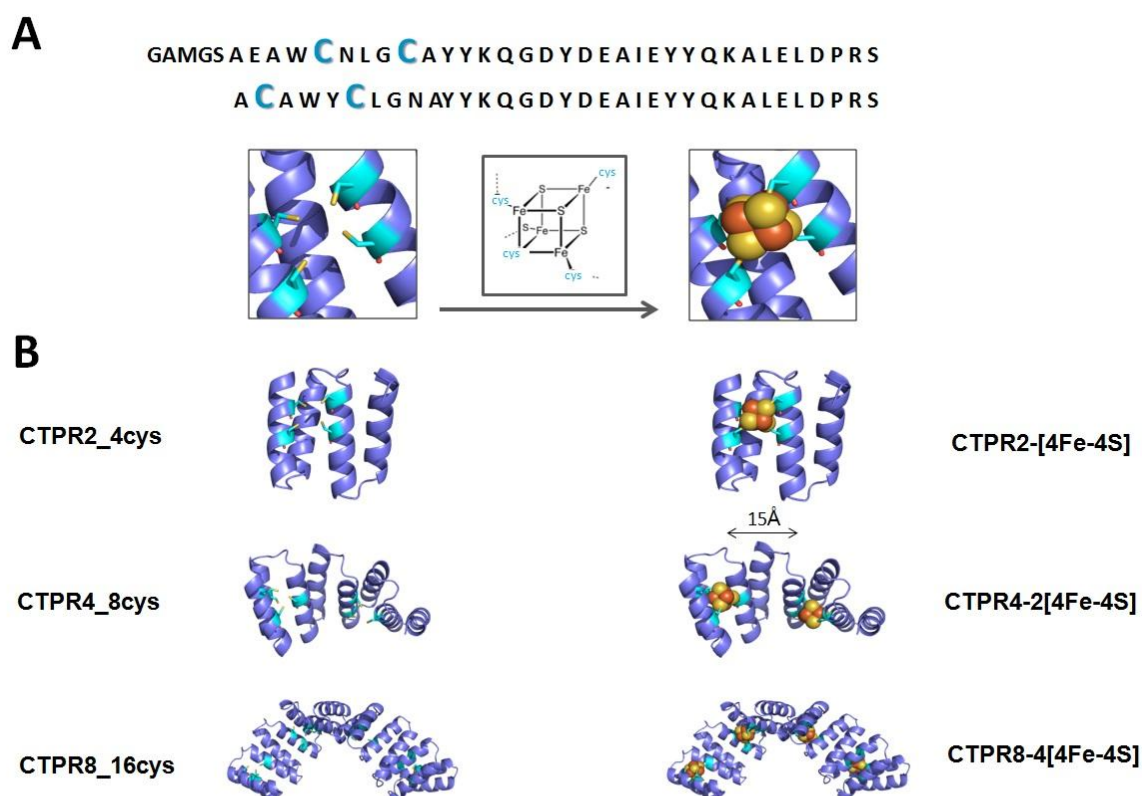


Figure 76. Design of CTPR protein for the generation of [4Fe-4S] cluster. A) Top, the sequence of modified CTPR2 protein in which the mutated positions are highlighted in blue: in CTPR1 Y5C, N9C and in CTPR2 E2C and N6C. Bottom, design of 4-Cys cavity on the left, and the model of the [4Fe-4S] cluster allocated in the cavity, on the right. B) On the left, CTPR proteins with different repeats with different number of 4cys cavities 1, 2 and 4 for CTPR2_4cys, CTPR4_8cys and CTPR8_16 Cys respectively. On the right, model of CTPR proteins with different number of [4Fe-4S] cluster 1,2 and 4 for CTPR2-[4Fe-4S], CTPR4-2[4Fe-4S] and CTPR8-4[4Fe-4S] respectively.

The cluster formation is done using the established in situ synthetic procedure under anaerobic conditions from inorganic precursors and mercaptoethanol [316] (*Materials and Methods M.2.3.2*). Before the cluster formation the His-tag of the protein is cleavage to avoid interactions between the histidines in the Hig-tag and the metal salts that are added for cluster formation (*Materials and Methods M.2.3.1*). The reaction is purified by PD-10 column to remove the excess of metal.

The UV-vis spectrum of the resulting CTPR2-[4Fe-4S] (Figure 77.A/red line), CTPR4-2[4Fe-4S] (Figure 77.B/green line) and CTPR8-4[4Fe-4S] (Figure 77.C/orange line) shows broad absorption peaks centered at 432 and 343 nm, compared with the spectra of the protein without cluster (Figure 77.A.B.C/black line) (*Materials and Methods M.2.3.3*). These peaks are characteristics of charge transfer excitations from sulfur to iron in [4Fe-4S]²⁺ clusters; the absorptions at 432 nm disappears when the cluster is reduced to [4Fe-4S]⁺ state with sodium dithionite in the three cases (Figure 77.A.B.C/blue line). The positions of the maxima are red-shifted compared with inorganic [4Fe-4S] clusters, indicating that the clusters are formed in a hydrophobic environment [316, 324, 325, 327-333]. These spectral features and redox dependent behavior are typical of iron-sulfur cluster proteins [318, 327].

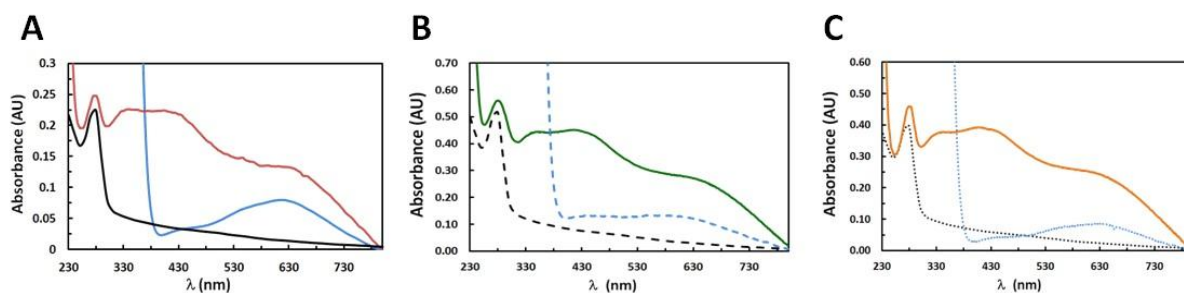


Figure 77. UV-vis characterization. A) UV-vis spectra of the CTPR2_4cys protein in black, CTPR2-[4Fe-4S] in red, and CTPR2-[4Fe-4S] reduced with dithionite in blue. B) UV-vis spectra of the CTPR4_8cys protein in dashed black, CTPR4-2[4Fe-4S] in green, and CTPR4-2[4Fe-4S] reduced with dithionite in dashed blue. C) UV-vis spectra of the CTPR8_16cys protein in pointed black, CTPR8-4[4Fe-4S] in orange, and CTPR8-4[4Fe-4S] reduced with dithionite in pointed blue.

The cluster incorporation is evaluated by the quantification of the iron (using ferrozine assay) and protein concentrations (using Bradford assay) as described in *Materials and Methods M.2.3.4*. The ratio of iron per protein (table 2) agrees with the number of cluster expected per protein: single cluster in CTPR2-[4Fe-4S]; two clusters in CTPR4-2[4Fe-4S]; four cluster in CTPR8-4[4Fe-4S].

Protein	Fe atoms /Protein
CTPR2-[4Fe-4S]	4.55 (± 0.95)
CTPR4-2[4Fe-4S]	8.42 (± 0.73)
CTPR8-4[4Fe-4S]	16.07 (± 0.86)

Table 2. Quantification of iron atoms per protein using Ferrozine method. On the left type of used protein and on the right iron atoms (Fe atoms) per protein with calculated standard deviation.

The effect of the cluster formation on the structure and stability of the protein is analyzed using circular dichroism (CD) (*Materials and Methods M.2.3.5*). The spectra show the same α -helical structure for the protein with and without cluster, meaning that the cluster formation does not disturb the structure of the protein in the three cases CTPR2-[4Fe-4S] (Figure 78.A); CTPR4-2[4Fe-4S] (Figure 78.B); CTPR8-4[4Fe-4S] (Figure 78.C).

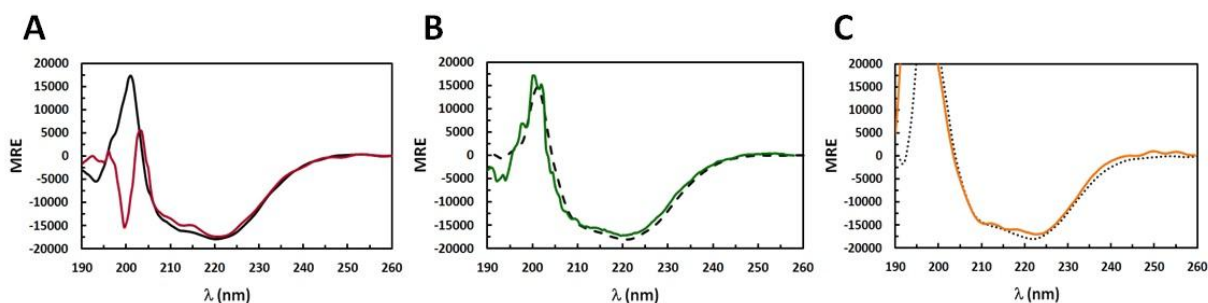


Figure 78. Characterization of the secondary structure. A) CD spectra of the CTPR2_4cys protein in black, and CTPR2-[4Fe-4S] in red. B) CD spectra of the CTPR4_8cys protein in dashed black, and CTPR4-2[4Fe-4S] in green. C) CD spectra of the CTPR8_16cys protein in dotted black, and CTPR8-4[4Fe-4S] in orange.

Furthermore, the stability of the protein with and without the cluster is evaluated performing thermal denaturations (*Materials and Methods M.2.3.5*). Thermal denaturations show that the cluster formation stabilizes the protein, being this effect more evident in the smaller CTPR proteins. The T_m values calculated for the samples with the cluster are larger than for the proteins without the cluster for the three CTPR protein: the T_m for the CTPR2 is 55°C, while the T_m for CTPR2-[4Fe-4S] is 65°C (Figure 79.A); the T_m for the CTPR4 is 73°C, while the T_m for CTPR4-BIS-[4Fe-4S] is 77°C (Figure 79.B); the T_m for the CTPR8 is 75°C, while the T_m for CTPR8-4-[4Fe-4S] is 76°C (Figure 79.C).

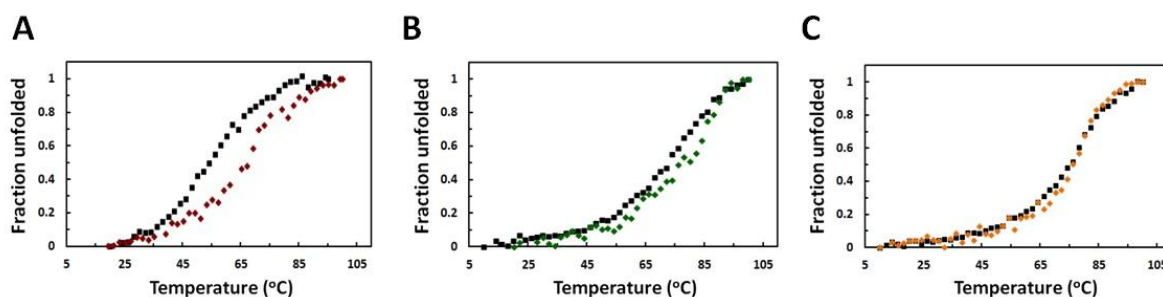


Figure 79. Characterization of the stability. A) Thermal denaturalization of the CTPR2_4cys in black, and CTPR2-[4Fe-4S] in red. B) Thermal denaturalization of the CTPR4_8cys in black and CTPR4-2[4Fe-4S] in green. C) Thermal denaturalization of the CTPR8_16cys in black, and CTPR8-4[4Fe-4S] in orange.

The electronic properties of the proteins with clusters are explored by electron paramagnetic resonance (EPR) spectroscopy (*Materials and Methods M.2.3.6*). As expected, the formation of the iron-sulfur cluster into the designed CTPR protein results in the formation of an EPR silent $[4\text{Fe-4S}]^{2+}$ resting state. After the reduction of the sample by dithionite, the three samples exhibit principal *g-factor* values of 1.89, 1.93 and 2.06 that coincides with the signal of $[4\text{Fe-4S}]^+$ cluster (Figure 80) [334].

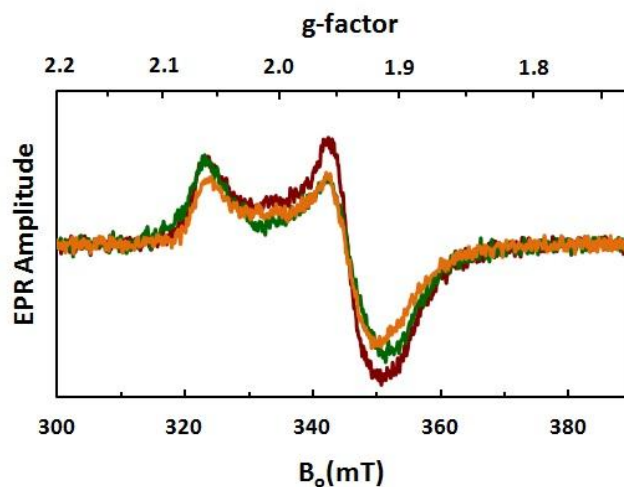


Figure 80. EPR characterization of the $[4\text{Fe-4S}]$ clusters conjugated into CTPRs. EPR spectra of the CTPR2-[4Fe-4S] in red, CTPR4-2[4Fe-4S] in green and CTPR8-4[4Fe-4S] in orange. The three samples are reduced with sodium dithionite. Experimental conditions: microwave frequency 9.336 GHz for CTPR2-[4Fe-4S], 9.333 GHz for CTPR4-2[4Fe-4S] and 9.335 GHz for CTPR8-4[4Fe-4S]; mw power, 2 mW; field modulation amplitude 0.5 mT; temperature 15 K.

Natural ferredoxin proteins containing two $[4\text{Fe-4S}]^{1+}$ within 10–15 Å of each other sometimes exhibit features in the EPR spectra attributable to a spin–spin interaction between the clusters [335, 336]. These features are not observed in the spectra of CTPR4-2[4Fe-4S] and CTPR8-4[4Fe-4S] cluster proteins. However, the lack of cluster interaction effects in the EPR spectra of dicluster ferredoxins is relatively common, for two main

reasons [316, 337]: first, the magnitude of the dipole and exchange interactions between two clusters strongly depends on their relative orientation, because of the mixed valence nature of the clusters [338-340]; the second reason, is the relatively low yield of reduced cluster, specially with concentrated samples that are use to the EPR assay (1mM of cluster).

The redox properties of CTPR2-[4Fe-4S] (Figure 81.A/red line), CTPR4-2[4Fe-4S] (Figure 81.B/green line) and CTPR8-4[4Fe-4S] (Figure 81.C/orange line) are probed using cyclic voltammetry (CV) scanning from 0 to -1 V vs Standard Hydrogen Electrode (SHE) (*Materials and Methods M.2.3.7*). The obtained spectra show peaks that correspond to the anodic (E_a) and cathodic (E_p) waves (Table 3) which are not observed in the scan with the CTPR protein without cluster (Figure 81.A.B.C/black line). For the three CTPR-[4Fe-4S] conjugates is observed a quasi-reversible process. However, the generated redox products in the process tend to be more stable while increases the number of cluster per protein, as the cell potential, E_p , calculated as explained in (*Materials and Methods M.2.3.6*), is more positive. The E_p values in Table 2: -0.242 V for CTPR2-[4Fe-4S], -0.202 V for CTPR4-2[4Fe-4S] and -0.190 V for CTPR8-4[4Fe-4S], show that the charge transfer between the cluster and the electrode is faster for proteins with more clusters. The half wave potential value ($E_{1/2}$) is related to the facility of the system to either loss an electron (oxidation) or gain an electron (reduction) by the [4Fe-4S] cluster. More positive values show greater tendency to reduction while lower values show greater tendency for oxidation. It is estimated a half-cell potential, $E_{1/2}$, of 0.435 V, 0.415 V and 0.320 V for CTPR2-[4Fe-4S], CTPR4-2[4Fe-4S] and CTPR8-4[4Fe-4S] respectively. This difference in the redox potential values probably is due to the difference in the environment between the different clusters. Thus, in CTPR2-[4Fe-4S] the cluster is more exposed to the solvent than in CTPR8-4[4Fe-4S], which can affect to the redox potential value. These values fall within the window expected for low potential [4Fe-4S] clusters in proteins [327, 329, 341, 342].

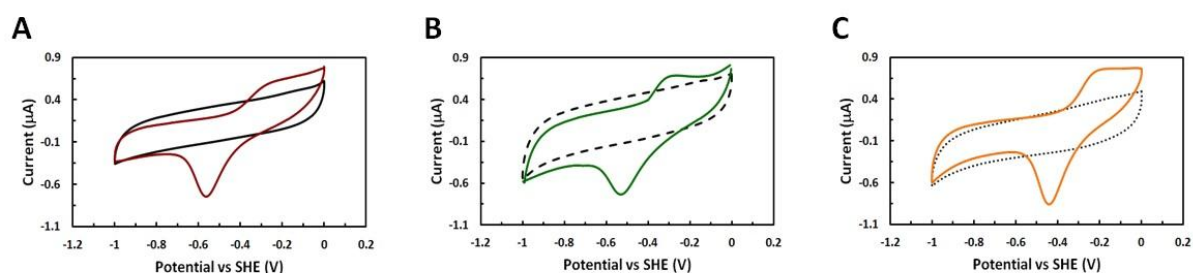


Figure 81. A) Cyclic voltammogram of CTPR2_4Cys in black and CTPR2-[4Fe-4S] in red. B) Cyclic voltammogram of CTPR4_8Cys in dashed black and CTPR4-2[4Fe-4S] in green. C) Cyclic voltammogram of CTPR8_16Cys in dotted black and CTPR8-4[4Fe-4S] in orange. In the three cases the experimental conditions are 100 mM sodium chloride, 100 mM Tris buffer at pH 7.5 at 100 mV/s scan rate with a Ag/AgCl reference electrode, glassy carbon working electrode, and platinum mesh counter electrode.

	E_{pc} (V)	E_{pa} (V)	E_p (V)	$ E_{1/2} $ (V)
CTPR2-[4Fe4S]	-0.314	-0.556	-0.242	0.435
CTPR4-2[4Fe4S]	-0.314	-0.516	-0.202	0.415
CTPR8-4[4Fe4S]	-0.225	-0.415	-0.190	0.320

Table 3. Redox properties of the CTPR-[4Fe-4S] conjugates. From left to right cathodic wave value (E_{pc}), anodic wave value (E_{pa}), cell potential (E_p), and half wave potential ($E_{1/2}$) in volts (V) for CTPR2-[4Fe-4S], CTPR4-2[4Fe-4S] and CTPR8-4[4Fe-4S].

The next step in order to achieve a multifunctional system is to integrate the antenna complex and electron transfer system in a single system. A CTPR2 protein is designed with two different reactivities to allocate a porphyrin to act as antenna complex and one [4Fe-4S] cluster to act as electron transfer system (Figure 75). Based on the CTPR2-4Cys protein a p-Azido-L-phenylalanine non-natural amino acid is introduced in the loop of this modified protein by R33Amber modification to obtain the CTPR2_CA protein (*Materials and Methods M.2.3.1*). The incorporation of the non-natural amino acid is done using the amber stop codon suppression methodology using an evolved mutant tRNA/aaRS pair from *Methanococcus jannaschii* developed by Schultz and coworkers [343, 344].

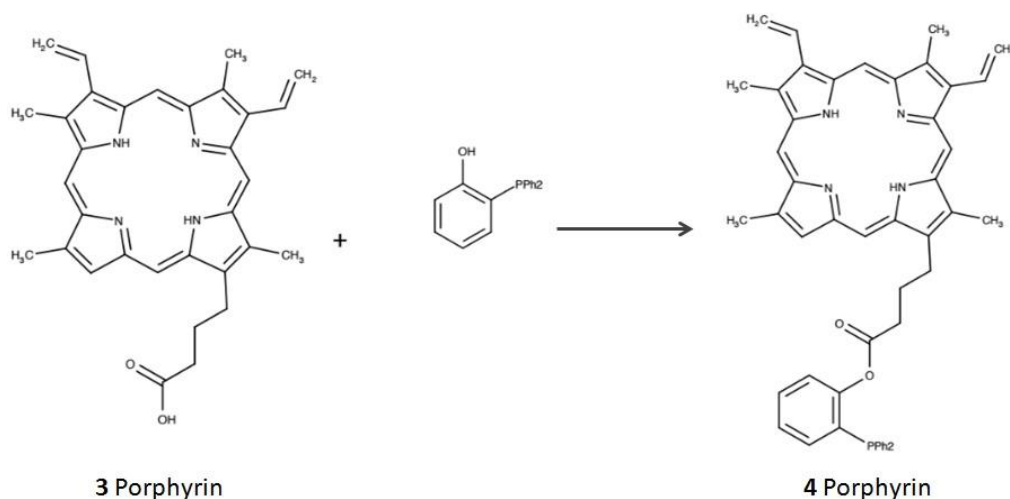


Figure 82. Preparation of a porphyrin with a phosphine group. On the left, **3** porphyrin, with a reactive carboxyl group exposed and (2-Hydroxyphenyl)diphenylphosphine with hydroxyl group exposed. In the right, **4** porphyrin with an exposed phosphine group as a results of esterification of carboxyl acids in **3** with the hydroxyl group of (2-Hydroxyphenyl) diphenylphosphine

The conjugation between the porphyrin and the protein is done by Staudinger ligation. So, it is necessary synthesized a porphyrin with an exposed phosphine group that reacts with the azide in the protein. As starting point a porphyrin **3** is used, synthesized by Dr. Ghirlanda's group (Arizona State University), that contains an exposed carboxyl group. The esterification of the carboxyl group of **3** is used to introduce the phosphine group in the porphyrin as explained in *Materials and Methods M.2.3.8* to form **4** (Figure 82).

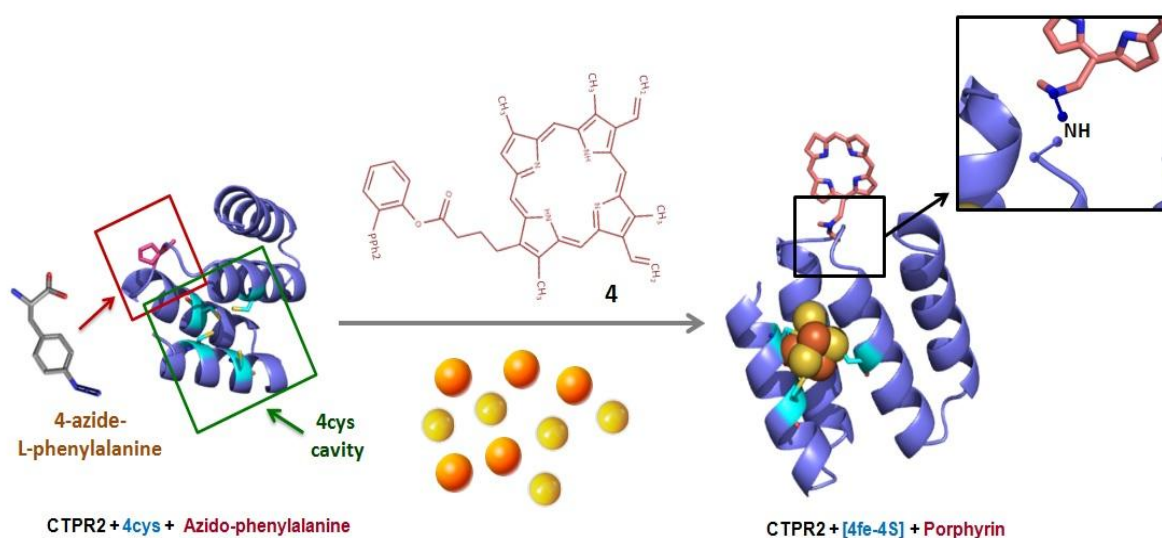


Figure 83. Strategy for the incorporation of antenna complex and electron transfer system in the CTPR2_CA. On the left, CTPR2 protein modified by 4-azide-L-phenylalananine in red and 4Cys cavity in light blue to obtain CTPR2_CA protein. On the right, a model of a CTPR2_CA protein with a porphyrin, as antenna complex, and [4Fe-4S] cluster as electron transfer system.

The ligation between the CTPR2_CA protein and the **4** is performed as explained in *Materials and Methods M.2.3.9* (Figure 83). The resulting purified conjugate is characterized using MALDI-TOF to confirm the porphyrin incorporation (Figure 84.A). The mass spectra shown clearly the incorporation of the porphyrin in the protein: the peak 11272 Da (Figure 84.A/black line) correspond to the CTR2_CA protein (the mass based on sequence is 11280 Da) and the peak of 11873 Da (Figure 84.A/red line) corresponds to the CTPR2-porphyrin conjugate (the mass calculated for the designed conjugate is 11729 Da). Moreover, in the UV-vis spectrum of the conjugate shows a peak that corresponds to the protein and the peak that corresponds to the porphyrin (Figure 84.B) (*Materials and Methods M.2.3.3*)

In parallel, it is performed the formation of the [4Fe-4S] cluster in the CTPR2_CA protein using the same protocol that is used for the formation of the cluster in the CTPR2_4cys, CTPR4_8cys and CTPR8_16cys (Figure 83) (*Materials and Methods M.2.3.2*). The UV-vis spectrum (*Materials and Methods M.2.3.3*) of the resulting CTPR2-[4Fe-4S] (Figure 84.C/red line) shows the typical broad absorptions for the [4Fe-4S]²⁺ state at 432

and 343 nm, compared with the spectra of the CTPR2_CA protein without cluster (Figure 84.C/black line) and the peak at 432 nm disappears when the cluster is reduced to $[4\text{Fe-4S}]^+$ state with sodium dithionite (Figure 84.C/blue line). These results confirm that the modified CTPR2_CA protein is well designed to allocate the elements, antenna system and electron transfer system, in the same protein.

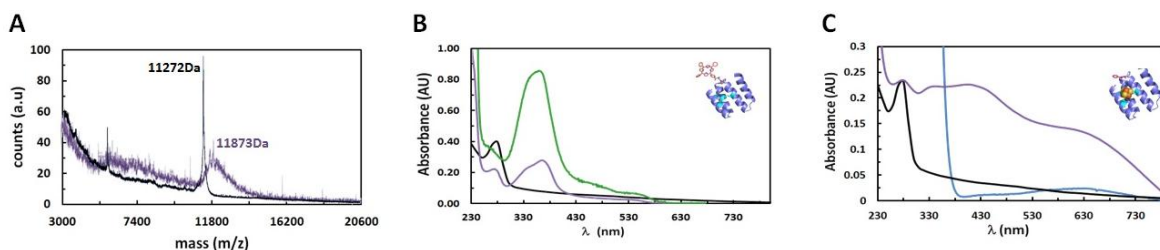


Figure 84. Characterization of the CTPR2-porphyrin conjugate. A) Mass spectra of the CTPR2-CA in black and CTPR2_CA-porphyrin conjugate in purple. B) UV-vis of the synthesized porphyrin in green, CTPR2_CA protein in black and CTPR2_CA-porphyrin conjugate in purple. C) UV-vis of the CTPR2_CA protein in black, CTPR2_CA-[4Fe-4S] in purple and CTPR2_CA-[4Fe-4S] reduced by dithionite in blue.

In this section we establish the basis to create a photosystem where the elements are integrated in a single system and works under the same conditions. The final step that includes the conjugation of both elements in the same system is already in progress.

2.4.3.3. Conclusions

In conclusion, we explore the use of CTPR proteins to mimic natural photosynthetic H_2 production. On one hand, we designed a system to generate arrays of redox active cluster using CTPR proteins, in which we can control the position of the different clusters in the protein. We confirmed using different techniques (UV-vis, EPR...) the correct formation of the cluster and confirmed that the clusters obtained are redox active clusters. On the other hand, we designed a CTPR2_CA protein to form a multifunctional system in which we form the antenna complex, by the introduction of a porphyrin and an electron transfer system, by the formation of the cluster in the CTPR2_CA protein. In this section, we set the basis to form a single integrated photosystem using CTPR proteins in which all the elements are active under the same conditions, a key feature to obtain efficient systems.

2.5. Chapter 2: Conclusions

2.5. Chapter 2. Conclusions

In this chapter, we show how it is possible to use proteins to get functional bio-hybrid structures. First of all, the deep understanding of the structure and stability of the CTPR proteins makes possible to use these proteins to organize elements at nanoscale by their arrangement onto the CTPR structure with atomic resolution. Furthermore, the control over the assembly, explored in the Chapter 1, allows to create functional structures where the element are organized at different length scales: nano, micro and macroscale. This aspect is very important for technological advances, for which is required a method to control the structure on the molecular and mesoscopic scales.

We explore the use of CTPR proteins as tools to organize gold nanoparticles, donor-acceptor molecules, carbon nanotubes, and metallic clusters. We form functional structures with specific characteristics that depend on the organization of the active elements in the system and that are useful in fields such as nanoelectronics, photovoltaics and energetic. In this study we establish the basis and protocols for the formation of functional structures using CTPR proteins as a template to organize different active elements that are useful in broad fields of nanotechnology.

General Conclusions

General conclusions

This thesis presents several examples on how it is possible to use proteins as tools for the generation of nano-devices and materials for different nanotechnological applications. The precise synthesis of nano-devices and materials with tailored complex structures and properties requires methods to control of structure of the materials at different length scales. However, nowadays, most of the methods for the generation of ordered materials lack the precision to determine their structure and properties in a control manner. In this sense, Nature has a lot to teach us. In Nature, there are materials with very different properties that arise from the organization of the composing elements in the material. Specifically, many of the most sophisticated materials in Nature are made by self-assembly of proteins. Therefore, we explore the possibility of using proteins as building blocks to generate structures in which active elements are organized at different length scales to provide specific properties to the material.

➤ The results obtained in the current thesis lead to the following main conclusions:

1. CTPR proteins can be rationally designed in order to control their self-assembly for the generation of different protein-based structures including nanofibers, protein monolayers, and protein nanotubes.
2. CTPR protein scaffolds can be functionalized by the introduction of reactive groups in the protein with atomic precision without disrupting the protein structure. These reactive moieties are then used to conjugate active elements, thus generating defined arrangements of the components within the structure. CTPR proteins have been used to organize different active elements including gold nanoparticles, donor-acceptor molecules and metallic clusters.
3. The CTPR proteins can be functionalized by several reactivities at the same time. This feature allows the conjugation of different active elements in the same CTPR protein thus generating multifunctional hybrid structures. As example, a CTPR that integrates an antenna system composed by a porphyrin, and an electron transfer system composed by metallic clusters has been generated.
4. Biohybrid structures based on proteins can be obtained by the combination of the control over the assembly properties and the functionalization of the repeated scaffolds. The active components are organized at nano, micro and macroscale in the biohybrid materials. The defined organization of the elements gives specific properties to the material in terms of conductivity, photo-activity or electrochemical activity for the different examples tested.

5. Repeat proteins have been demonstrated to be robust scaffolds for fabrication of complex hybrid nanostructures and materials.
6. The templating approaches developed can be expanded to other classes of active elements. Thus, this thesis opens the door to generate hybrid nanostructures with applications in different fields of nanotechnology.
7. The methodology established in this thesis can be translated to other protein systems that present different structural and self-assembly features in order to obtain functional structures with different properties.

Conclusiones Generales

Conclusiones generales

En esta tesis se presentan algunos ejemplos de cómo es posible utilizar proteínas para generar nano-dispositivos y materiales funcionales con diferentes aplicaciones nanotecnológicas. La generación de estos nano-dispositivos y materiales funcionales requiere el control preciso de los elementos funcionales que lo componen en diferentes escalas de tamaño. Sin embargo, hoy en día, los métodos para generar este tipo de materiales ordenados carecen de la precisión necesaria para determinar su estructura y propiedades de una forma controlada. En este sentido, la naturaleza tiene mucho que enseñarnos. En la naturaleza existen materiales con características muy diferentes cuyas propiedades vienen de la organización de los elementos que componen el material. Especialmente, los materiales más sofisticados en la naturaleza están hechos mediante el auto-ensamblaje de proteínas. Es por ello, que en esta tesis se usan las proteínas como bloque de construcción para generar estructuras en el que elementos activos están organizados en diferentes escalas de tamaños. Las propiedades de los materiales generados dependen de la organización de dichos elementos.

➤ De los resultados de esta tesis obtenemos las siguiente conclusiones:

1. Las proteínas CTPR se pueden diseñar de forma racional de tal forma que podemos controlar el auto-ensamblaje de estas para generar estructuras supramoleculares como nanofibras, monocapas o nanotubos de proteínas.
2. Las proteínas CTPR se pueden funcionalizar para introducir grupos reactivos a lo largo de las proteínas con precisión atómica sin perder la estructura de estas. Utilizando estos grupos reactivos, se conjugan diferentes elementos activos a lo largo de la proteína de forma controlada. En esta tesis, las CTPR se utilizan para organizar partículas de oro, moléculas donoras-aceptoras y clústeres metálicos.
3. Se pueden introducir diferentes reactividades en una misma proteína CTPR. Esto nos permite conjugar diferentes elementos activos en la misma proteína y de esta forma generar estructuras híbridas multifuncionales. Como ejemplo, generamos una CTPR que integra un sistema de antena de captación de luz, compuesto por una porfirina, y una cadena transportadora de electrones, compuesto por clústeres metálicos.
4. Podemos obtener estructuras funcionales biohíbridas basadas en proteínas combinando el control en el auto-ensamblaje y el control en la funcionalización de las proteínas CTPR. De esta forma, podemos organizar elementos activos en diferentes escalas de tamaño (nano, micro y macroescala). Las propiedades tales como conductividad, foto-actividad o actividad electroquímica de los materiales biohíbridos generados proviene de la organización de los elementos activos en dicho material.

5. En esta tesis hemos demostrado que las proteínas de repetición son un bloque de construcción versátil y robusto para la generación de nanoestructuras híbridas complejas y materiales funcionales.
6. Las estrategias desarrolladas en la tesis para conjugar partículas de oro, moléculas donoras-aceptoras y clústeres netálicos se pueden expandir para conjugar diferentes elementos activos a las proteínas CTPR. Por tanto, en esta tesis se han desarrollado las bases para generar nanoestructuras híbridas con aplicaciones en diferentes campos nanotecnológicos.
7. La metodología desarrollada en esta tesis, se puede trasladar a otros sistemas de proteínas que presenten diferentes características estructurales y de auto-ensamblaje para obtener estructuras funcionales con distintas propiedades.

Materials and Methods

Materials and Methods

Basic consensus CTPR1WT is use as starting sequence to introduce all the mutations of all the designed CTPR mutants. CTPR1WT protein DNA sequence:

```
ATGTCGTACTACCATCACCATCACCATCACGATTACGATATCCCAACGACCGAAAACCTGTA
TTTTCAGGGCGCCATGGGATCCGCTGAGGCATGGTACAACCTGGGTAACGCTTACTACAA
ACAG GGTG ACTACGACGA AGCTATCGAATACTACCAGAAGGCTCTCGAGCTGGACCCGA
GATCCG CTGAAGCTAAACAGAACCTGGGT AACGCTAAACAGAAACAGGGTTGA
```

CTPR1WT consensus protein amino acid sequence:

```
MSYYHHHHHDYDIPTTENLYFQGAMGSAEAWYNLGNAYYKQG
DYDEAIEYYQKALELDP RSAEAKQNLGN AKQKQG*
```

M.1. Chapter 1. CTPR protein design for protein self-assembly

M.1.1. Controlled nanometric fibers of self-assembled CTPR

M.1.1.1. Protein design and purification

Based on consensus CTPR20 protein cloned into pPro-EXHTa [76] two single Cys residues at each end of the protein and stop codon to remove the solvating helix are introduced by quick-change site directed mutagenesis using the following oligonucleotide sequences:

Cys-N-terminal_forward: CTGTATTTTCAGGGCTGTGCCATGGGATCCGC

Cys N-terminal_reverse: GCGGATCCCATGGCACAGCCCTGAAAATACAG

Cys-C-terminal_forward: CTGGACCCGAGATCTTGTTGAGCTAAAGC

Cys-C-terminal_reverse: GCTTCAGCTCAACAAGATCTCGGGTCCAG

measurements. The polymerization is initiated immediately by placing the protein in the DLS instrument at constant temperature (25°C or 42°C).

M.1.1.3. Dynamic light scattering measurements (DLS)

To monitor the increase in size upon protein polymerization by DLS is used a Zetasizer NanoZS instrument (Malvern Instruments, Malvern, UK), which measures scattering in small volumes at low concentration (from 0.1 mg/mL) and sizes from 0.3 nm (diameter). An incident light wavelength of 532 nm is used and the scattered light is collected at a fix angle of 173°, which is optimal for low sample concentration.

The distributions of the hydrodynamic sizes in the sample are collected periodically from time 0 at the different starting protein concentrations and polymerization temperatures. Three independent measurements are taken at each data point to calculate the error.

M.1.1.4. Transmission electron microscopy (TEM)

Monomeric sample are deposited at 1 μM and polymerized CTPR protein samples at 5 μM protein concentration are deposited on glow discharged Cu/Rh grids coated with carbon, and negatively stained with 2% uranyl acetate. Micrographs are recorded using Kodak SO-163 film, in a JEOL JEM1200EXII electron microscope with a tungsten filament operated at 100 kV and at 60 K magnifications. The particle sizes are the average size of 5 molecules measured using ImageJ program in order to obtain SD. The polymeric samples are reduced in the presence of 1 mM DTT during 10 minutes at room temperature and are deposited using the same protocol.

M.1.1.5. Hydrodynamic size calculations of CTPR proteins

The experimental diffusion coefficients, and therefore, hydrodynamic sizes of CTPR proteins with different number of repeats (2-20) are previously calculated by fluorescence correlation spectroscopy (FCS) [127]. The theoretical diffusion coefficients and the hydrodynamic radius for CTPR proteins of different lengths are calculated from the x-ray crystal structure coordinates [70, 76] using the program Hydropro version 7c19 <http://Leonardo.fcu.um.es/macromol/programs/hydropro/hydropro.htm> [128].

The correlation between number of repeats and the hydrodynamic size can be fitted to the equation $R_h = R_0 N^v$, where R_h is the hydrodynamic radius and N the number of repeats per protein.

absorption using a Cary 50 UV–vis spectrophotometer (Varian) at 280 nm using the extinction coefficient calculated from the amino acid composition [345] ($294760 \text{ M}^{-1}\text{cm}^{-1}$).

M.1.2.2. Protein reduction for gold immobilization experiments

Prior to any immobilization experiment on gold, purified CTPR20-Cys at a protein concentration of about $23 \mu\text{M}$ is freshly reduced with 1 mM dithiothreitol (DTT) during 20 min to ensure full reduction of the sample. DTT is removed by buffer exchange over a NAP-5 column (GE Healthcare Life Science) in PBS buffer with lower salt and buffer concentration (10 mM NaCl, 10 mM phosphate buffer pH 7.4) or water. Protein fractions without DTT are collected at 4°C and their concentration measured by UV absorption spectroscopy.

M.1.2.3. Quartz crystal microbalance (QCM)

QCM is an extremely sensitive mass sensor capable of providing real-time monitoring of immobilization events in the nanogram range and determination of kinetic rate constant. QCM measures frequency changes in the resonant frequency of the quartz crystal according to Sauerbrey's equation (Equation (8)), where a decrease in mass corresponds to an increase in frequency and vice versa. Moreover, QCM with motional resistance monitoring also allows obtaining information related to the structure, conformation, and interactions of the molecules on the surface through the acoustic dissipation. In situ mass changes are measured with a SRS QCM200 Quartz Crystal Microbalance from SRS Instruments (Sunnyvale, CA., USA) with AT-cut quartz crystals (5 MHz) of 25 mm diameter with gold electrodes deposited over a Chromium adhesion layer. An asymmetric electrode format is used with the side having the larger gold area facing the solution. The electroactive working area (front side) is 1.370 cm^2 and the piezoelectric area (backside) was 0.317 cm^2 .

Prior to use, the quartz crystals are cleaned by exposure to "piranha" solution (3:1 concentrated $\text{H}_2\text{SO}_4/30\% \text{H}_2\text{O}_2$) followed by exhaustive rinsing with distilled water and a final rinse with ethanol/water (2:1). *Caution: piranha solution reacts violently with most organic materials and must be handled with extreme care.* The quartz crystal resonator is set in a probe made of Teflon and immersed in water-jacketed beaker thermostated at 25°C . The crystals sensitivity is $17.68 \cdot 10^{-9} \text{ g Hz}^{-1}\text{cm}^{-2}$ [346-348].

Under constant flow protein samples in PBS buffer at concentrations from 5 nM to $2.5 \mu\text{M}$ are injected in the system and both the changes in motional frequency and in resistance are recorded until saturation is reached.

M.1.2.4. Surface Plasmon resonance (SPR) measurements

A gold film of approximately 30 nm is first deposited on sodalime glass substrate (1 mm thick) by electron beam evaporation in a home-based deposition chamber with a base pressure of $P_b = 1 \times 10^{-8}$ Torr. The SPR spectrum of the gold surface is measured to calculate the exact thickness of the film. Then, a CTPR20-Cys protein solution at 2.5 μM concentration is deposited on the gold surface. After incubation for 30 min at room temperature the SPR spectrum of the protein film in solution is acquired.

SPR spectra are acquired using the Kretschmann–Raether configuration [349, 350], with a home-made device described elsewhere [147]. SPR is excited with a 632.8 nm laser. The angular response of the photodiode is corrected as described before [148]. For each sample 4 scans are recorded from different spatial locations. The spectra shown here correspond to the average (with the line thickness being the standard deviation). Any possible drift in the motor position is corrected by fixing the position of the critical angle to 39.5° ; this value depends only on the glass and the surrounded air, and it is independent of the gold and protein films. To calculate the total mass of protein immobilized on the gold surface, a shift of 0.1° on the resonance angle is related to a mass of 1 ng mm^{-2} , as described before for proteins [351]. SPR curve simulation is carried out using Winspill freeware by RES-TEC (<http://www.res-tec.de/applications.html>) including the correction of refraction for triangular prisms.

M.1.2.5. Atomic force microscopy (AFM)

For AFM measurements CTPR20-Cys protein was immobilized on gold surfaces. The gold substrates were commercial gold-coated borosilicate glass substrates (Gold Arrandee™) annealed using a Bunsen burner prior to use. First, we monitored gold surfaces under different annealing conditions without protein to achieve sub-nanometric roughness and ensure that the surface roughness is small enough to clearly image the nanometric-scale protein molecules. CTPR20-Cys freshly reduced protein as described above is incubated on the gold surface during 1 h. After incubation the surface is thoroughly washed with water. The protein coated surface is imaged in solution using a Nanotec Cervantes system in jumping mode [352]. Olympus standard silicon nitride probes of 0.05 N m^{-1} and 18 kHz (OMCL-RC800PSA) are employed. Images are processed using the WSxM software (Nanotech Electronica, Madrid, Spain <http://www.nanotec.es/>) [353].

M.1.2.6. Profilometry measurements

Measurements CTPR20-Cys protein is immobilized on gold surfaces using the same protocol used for the generation of the AFM samples. Three different layers are obtained by incubation of 2.5 μ M protein. The layer thickness is measured using a “Dektak XT” mechanical profilometer with a 2.5 μ m radio stylus and 1 mg force (the minimal force).

M.1.3. CTPR nanotubes: 3D structure

M.1.3.1. Protein design and purification

Based on a consensus CTPR6 protein, two leucine residues are introduced in each CTPR repeat at positions 15 and 31. The mutations are introduced in CTPR1WT by quick-change site directed mutagenesis using the following oligonucleotide sequences:

G15L_forward: CTTACTACAAACAGCTTGACTACGACGAAGC

G15L_reverse: GCTTCGTCGTAGTCAAGCTGTTTGTAGTAAG

G31L_forward: GGCTCTCGAGCTGCTCCCAAGATCCGCTGAG

G31L_reverse: CTCAGCGGATCTTGGGAGCAGCTCGAGAGCC

The C6L gene is generated from the CTPR1 mutated gene by sequential additions of identical mutated repeats and cloned into pPro-EXHTa vector. The sequence of the obtained C6L protein is:

```

M S Y Y H H H H H D Y D I P T T E N L Y F Q G A M G S A E A W Y N
L G N A Y Y K Q L D Y D E A I E Y Y Q K A L E L L P R S A E A W Y N
L G N A Y Y K Q L D Y D E A I E Y Y Q K A L E L L P R S A E A W Y N
L G N A Y Y K Q L D Y D E A I E Y Y Q K A L E L L P R S A E A W Y N
L G N A Y Y K Q L D Y D E A I E Y Y Q K A L E L L P R S A E A W Y N
L G N A Y Y K Q L D Y D E A I E Y Y Q K A L E L L P R S A E A W Y N
L G N A Y Y K Q L D Y D E A I E Y Y Q K A L E L L P R S A E A W Y N
L G N A K Q K Q G

```

The protein is expressed as His-tagged fusion and purified using standard affinity chromatography methods based on previously published protocols [71] for His-tagged CTPR proteins using 0.2% NP-40 detergent and, depending on the sample, 0.2% SDS in the lysis buffer. The protein is dialyzed into PBS buffer (150 mM NaCl, 50 mM phosphate buffer pH 7.4) and stored frozen at -20°C. The protein concentration is determined by UV-

absorbance at 280 nm using the extinction coefficient calculated from the amino acid composition [345] ($92600 \text{ M}^{-1} \text{ cm}^{-1}$).

The C6L_2 gene is generated from the CTPR1 mutated gene by sequential additions of CTPR1 modified repeats, where the repeats 1 and 2 have a mutation in D15L; the repeats 3 and 4 have a mutation in G15L and D31L; the repeats 5 and 6 have a mutation in G31L. The sequence of the obtained C6L_2 protein is:

```
M S Y Y H H H H H D Y D I P T T E N L Y F Q G A M G S A E A W Y N
L G N A Y Y K Q G D Y D E A I E Y Y Q K A L E L L P R S A E A W Y N
L G N A Y Y K Q G D Y D E A I E Y Y Q K A L E L L P R S A E A W Y N
L G N A Y Y K Q L D Y D E A I E Y Y Q K A L E L L P R S A E A W Y N
L G N A Y Y K Q L D Y D E A I E Y Y Q K A L E L L P R S A E A W Y N
L G N A Y Y K Q L D Y D E A I E Y Y Q K A L E L D P R S A E A W Y N
L G N A Y Y K Q L D Y D E A I E Y Y Q K A L E L D P R S A E A K Q N
L G N A K Q K Q G
```

The protein is expressed as His-tagged fusion and purified using standard affinity chromatography methods based on previously published protocols [71] for His-tagged CTPR proteins. The protein is dialyzed into PBS buffer (150 mM NaCl, 50 mM phosphate buffer pH 7.4) and stored frozen at -20°C . The protein concentration is determined by UV-absorbance at 280 nm using the extinction coefficient calculated from the amino acid composition [345] ($92600 \text{ M}^{-1} \text{ cm}^{-1}$).

M.1.3.2. Protein design and purification

Standard 15% acrylamide SDS-PAGE electrophoresis gels are used to run the C6L and C6L_2 purified samples to see the dimeric and monomeric percentage in the samples. The gels are stained with Coomassie blue.

The native gel, is done using a 15% acrylamide gel and the sample loading buffer is made with the same recipe as standard SDS-PAGE gel, but without adding the SDS. The gels are stained with Coomassie blue.

The 15% acrylamide gel with 0.2% SDS gel and sample loading buffer is done using the same recipe as standard SDS-PAGE gel, but adding only 0.2% of SDS.

M.1.3.3. Size exclusion chromatography

To separate the different species in the sample gel filtration chromatography was performed using an AKTA prime plus Fast Protein Liquid Chromatography (FPLC) equipment (GE Healthcare). The purified sample is injected into a Superdex 75 HR 10/30 size exclusion chromatography column (GE Healthcare) and run at 0.5 mL/min in PBS at 4°C or PBS with 0.2% SDS at room temperature, depending on the sample. The purified samples are collected in 0.5 mL fractions and stored protected from light.

M.1.3.4. MALDI-TOFF mass spectroscopy

The samples are analyzed using ABI 4800 MALDI TOF/TOF mass spectrometer. 4-Hydroxy-3-5-dimethoxycinnamic acid is used as matrix at 10 mg/ml dissolved in a 70% acetonitrile and 0,1% TFA solution. The samples are prepared at a sample to matrix solution ratio of 1:4 (v/v) and 1µl of the mixture deposited on the sample plate. When the sample dried, 1 µl extra of matrix is deposited over the sample.

M.1.3.5. Dynamic light scattering measurements (DLS)

The hydrodynamic radius of the different samples is measured by DLS in a Zetasizer NanoZS instrument (Malvern Instruments, Malvern, UK), which measures scattering in small volumes at low concentration (from 0.1 mg/mL) and sizes from 0.3 nm (diameter). An incident light wavelength of 532 nm is used and the scattered light is collected at a fix angle of 173°, which is optimal for low sample concentration.

M.1.3.6. Hydrodynamic size calculations

The hydrodynamic radius of the monomer based on the CTPR8 PDB (ID: 2AVP) and of the dimer based on the design are calculated using the program Hydropro version 7c19 <http://Leonardo.fcu.um.es/macromol/programs/hydropro/hydropro.htm> [128].

M.1.3.7. Transmission electron microscopy (TEM)

The samples in PBS or PBS+0.2% SDS are deposited at 0.8 µM protein concentration on glow discharged Cu/Rh grids coated with carbon, and negatively stained with 2% uranyl acetate. Micrographs are recorded using Kodak SO-163 film, in a JEOL JEM1200EXII electron microscope with a tungsten filament operated at 100 kV and at 60 K magnifications.

M.1.3.8. TEM image processing, particle selection and 2D classification

In all cases, the CTF (contrast transfer function) is corrected using the CTFFIND3 program [354], which also calculates potential astigmatism. Micrographs with visible drift and astigmatism are discarded. Single particles are selected manually, extracted from micrographs and normalized using the XMIPP software package [355]. Two types of algorithms implemented in XMIPP are used to classify single images: ML2D [356] and CL2D [357]. In the main text CL2D classification is presented.

For the theoretical projections of the C6L dimer based on the designed model, EMAN software [358] is used.

M.1.3.9. Circular dichroism (CD) measurements

CD spectra are measured using a Jasco J-815 CD Spectrometer. CD spectra the samples in PBS or PBS+0.2% SDS buffer are acquired in a 1 cm path length quartz cuvette. All the CD spectra are recorded with a band-width of 1 nm at 1 nm increments and 10 second average time.

Thermal denaturalization is performed in a 0.1 cm path length quartz cuvette in PBS or PBS+0.2% SDS sample. The curves are monitored by following the CD signal at 222 nm wave length as a function of temperature from 10°C to 95°C.

M.1.3.10. Docking of the C6L protein using Cluspro program

Using Cluspro [191-194] docking is performed between C6L monomer-C6L monomer. The obtained structures are used as an input in the next round of docking where the most probable dimers are docked in pairs using Cluspro, obtaining different possible tetramers as results. The obtained tetramers are used as input in the next docking round where the most probable tetramers are docked with C6L dimers (obtained in the second round) and C6L tetramers obtained in the third round. In the Figure 34 some of the most stable results are represented.

M.2. Chapter 2. CTPR protein based ordered bio-hybrid functional structures

M.2.1. Templating gold nanoparticles by CTPR proteins

M.2.1.a. Performing AuNPs monolayers templated by CTPR proteins

M.2.1.a.1. 12 nm diameter AuNPs synthesis

AuNPs are synthesized following the well-established Turkevich protocol [359] with slight adjustments as described previously [360]. 100 mL of an aqueous solution of HAuCl₄ (Sigma–Aldrich) at 1 mM is refluxed at 140°C and reduced for 15 min with 10 mL of an aqueous solution of 55 mM sodium citrate. The reaction lead rapidly to a red-wine solution of citrate based AuNPs. The sample is then cooled to room temperature and filter through 0.2 μm pore-size filter to eliminate large aggregates. The final concentration of AuNPs (9 nM) is determined by UV–vis spectrometry using an extinction coefficient at 520 nm, $\epsilon = 2.7 \times 10^8 \text{ M}^{-1}\text{cm}^{-1}$. The as-modified Turkevich protocol for the synthesis of AuNPs leads to a well-dispersed colloidal suspension of citrate-stabilized AuNPs with an average diameter of 12.4 ± 1.2 nm (measured by TEM). A second step of ligand exchange with thioctic acid is carried out to introduce a more stable surfactant with carboxylic groups. To 10 mL of AuNPs at 9 nM is added 1.15 mL of thioctic acid at 10 mM in ethanol to react overnight at room temperature. Then, sample is centrifuged 40 min at $10,000 \times g$ and supernatant discarded. The red-wine pellet is redispersed in 7 mL milliQ water, resulting in a light purple suspension. Then, drops of NaOH 10 mM are added until the solution turned to red-wine color, indicating a good colloidal dispersion of thioctic acid stabilized AuNPs (TA-GNPs) due to negative charges of the carboxylic acids.

M.1.1.a.2. AuNPs conjugation with CTPR20-Cys immobilized onto gold surface

The conjugation between TA-AuNPs and CTPR20-Cys, previously immobilized onto gold surface by the C-terminal Cys residue, is carried out through the reaction between carboxylic acid functions of the AuNPs ligands and the free exposed terminal amine of the protein. First, 9 μL of ethyl(dimethylaminopropyl) carbodiimide (EDC, 1 mM) and 6 μL of N-hydroxysuccinimide (NHS, 1 mM) are added to 1 mL of TA-AuNPs at 30 nM for 10 min to perform the activation of the carboxylic groups. Then, 200 μL of activated TA-AuNPs are incubated onto the gold surface covered by CTPR20-Cys during 3 days at room temperature. Controls of the same experiment without adding EDC/NHS coupling agents and the immobilization of TA-AuNPs onto protein-free gold surface are also carried out under the same experimental conditions. After the incubation time, the liquid phase is

removed from the gold surface and the surface was washed several times with MilliQ water and dried under nitrogen flux.

M.1.1.a.3. Scanning electron microscopy (SEM)

SEM measurements of the surfaces with gold nanoparticles are carried out at low acceleration voltages with an ultra high resolution Scanning Electron Microscope (SEM), Auriga Cross Beam FIB–SEM from Carl Zeiss equipped with a Gemini electron beam column and in-lens detector technology, at IMDEA Nanociencia. The samples are deposited on gold substrates and imaged using a 2–3 kV beam acceleration voltages.

M.2.1.b. Performing AuNPs conductive films template by CTPR proteins

M.2.1.b.1. protein expression and purification

Based on consensus CTPR16 protein, four cysteine residues are introduced in a loop position 33 of the CTPR repeats 2, 6, 10 and 14 to form CTPR₁₆-4cys protein. The mutation is introduced in CTPR1WT by quick-change site directed mutagenesis using the following oligonucleotide sequences:

R33C_forward: CTCGAGCTGGACCCGTGCTCCGCTGAGGCATGG

R33C_reverse: CCATGCCTCAGCGGAGCACGGGTCCAGCTCGAG

The CTPR₁₆-4Cys gene is generated from the CTPR1 gene by sequential additions of CTPR1WT or mutated repeats, depending on the CTPR repetition number, and cloned into pPro-EXHTa vector. The sequence of the obtained CTPR₁₆-4Cys protein is:

```

M S Y Y H H H H H D Y D I P T T E N L Y F Q G A M G S A E A W Y N
L G N A Y Y K Q G D Y D E A I E Y Y Q K A L E L D P R S A E A W Y N
L G N A Y Y K Q G D Y D E A I E Y Y Q K A L E L C P R S A E A W Y N
L G N A Y Y K Q G D Y D E A I E Y Y Q K A L E L D P R S A E A W Y N
L G N A Y Y K Q G D Y D E A I E Y Y Q K A L E D P R S A E A W Y N L
G N A Y Y K Q G D Y D E A I E Y Y Q K A L E L D P R S A E A W Y N L G
N A Y Y K Q G D Y D E A I E Y Y Q K A L E L C P R S A E A W Y N L G
N A Y Y K Q G D Y D E A I E Y Y Q K A L E L D P R S A E A W Y N L G
N A Y Y K Q G D Y D E A I E Y Y Q K A L E L D P R S A E A W Y N L G N A
Y Y K Q G D Y D E A I E Y Y Q K A L E L D P R S A E A W Y N L G N A Y
Y K Q G D Y D E A I E Y Y Q K A L E L C P R S A E A W Y N L G N A Y Y
K Q G D Y D E A I E Y Y Q K A L E L D P R S A E A W Y N L G N A Y Y
K Q G D Y D E A I E Y Y Q K A L E D P R S A E A W Y N L G N A Y Y K
Q G D Y D E A I E Y Y Q K A L E L D P R S A E A W Y N L G N A Y Y K
Q G D Y D E A I E Y Y Q K A L E L C P R S A E A W Y N L G N A Y Y K
Q G D Y D E A I E Y Y Q K A L E L D P R S A E A W Y N L G N A Y Y K
Q G D Y D E A I E Y Y Q K A L E D P R S A E A K Q N L G N A K Q K Q
G

```

The protein is expressed as His-tagged fusion and purified using standard affinity chromatography methods as previously described [71]. The protein is dialyzed into PBS buffer (150 mM NaCl, 50 mM phosphate buffer pH 7.4) with 2 mM β -mercaptoethanol and stored frozen at -20°C. The protein concentration was determined by UV-absorbance at 280 nm using the extinction coefficient calculated from the amino acid composition [345].

M.2.1.b.2. AuNPs

The AuNPs are obtained from NITparticles (nanoimmunotech): NIT gold carboxylated cluster 1.5 nm. CAS=231-791-2. N^o cluster/mL : $\sim 1 \times 10^{15}$ clusters/mL.

M.2.1.b.3. 2-(Pyridyldithio)-ethylaminehydrochloride (PDA*HCl)

PDA*HCl is synthesized as reported [361] with some modifications. To a stirred solution of aldrithiol (213 mg, 0.96 mmol) in MeOH (1.1 mL), 2-mercaptoethylamine hydrochloride (109 mg, 0.96 mmol) is added. After stirring 1h, the solvent is evaporated and the residue washed with cold AcOEt three times. PDA*HCl is obtained as a white solid in 51% yield: ¹H NMR (300 MHz, d⁶-CD₃OD) δ 8.57 (d, *J* = 5.0 Hz, 1H), 7.83 (t, *J* = 7.7 Hz, 1H), 7.69 (d, *J* = 8.0 Hz, 1H), 7.35 (dd, *J* = 7.5, 5.0 Hz, 1H), 3.18 (t, *J* = 6.1 Hz, 2H), 3.07 (t, *J* = 6.8

Hz, 2H); MS (ESI):*m/z* (%) 107 (100), 153 (79), 187 ($M^+ - Cl$, 12); HRMS (ESI) calcd for $C_7H_{11}NS_2(M^+ + 1)$ 187.0366, found 187.0391.

M.2.1.b.4. Modification of AuNPs by linkers

1 mL of an aqueous dispersion of AuNPs coated with citrate is mixed with 0.3 μmol of 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide, EDC, and 0.2 μmol of N-hydroxysuccinimide, NHS, for 10 min. In parallel, 0.1 μmol of cite-amina linker is mixed with 1 equivalent NaOH 1M for 10 min. After that time, both solutions are mixed and left to react at room temperature for at least 18 hrs. Then, the sample is washed several times with phosphate buffer saline, PBS pH 7.4, by ultrafiltration (3 kDa).

*M.2.1.b.5. Quantification of immobilized PDA*HCl linker*

During the washes by ultrafiltration, the filtrates are kept and are added 1,4-dithio-DL-treitol (DTT) (5 μL at 20 mM) and UV-Visible spectroscopy is done. By measuring UV-visible absorption of pyridine-2-thione at 343 nm (in PBS $\epsilon_{343\text{nm}} = 8080 \text{ M}^{-1}\text{cm}^{-1}$) the concentration of unbound linker is quantified in each filtrate. Washes are stopped after observing 2 consecutive filtrates without pyridine-2-thione released and immobilization yield is calculated.

M.2.1.b.6. Conjugation of linker modified AuNPs and CTPR₁₆-4Cys protein

Prior to any conjugation, purified CTPR₁₆-4Cys at a protein concentration of about 50 μM is freshly reduced with 5 mM DTT during 20 minutes to ensure full reduction of the sample. DTT is removed by buffer exchange over a NAP-5 column (GE Healthcare Life Science) in PBS buffer. Protein fractions without DTT are collected at 4°C. The protein concentration was measured by UV-absorbance.

To do the conjugation of the AuNPs with CTPR₁₆-4Cys protein 1 mL of an aqueous dispersion of linker modified AuNPs was mixed with 500 μL of CTPR₁₆-4Cys dissolved in PBS at 17 μM and is left to react for at least 48 h at room temperature (RT). After that time, sample is washed three times by ultrafiltration.

M.2.1.b.7. Co-NTA column

Co-NTA agarose affinity chromatography is used to remove the excess of nanoparticles and purify the CTPR₁₆-AuNPs conjugate, taking advantage of the His-tag of the CTPR₁₆ protein. 300 µl of the reaction mixture is incubated with 200 µl of cobalt resin (ABT 6BCL-QHCo-100 (Agarose Bead Technology)) during 45 minutes. After that, the first elution is collected and named as Flow Through. Then several washes with 500 µL of water each are done until getting a clear transparent elution. The protein is eluted using 500 µL of an aqueous solution of imidazole 300 mM. 200 µl fractions are collected. The eluted fractions are dialyzed into PBS buffer (150 mM NaCl, 50 mM phosphate buffer pH 7.4) buffer. The fractions are stored at 4°C.

M.2.1.b.8. Gel filtration chromatography

Gel filtration chromatography is performed using an AKTA prime plus Fast Protein Liquid Chromatography (FPLC) equipment (GE Healthcare). After Co-NTA column purification, the dialyzed elution fractions are injected into a Superdex 75 HR 10/30 size exclusion chromatography column (GE Healthcare) and run at 0.5 ml/min in PBS buffer with a detection UV absorption at 280 nm. The samples are collected in 0.5 ml fractions and stored at 4°C.

M.2.1.b.9. Analysis by agarose gel electrophoresis

To analyze the different mobilities of the samples 1% agarose electrophoresis gels are used prepared with 0.5x Tris-Borate-EDTA (TBE) buffer. 15 µL of concentrated CTPR₁₆-AuNPs conjugate sample is mixed with 5 µL of glycerol 80% and loaded in the gel. Samples of the CTPR₁₆-4Cys protein control are prepared using a 5 µl of 4x DNA standard loading buffer and 15 µl of the sample. AuNPs alone are prepared adding 25% glycerol to the sample. The samples are run at 100V during 120 minutes. The gel is stained with Coomassie solution (0.5% Coomassie blue, 40% methanol, 10% acetic acid) during 1 hour and unstained during over night in unstaining solution (40% methanol, 10% acetic acid).

M.2.1.b.10. High Resolution Transmission Electron Microscopy (HR-TEM)

TEM measurements are conducted on JEOL JEM 2100 microscope. All the carbon grids are exposed to a glow discharge just before use. The samples for TEM are prepared by depositing 10 µL of the sample solution on a carbon-coated TEM grid (~1 µM of CTPR₁₆-4cys protein in the conjugate). After 3 min, the excess solution is removed from the grid

using filter paper. To remove the deposited salt, the grid is washed with a droplet of water and the excess water is dried using filter paper. Then a droplet of a uranyl acetate solution is used to treat the grid for 1 min and the excess solution is removed using filter paper.

M.2.1.b.11. Scanning Tunnel Microscope (STM)

STM images are obtained with a home-built Scanning Tunnelling Microscope designed for room-temperature experiments [362]. All the images are recorded in ambient conditions using commercial gold substrates (Arrandee) cleaned prior sample deposition by flame-annealing. Freshly cut gold wires (99.99%) are used as tips. Samples are prepared by the drop casting technique from an aqueous solution of CTPR₁₆-AuNPs. For individual molecule imaging, 0.1-10 nM concentrations of proteins are used, while for layer formation, 1 μ M protein concentration is used. A drop of 200-300 μ L of the solution is deposited over our 1 cm² gold substrates, which are rinsed with water several times after assembly periods of 10 to 30 minutes. The gold surface is then dried under N₂ flow. Images are recorded using bias voltage values between 0.1 V and 1.5 V and a typical set point current of 0.1 nA. The body of the protein could not be imaged in range of voltages studied reflecting its low electrical conductance. The typical apparent height of the AuNPs is 0.4 nm, accounting for their expected poor coupling to the gold substrate.

M.2.1.b.12. Film formation

Protein solid ordered films are generated as previously described [220]. CTPR₁₆-4Cys protein alone and CTPR₁₆-AuNPs conjugate are diluted to 3% (w/v) protein concentration in 10 mM NaCl, 10 mM Na phosphate pH 7.0 buffer. The solutions are deposited on different surfaces, depending on the experiments to be performed. Glass surface for conductivity measurements and silicon wafer for X-ray diffraction (XRD) analysis. The drop volumes also vary between 10 to 30 μ l. The solvent is evaporated at room temperature during 12 hours, resulting in solid thin films.

M.2.1.b.13. X-ray diffraction (XRD)

X-ray diffraction is performed in a Panalytical X'Pert PRO diffractometer with Cu tube (λ K α =1.54187 Å) operated at 45 kV, 40 mA, Ni beta filter, programmable divergence and anti-scatter slits working in fixed mode, and fast linear detector (X'Celerator) working in scanning mode.

M.2.1.b.14. Lithography of the electrodes

The electrode system is fabricated through clean room processes of "maskless" lithography ("Heidelberg DWL66fs" model) and thermal evaporation. The silicon surface, covered by SiO₂, is coated by 2 μm "AZ 1512HS" (MicroChemicals GmbH) positive resin that is exposed to λ =405 nm laser that prints the designed electrode system.

The printed resin is introduced in a AZ351B developer (1:4 developer:water) to remove the resin from the part of the resin that is exposed to the laser. To ensure that all the resin is removed, 30 s plasma (50 W). The revealed sample is introduced in a thermal evaporator (Nanosphere, de Oxford Vaccum Science model) where it is evaporated a chromium I 5-10 nm thickness layer first to increase the adhesion to the substrate and, above 50 nm thickness gold layer.

To finish, a lift-off process is performed, introducing the sample in acetone to remove all the resin in the sample and obtaining the desired pattern.

M.2.1.b.15. Electrical conductivity measurements

In order to study the electrical properties of the CTPR₁₆-AuNPs conjugate, thin film devices over silicon wafers are prepared. Gold electrodes are deposited over the SiO₂ surface by mask-less optical lithography (Heidelberg DWL66fs) followed by thermal evaporation. For the lithography step, 17 μm spin-coated layer of AZ 1512HS positive resist (MicroChemicals GmbH) is used, which is exposed to a laser of 405 nm wavelength. Then a gold layer of 50 nm is evaporated, preceded by a Chromium layer of about 5-10 nm in thickness. Interdigitated electrodes are depicted, which defined a channel of 20 μm in length L and 11.8 μm in width W. This results in a ration width to length (W/L) of nearly 600, which allowed to obtained measurable currents even for low-conducting plain CTPR₁₆ films, and reduced the effect of edge currents between the electrodes.

Using a Keithley 4200-SCS, the current versus voltage curves are recorded in the interval (-1 V, +1 V) for both the CTPR₁₆-4Cys protein and the CTPR₁₆-AuNps conjugate. No lager voltages are used due to the pronounced current increase observed when raising the voltage due to the film heating. Due to the low-conductance of the CTPR₁₆-4Cys, a thicker film of this protein than of the CTPR₁₆-AuNPs conjugate is prepared. In particular, 4-5 μm thick films are prepared for CTPR₁₆-4Cys protein, and 20-100 nm thick films for the CTPR₁₆-AuNPs conjugate.

M.2.1.b.16. Thickness of the film

The thickness of the CTPR₁₆ film samples is measured using a profilometer. The layer thickness is measured using a “Dektak XT” mechanical profilometer with a 2.5 μm radius stylus and 1 mg force (the minimal force).

The thickness of the films is measured using a JPK Nanowizard II AFM operating in air. First, the films that are used to perform the conductivity measurements are scratched forming a thin line where there is removed the entire sample. Rectangular HQ:XSC11/Hard/AIBS cantilevers from MikroMasch are used in contact mode. Their nominal spring constant is around 0.2 N m⁻¹ and their typical tip radius is less than 20 nm. JPK SPM Data Processing software is employed to perform linear background subtraction on the images and to extract profiles and roughness analysis from them.

M.2.2. Generation of electron donor-acceptor systems

M.2.2.a. Organizing porphyrins in solution and solid state

M.2.2.a.1. Protein design and purification

Based on a consensus CTPR4 protein, two cysteine residues are introduced in each CTPR repeat at positions 15 and 17 within the loop connecting the helix A and B within the repeat. Also, a stop codon is included in the C-terminal of the CTPR protein to remove the solvating helix. The mutations are introduced in the gene encoding CTPR1WT by quick-change site directed mutagenesis using the following oligonucleotide sequences.

The protein is expressed as His-tagged fusion and purified using standard affinity chromatography methods based on previously published protocols [71] for His-tagged CTPR proteins using 0.5% deoxycholic acid in the lysis buffer. The protein is dialyzed into PBS buffer (150 mM NaCl, 50 mM phosphate buffer pH 7.4) and stored frozen at -20°C. The protein concentration is determined by UV-absorbance at 280 nm using the extinction coefficient calculated from the amino acid composition [345] (51800 M⁻¹ cm⁻¹).

M.2.2.a.2.CTPR4_CP-1/CTPR4_CP-2 conjugation

Prior to any conjugation, purified CTPR4_CP at a protein concentration of about 100 μM is freshly reduced with 1 mM 1,4-dithio-DL-treitol (DTT) during 20 minutes to ensure full reduction of the cysteine residues. DTT is removed by buffer exchange over a NAP-5 column (GE Healthcare Life Science) in PBS buffer. Protein fractions without DTT are

collected at 4°C and 1 mM TCEP is added to avoid the formation of disulfide bonds between cysteines during the conjugation reaction. The protein concentration is measured by UV-absorbance. 300 µl of 50 µM of freshly reduced protein is mixed with 40 equivalents of **1** or **2** (Figure 53) (around 1.9 mg) giving a ratio of 1:5 cysteine:maleimide and mixed gently. After an hour, an extra of 20 equivalents of **1** or **2** is added to the mixture (around 1 mg) giving a final ratio 1:7.5 cysteine:maleimide. The reaction mixture is incubated during 3 hours shaking and protected from light. 15% SDS-PAGE electrophoresis gels are used to monitor the conjugation process. Samples of the CTPR4_CP protein control and marker are prepared using Amresco EZ-vision loading buffer and the conjugates are mixed with SDS loading buffer. The gel prior staining is imaged using UV-light to monitor the fluorescence of the porphyrins. Then, the gels are stained with Coomassie Blue.

M.2.2.a.3. MALDI-TOF Mass spectroscopy.

The samples are analyzed using ABi 4800 MALDI TOF/TOF mass spectrometer. 4-Hydroxy-3-5-dimethoxycinnamic acid is used as matrix at 10 mg/ml dissolved in a 70% acetonitrile and 0,1% TFA solution. The samples are prepared at a sample to matrix solution ratio of 1:4 (v/v) and 1 µl of the mixture deposited on the sample plate. When the sample dried, 1 µl extra of matrix is deposited over the sample.

M.2.2.a.4. Size exclusion chromatography

To purify the conjugate from the free porphyrins gel filtration chromatography is performed using an AKTA prime plus Fast Protein Liquid Chromatography (FPLC) equipment (GE Healthcare). The conjugation reaction is injected into a Superdex 75 HR 10/30 size exclusion chromatography column (GE Healthcare) and run at 0.5 mL/min in PBS buffer with 2 mM β-mercaptoethanol. The purified samples are collected in 0.5 mL fractions and stored protected from light.

M.2.2.a.5. Absorbance measurements

Absorbance spectra are recorded using a VARIAN-80 UV-vis spectrophotometer. The absorbance spectra of the protein, porphyrin, and protein-porphyrin conjugates from 230 nm to 1000 nm are acquired in a 1 cm pathlength quartz cuvette using a 4 nm slit-width. For the denaturalization experiments, absorption spectra of the conjugates in different percentage of methanol are recorded. The equilibrium denaturation studies are

performed by preparing two stock solutions at 5 μM of CTPR4_CP-1 and CTPR4_CP-2 in a PBS buffer and in methanol. The two solutions are mixed at different ratios to obtain solutions of conjugates at the same concentration and at different percentage of methanol (from 0 to 100%).

M.2.2.a.6. Circular dichroism (CD) measurements

CD spectra are measured using a Jasco J-815 CD Spectrometer. CD spectra of CTPR protein and CTPR4_CP-1 and CTPR4_CP-2 conjugates in PBS buffer are acquired in a 1 cm path length quartz cuvette at protein a concentration of 1.7 μM . CD spectra of **1** and **2** are acquired under the same conditions using 1.7 μM of porphyrin concentration. All the CD spectra are recorded with a band-width of 1 nm at 1 nm increments and 10 second average time. CD spectra of CTPR protein films are acquired in 0.01 cm pathlength quartz cuvette. 10 μl of the conjugate at 1% (w/v) protein concentration in 10 mM NaCl, 10mM Na phosphate pH 7.0 buffer are deposited on the quartz cuvette, and the solvent is left to evaporate. The CD spectra are recorded at 1 nm increments and 10 second average time. 10 scans are accumulated.

M.2.2.a.7. Film formation

Solid CTPR protein ordered films are generated as previously described [103]. CTPR4_CP protein alone and porphyrin conjugates CTPR4_CP-1 and CTPR4_CP-2 are diluted to 3% (w/v) protein concentration in 10 mM NaCl, 10 mM Na phosphate pH 7.0 buffer. The solutions are deposited on different surfaces, depending on the experiments to be performed. Quartz cuvette is used for CD analysis and fluorescence spectra, quartz plate for conductivity measurements, and silicon wafer for XRD analysis. The drop volumes also vary between 10 to 30 μl . The solvent evaporated at room temperature during 12 hours, resulting in solid thin films.

M.2.2.a.8. Fluorescence anisotropy measurements

Fluorescence anisotropy experiments are recorded in a Fluorolog–TCSPC spectrofluorometer (Horiba) with excitation and emission polarizers. Protein films are placed on a quartz sandwich cuvette and the fluorescence intensity signal of the film of the CTPR4_CP-1 and CTPR4_CP-2 conjugates is monitored. The samples are excited at 420 nm with a 4 nm slit-width and the fluorescence emission of the porphyrins is recorded at 650 nm with slit width of 4 nm. The fluorescence intensity is recorded with a fixed polarizer at 0 degrees in the excitation path and varying the angle of the polarizer placed in the emission path. The change of the intensity of the porphyrin in solution is recorded

to account the difference in transmission efficiencies when the emission polarizer is placed at different angles. The intensity signal of the film is corrected for this change and then calculated the intensity change with respect to the fluorescence intensity (F_1) when both polarizer are placed at 0 degrees (F_1/F_{10}).

M.2.2.a.9. X-Ray diffraction (XRD)

X-ray diffraction is performed in a Panalytical X'Pert PRO diffractometer with Cu tube ($\lambda = 1.54187 \text{ \AA}$) operated S7 at 45 kV, 40 mA, Ni beta filter, programmable divergence and anti-scatter slits working in fixed mode, and fast linear detector (X'Celerator) working in scanning mode.

M.2.2.a.10. FP-TRMC and TAS measurements

Charge carrier mobility is evaluated by flash-photolysis time resolved microwave conductivity (FP-TRMC) and transient absorption spectroscopy (TAS) techniques at room temperature under air. Solid CTPR4_CP-1 and CTPR4_CP-2 films are prepared by dropcasting of their Milli-Q water solution. Charge carriers are photochemically generated using a third harmonic generation ($\lambda = 355 \text{ nm}$) of a Spectra Physics model INDI-HG Nd:YAG laser with a pulse duration of 5–8 ns and frequency of 10 Hz. The photon density of a 355 nm pulse was $9.1 \times 10^{15} \text{ photons cm}^{-2} \text{ pulse}^{-1}$. The microwave frequency and power are set at $\sim 9.1 \text{ GHz}$ and 3 mW, respectively. The TRMC signal, picked up by a diode (rise time $< 1 \text{ ns}$), is monitored by a Tektronics model TDS3032B digital oscilloscope. The observed conductivities are normalized, given by a photocarrier generation yield (Φ) multiplied by sum of the charge carrier mobilities ($\Sigma\mu$), according to the Equation (11):

$$\Phi \Sigma\mu = \left(\frac{1}{eA I_o F_{light}} \right) \frac{\Delta P_r}{P_r} \quad (11)$$

Where, e , A , I_o , F_{light} , P_r and ΔP_r are unit charge of a single electron, sensitivity factor ($S^{-1} \text{ cm}$), incident photon density of the excitation laser (photon cm^{-2}), correction (or filling) factor (cm^{-1}), and reflected microwave power and its change, respectively.

TAS measurements are carried out at room temperature under air. The identical drop-cast films used for FP-TRMC measurements are used for TAS measurements. The film is photoexcited using a third harmonic generation ($\lambda = 355 \text{ nm}$) of a Spectra Physics model INDI-HG Nd:YAG laser with a pulse duration of 5–8 ns and frequency of 10 Hz, where the photon density of a 355 nm pulse was $9.1 \times 10^{15} \text{ photons cm}^{-2} \text{ pulse}^{-1}$. A white light continuum from a Xe lamp is used as a probe light source for transient absorption

spectroscopy. The monochromated probe light is guided into a Hamamatsu model C7700 wide-dynamic-range streak camera system, which collected a two-dimensional image of the spectral and temporal profiles of light intensity.

M.2.2.b. Generation of different sizes CTPR porphyrin arrays

M.2.2.b.1. Protein design and purification

The CTPR8 and CTPR16 gene is generated from the CTPR4_CP mutated gene by sequential additions of identical mutated repeats and cloned into pPro-EXHTa vector.

The protein is expressed as His-tagged fusion and purified using standard affinity chromatography methods based on previously published protocols [71] for His-tagged CTPR proteins using 0.5% deoxycholic acid in the lysis buffer. The protein is dialyzed into PBS buffer (150 mM NaCl, 50 mM phosphate buffer pH 7.4) and stored frozen at -20°C . The protein concentration is determined by UV-absorbance at 280 nm using the extinction coefficient calculated from the amino acid composition [345] ($98640\text{ M}^{-1}\text{ cm}^{-1}$ for CTPR8_CP protein and $189320\text{ M}^{-1}\text{ cm}^{-1}$ for CTPR16_CP protein).

M.2.2.b.2. CTPR4/8/16_CP-2 conjugations

Prior to any conjugation, purified CTPR4/8/16_CP at a protein concentration of about $100\mu\text{M}$ is freshly reduced with 1 mM 1,4-dithio-DL-treitol (DTT) during 20 minutes to ensure full reduction of the cysteine residues. DTT is removed by buffer exchange over a NAP-5 column (GE Healthcare Life Science) in PBS buffer. Protein fractions without DTT are collected at 4°C and 1mM TCEP is added to avoid the formation of disulfide bonds between cysteines during the conjugation reaction. The protein concentration is measured by UV-absorbance. $300\mu\text{l}$ of $50\mu\text{M}$, $25\mu\text{M}$ or $12.5\mu\text{M}$ of freshly reduced CTPR4_CP, CTPR8_CP or CTPR16_CP protein, respectively, is mixed with 40 equivalents of **2** (around 1.9 mg) giving a ratio of 1:5 cysteine:maleimide and mixed gently. After an hour, an extra of 20 equivalents of **2** is added to the mixture (around 1 mg) giving a final ratio 1:7.5 cysteine:maleimide. The reaction mixture is incubated during 3 hours shaking and protected from light. 15% SDS-PAGE electrophoresis gels are used to monitor the conjugation process. Samples of the CTPR4/8/16_CP protein controls and marker are prepared using Amresco EZ-vision loading buffer and the conjugates are mixed with SDS loading buffer. The gel prior staining is imaged using UV-light to monitor the fluorescence of the porphyrin **2**. Then, the gels are stained with Coomassie Blue.

M.2.2.b.3. Gel filtration chromatography

To purify the conjugates from the free porphyrin gel filtration chromatography is performed using an AKTA prime plus Fast Protein Liquid Chromatography (FPLC) equipment (GE Healthcare). The conjugation reaction is injected into a Superdex 75 HR 10/30 size exclusion chromatography column (GE Healthcare) and run at 0.5 mL/min in PBS buffer with 2mM β -mercaptoethanol. The purified samples are collected in 0.5 mL fractions and stored protected from light.

M.2.2.b.4. Absorbance measurements

Absorbance spectra are recorded using a VARIAN-80 UV-vis spectrophotometer. The absorbance spectra of the protein, porphyrin, and protein-porphyrin conjugates from 230 nm to 1000 nm are acquired in a 1 cm pathlength quartz cuvette using a 4 nm slit-width.

M.2.2.b.5. Circular dichroism (CD) spectra

CD spectra are measured using a Jasco J-815 CD Spectrometer. CD spectra of proteins, conjugates and nanohybrids in water are acquired in a 1 mm path length quartz cuvette at a protein concentration of around 10 μ M. All the CD spectra are recorded with a bandwidth of 1 nm at 1 nm increments and 4 second average time.

M.2.2.c. Donor/acceptor nanohybrids based on superhelical protein-porphyrin conjugates and SWCNT

M.2.2.c.1. Protein synthesis and purification

Based on a consensus CTPR1_CP protein, that contains the Y17C and Q14C mutations and the stop codon to remove the solvating helix, two histidines are introduced in the back face of the CTPR1_CP protein. The CTPR4/8/16_CP gene is generated from the CTPR1 mutated gene by sequential additions of identical mutated repeats cloned into pPro-EXHTa vector.

The proteins are expressed as His-tagged fusion and purified using standard affinity chromatography methods based on previously published protocols [71] for His-tagged CTPR proteins using 0.5% deoxycholic acid in the lysis buffer. The proteins are dialyzed into PBS buffer (150 mM NaCl, 50 mM phosphate buffer pH 7.4) and stored frozen at -20°C. The protein concentration is determined by UV-absorbance at 280 nm using the

extinction coefficient calculated from the amino acid composition [345] ($51800 \text{ M}^{-1} \text{ cm}^{-1}$ for CTPR4_CP, $98640 \text{ M}^{-1} \text{ cm}^{-1}$ for CTPR8_CP protein and $189320 \text{ M}^{-1} \text{ cm}^{-1}$ for CTPR16_CP protein).

M.2.2.c.2. Circular dichroism (CD) measurement

CD spectra are measured using a Jasco J-815 CD Spectrometer. CD spectra of proteins, conjugates and nanohybrids in water are acquired in a 1 mm path length quartz cuvette at a protein concentration of around $10 \mu\text{M}$. All the CD spectra are recorded with a bandwidth of 1 nm at 1 nm increments and 4 second average time.

M.2.2.c.3. Wrapping of SWCNT with proteins and conjugates

Firstly, the absorbance at $\lambda = 280 \text{ nm}$ of the corresponding CTPR4_CP, CTPR8_CP and CTPR16_CP mutated proteins is adjusted to 0.13, that is, a final concentration of $2.5 \mu\text{M}$, $1.25 \mu\text{M}$ and $0.63 \mu\text{M}$, respectively, in PBS buffer. This molar concentration implied a weight/volume percent of 0.005% (w/v) for all the samples. Meanwhile, in the case of CTPR4_CP-2, CTPR8_CP-2 and CTPR16_CP-2 conjugates, the same strategy is followed, adjusting the absorbance at $\lambda = 280 \text{ nm}$ to 0.13.

Secondly, 2 mL of each sample are transferred to a glass vial and are sonicated in an ultrasound bath in the presence of around 1 mg of (7,6)-enriched SWCNTs. The temperature of the bath is maintained at around $15\text{-}20^\circ\text{C}$ to avoid denaturation of the proteins. Every 15 min, each dispersion is transferred to a microcentrifuge tube and is centrifuged at 12000 r.p.m. for 3 min. The 75% of the supernatant is recovered from the centrifugation tubes, avoiding sediment at the bottom, and is transferred to an absorbance cuvette for monitoring the amount of wrapped SWCNTs. This process is repeated until no more changes in the UV-vis spectra are observed.

Finally, the supernatant obtained for all nanohybrids is centrifuged 30 min more at 5000 r.p.m. to perfectly remove the non-functionalized SWCNTs, obtaining clear solutions in all the cases. These solutions are kept in the fridge where they remained stable over months.

M.2.2.c.4. Absorbance measurements

Absorbance spectra are recorded using a VARIAN-80 UV-vis spectrophotometer. The absorbance spectra of the protein, porphyrin, and protein-porphyrin conjugates from 230 nm to 1000 nm are acquired in a 1 cm pathlength quartz cuvette using a 4 nm slit-width.

M.2.2.c.5. Raman measurements

CTPR16_CP/SWCNT and CTPR16_CP-2/SWCNT are characterized by using a Renishaw inVia Raman microscope (laser power 0.05%, 785 nm, optical lens 1000x) on a glass microscope slide.

M.2.2.c.6. Transmission electron microscopy (TEM)

TEM images are performed in a JEOL JEM 1011 electron microscope operated at 100 kV. Images were directly recorded using a GATAN Erlangshen ES 1000W camera attached to the microscope.

M.2.2.c.7. X-ray diffraction (XRD)

X-ray diffraction was performed in a Panalytical X'Pert PRO diffractometer with Cu tube ($\lambda = 1.54187 \text{ \AA}$) operated at 45 kV, 40 mA, Ni beta filter, programmable divergence and anti-scatter slits working in fixed mode, and fast linear detector (X'Celerator) working in scanning mode.

M.2.2.c.8. FP-TRMC and TAS measurements

The charge carrier transport property along the π -stacked porphyrin assemblies is evaluated by FP-TRMC technique at room temperature under air. Uniform thin films are prepared on a quartz plate by dropcasting of deionized water solution of CTPR16, CTPR16/SWCNT, and CTPR16-2/SWCNT. Transient charge carriers are generated through photoexcitation by laser pulses of third harmonic generation ($\lambda = 355 \text{ nm}$) from a Spectra Physics model INDI-HG Nd:YAG laser with a pulse duration of 5–8 ns. Light pulses at 420 nm are also used as excitation light source from a Spectra Physics model versaScan optical parametric oscillator unit pumped with THG pulses of an identical laser system. The photon density of a 355 nm and 420 nm pulse is set at 4.6×10^{15} and 8.1×10^{15} photons

$\text{cm}^{-2} \text{ pulse}^{-1}$, respectively. The probing microwave frequency and power are set at 9.1 GHz and 3 mW, respectively. Photoconductivity transients, demodulated through a GaAs crystal-diode with Schottky-barriers (rise time < 1 ns), are monitored by a Tektronix model TDS3032B digital oscilloscope. The observed conductivities are normalized, given by a photocarrier generation yield (ϕ) multiplied by sum of the charge carrier mobilities ($\Sigma\mu$), according to the Equation (11).

Transient absorption spectroscopy (TAS) measurements are carried out at room temperature under air. The identical dropcast films used for FP-TRMC measurements are used for TAS measurements. The film is photoexcited using the third harmonic generation ($\lambda = 355 \text{ nm}$) from the identical Spectra Physics laser, where the photon density of a 355 nm pulse was $4.6 \times 10^{15} \text{ photons cm}^{-2} \text{ pulse}^{-1}$. A white light continuum from a Xe lamp is used as a probe light source for transient absorption spectroscopy. The monochromated probe light is guided into a Hamamatsu model C7700 wide-dynamic-range streak camera system, which collected a two-dimensional image of the spectral and temporal profiles of light intensity.

M.2.3. Mimicking natural phosystems

M.2.3.1. Protein design and purification

Based on a consensus CTPR2 protein, 4 cysteine cavity is formed based on the design done in the Giovanna's lab [363]. A four distalcysteine cubic cavity is formed by introducing four cysteines between the A helix of two consecutive CTPR repeats: changing Y5C and N9C in the first repeat and E2C and N6C in the second repeat. In the second repeat a stop codon is introduced to remove the solvating helix. The mutations are introduced in CTPR1WT blocks to generate two mutated repeats C1 and C2 by quick-change site directed mutagenesis. Using the two mutated CTPR1 units (C1 and C2) are then fused to create the CTPR2_4cys gene by sequential addition of C1 and C2 cloned into pPro-EXHTa vector. The CTPR4_8cys and CTPR8_16cys are generated by sequential addition of CTPR2_4cys protein cloned into pPro-EXHTa vector.

The protein is expressed as His-tagged fusion and purified using standard affinity chromatography methods based on previously published protocols [71] for His-tagged CTPR proteins using 0.5 M UREA in the lysis buffer. The protein is dialyzed into PBS buffer (150 mM NaCl, 50 mM phosphate buffer pH 7.4). The protein concentration is determined by UV-absorbance at 280 nm using the extinction coefficient calculated from the amino acid composition [345] ($32110 \text{ M}^{-1} \text{ cm}^{-1}$ for CTPR2_4cys, $59750 \text{ M}^{-1} \text{ cm}^{-1}$ for CTPR4_8cys and $115030 \text{ M}^{-1} \text{ cm}^{-1}$ for CTPR8_16cys). The His-tag of CTPR protein is cleaved using TEV protease. In the PBS buffer containing the proteins 1 mM DTT and

0.5 mM EDTA. In the mixture TEV protease is added in protein ratio (w/w) 1:4 CTPR protein: TEV. The cleavage is performed at 4°C overnight. The TEV is removed using standard affinity chromatography, where the CTPR protein without his-tag is collected in the flowthrough, while the TEV protease is joined in the affinity resin. The cleavage CTPR protein is dialyzed into 100 mM Tris-HCl, pH 8.5 buffer and stored at -20°C.

For porphyrin incorporation a CTPR2_CA protein is designed based on CTPR2_4cys protein. In CTPR2_4cys protein a p-Azido-L-phenylalanine non-natural amino acid is incorporated in the loop between helix A of the repeat 1 and helix A of repeat 2. The non-natural amino acid incorporation is done by the amber stop codon suppression method using an evolved mutant tRNA/aaRS pair from *Methanococcus jannaschii* [344]. The amber stop codon is introduced in CTPR2_4cys gene by quick-change site directed mutagenesis.

To obtain the CTPR2_CA protein the CTPR2_CA gene, cloned into pPro-EXHTa, and pEVOL-pAz plasmid are cotransformed into C41(DE3) *E. coli* and a single colony is used to inoculate an overnight culture of 10 mL of fresh LB medium containing 100 µg/mL of ampicillin and 35 µg/ml of chloramphenicol and then grow overnight at 37°C in a shaking incubator. 5 mL of overnight culture is used to inoculate 1 L of fresh LB medium containing 100 µg/mL of ampicillin 34 µg/ml of chloramphenicol. When the culture reached an optical density at 600 nm of 0.4-0.5, 0.5 g of p-azido-L-phenylalanine amino acid (BACHEM. CAS number 33173-53-4) is added. When the culture reached an optical density at 600nm of 0.6-0.8, the expression is induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) (final concentration 0.6 mM) and L-Arabinose (final concentration 0.2%). Expression is done during 5 hours at 30°C. The protein is purified using standard affinity chromatography methods based on previously published protocols [71] for His-tagged CTPR proteins using 0.5 M UREA in the lysis buffer. The protein is dialyzed into PBS buffer (150 mM NaCl, 50 mM phosphate buffer pH 7.4) and stored frozen at -20°C. The protein concentration is determined by UV-absorbance at 280 nm using the extinction coefficient calculated from the amino acid composition [345] ($32110 \text{ M}^{-1} \text{ cm}^{-1}$). The His-tag of CTPR protein is cleaved using TEV protease. In the PBS buffer containing the proteins 1 mM DTT and 0.5 mM EDTA. In the mixture TEV protease is added in protein ratio (w/w) 1:4 CTPR protein: TEV. The cleavage is performed at 4°C overnight. The His-tag and tagged TEV protease is removed by a second affinity chromatography, where the CTPR protein without His-tag is collected in the flowthrough, while the TEV protease is joined to the affinity resin. The cleavage CTPR protein is dialyzed in 100 mM Tris-HCl, pH 8.5 buffer and stored at -20°C.

M.2.3.2. Direct incorporation of [4Fe-4S] cluster

Iron–sulfur clusters are incorporated into peptide variants by adapting well-established methodologies [364, 365]. All the reactions are performed in an anaerobic chamber (Coy Scientific) under an inert atmosphere (95% N₂ and 5% H₂ environment). To a solution of 50 μM CTPR protein in 100 mM Tris-HCl, pH 8.5, the following reagents are added sequentially at 20 min intervals to a final concentration: 0.8% (v/v) β-mercaptoethanol, freshly prepared 3 mM ferric chloride (FeCl₃), and freshly prepared 3 mM sodium sulfide (Na₂S). The mixture is incubated overnight at 4°C in the anaerobic chamber. The resulting dark brown solution is subjected to desalting with a PD10 G25 column (GE Healthcare) that is pre-equilibrated with 100 mM Tris at pH 7.5 to exclude all non-protein low molecular mass contaminants and salts. Appropriate fractions are collected where the protein elutes with the cluster.

M.2.3.3. Absorbance measurements

UV–vis spectra are recorded on a Perkin-Elmer Lambda-35 spectrophotometer under anaerobic Conditions. The absorption spectra are obtained from 230 nm to 800 nm are acquired in a 1 cm pathlength quartz cuvette using a 4 nm slit-width.

M.2.3.4. Cluster quantification

Cluster incorporation is assessed quantitatively by measuring independently the concentration of iron and of peptide in identical samples. The CTPR2-[4Fe-2S], CTPR4-2[4Fe-4S] and CTPR8-4[4Fe-4S] samples are split in two portions: one is used to measure peptide concentration (standard Bradford assay), and the second is used to determine iron concentration (ferrozine assay) [366, 367]. Before doing the Bradford assay, the protein is denatured at 95°C during 15 minutes to ensure that the Bradford reagent interacts efficiently with the amino acids in the protein. The estimated concentration of iron is then compared with the CTPR2_4cys, CTPR4_8cys and CTPR8_16cys protein concentration evaluated by Bradford assay.

M.2.3.5. Circular dichroism (CD) spectroscopy

CD spectroscopy of the CTPR2_4cys, CTPR4_8cys and CTPR8_16cys protein and CTPR2-[4Fe-4S], CTPR4-2[4Fe-4S] and CTPR8-4[4Fe-4S] in 10 mM Tris-HCl (pH 7.5) buffer is performed on a JASCO J-815 spectropolarimeter. Spectra are recorded from 260 to 190 nm in 1 nm increments using a quartz cuvette with a path length of 1 cm. Thermal

unfolding curves are monitored by following the CD signal at 222 nm as a function of temperature from 10 to 95°C in a 0.1 cm path length cuvette. The CTPR2-[4Fe-4S], CTPR4-2[4Fe-4S] and CTPR8-4[4Fe-4S] samples are handled in an airtight cuvette to exclude air during the course of the experiment. CD spectra of CTPR2_4cys, CTPR4_8cys and CTPR8_16cys proteins are recorded in the presence of excess reducing agent (TCEP) to prevent oxidation of the cysteine side chains.

M.2.3.6. Electron paramagnetic resonance (EPR) spectroscopy

Appropriate fractions from the PD10 column are concentrated using a centrifuge concentrator with a molecular mass cutoff of 3000 Da (GE Healthcare) to approximately 1mM protein concentration. Reduced samples are prepared by addition of 100 mM sodium dithionite in 1 M glycine buffer, pH 10, to a final concentration of 20 mM dithionite. EPR samples are prepared by addition of 10% (v/v) glycerol as a cryoprotectant and placed in quartz EPR tubes, after which the samples are flash frozen and stored under liquid N₂ until measurements.

EPR experiments are carried out on a X-band EPR spectrometer Elexsys E500 (Bruker) equipped with the ESR900 flow cryostat (Oxford Instruments). The measurement temperature is 15 K. Experimental conditions: mw frequency 9.336 GHz for CTPR2-[4Fe-4S], 9.333 GHz for CTPR4-2[4Fe-4S] and 9.335 GHz for CTPR8-4[4Fe-4S]; mw power, 2 mW; field modulation amplitude 0.5 mT.

M.2.3.7. Cyclic voltammetry (CV)

Electrochemical experiments are carried out using a CHinstruments 1242B potentiostat. For all electrochemical measurements, a three-electrode system is used: a 3-mm diameter glassy carbon working electrode with a surface area of 0.28 cm², platinum mesh counter electrode, and a Ag/AgCl reference electrode. Electrolyte solutions are degassed by incubation in anaerobic chamber for 2 days prior to use. Working electrodes are polished with 1 μm alumina for 5 min followed by 10 min of sonication. Electrodes are cleaned electrochemically by cycling 40 times between 1.2 and -1.2 V at 100 mV sec, followed by extensive washing with water, prior to use. Experimental conditions: 10 μM CTPR protein concentration; 100 mM sodium chloride, 100 mM Tris-HCl pH 7.5; 100 mV scan rate.

To calculate the parameters shown in table 3 is used the following equation:

E_{pc} = Cathodic wave = voltage value in the maximum of the oxidation peak

E_{pa} = Anodic wave = voltage value in the maximum of the reduction peak

E_p = cell potential = $E_{pc} - E_{pa}$

$E_{1/2}$ = Half wave potential = $\left| \frac{E_{pc} + E_{pa}}{2} \right|$

M.2.3.8. Activation of carboxyl group of porphyrin 3 with a phosphine group to obtain porphyrin 4

The reaction is carried out in a glove box (95% N₂ and 5% H₂) to ensure that no oxygen is present porphyrin **3** (Figure 81) (100 mg, 0.153 mmol), synthesized by Dr. Ghirlanda's group (Arizona State University), is dissolved in 100 mL of THF and (2-hydroxyphenyl) and (2-Hydroxyphenyl)diphenylphosphine (SIGMA-Aldrich CAS= 60254-10-6) (28.4 mg, 0.102 mmol) is added. Then, N,N-diisopropylethylamine (DIPEA) (36 μ l, 0.204 mmol), 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDCI) (58.7 mg, 0.306 mmol), and then 4-6-dimethylaminopyridine (DMAP) (62.3 mg, 0.510 mmol) are added. The reaction is stirred for 18 h at room temperature. The product, porphyrin **4**, is isolated by flash column chromatography using 5:15:80 methanol/ethyl acetate/dichloromethane obtaining a pure porphyrin **4** (Figure 81).

M.2.3.9. Conjugation between porphyrin 4 and CTPR2_CA protein

The conjugation between porphyrin **4** (Figure 81) and CTPR2_CA protein is performed in 50/50 100 mM Tris-HCl pH 8.5/Acetonitrile. The reaction is performed in a shaker at 60°C overnight under an inert atmosphere (95% N₂ and 5% H₂). The mixture is incubated overnight at 4°C in the anaerobic chamber. The resulting red-wine color solution is subjected to desalting with a PD10 G25 column (GE Healthcare) that is pre-equilibrated with 100 mM Tris pH 8.5 to exclude all non-protein low molecular mass contaminants and salts. Appropriate fractions are collected where the protein elutes with the porphyrin.

M.2.3.10. MALDI-TOF Mass spectroscopy.

The samples are analyzed using ABi 4800 MALDI TOF/TOF mass spectrometer. 4-Hydroxy-3-5-dimethoxycinnamic acid is used as matrix at 10 mg/ml dissolved in a 70% acetonitrile and 0.1% TFA solution. The samples are prepared at a sample to matrix solution ratio of 1:4 (v/v) and 1 μ l of the mixture deposited on the sample plate. When the sample dried, 1 μ l extra of matrix is deposited over the sample.

References

References

1. Plasticity and Toughness in Bone. *Physics Today*. 2009. **62**(41-47).
2. The mechanical efficiency of natural materials. *Philosophical Magazine*. 2004. **84**(2167-2186).
3. Adhesion and friction in gecko toe attachment and detachment. *Proc Natl Acad Sci USA*. 2006. **103**(19320–19325).
4. Effects of micro- and nano-structures on the self-cleaning behaviour of lotus leaves. *Nanotechnology*. 2006. **17**(1359-1362).
5. Adhesive force of a single gecko foot-hair. *Nature*. 2000. **405**(681-685).
6. Evidence for van der Waals adhesion in gecko setae. *Proc. Natl. Acad. Sci. USA*. 2002. **99**(12252-12256).
7. Recent developments in bio-inspired special wettability. *Chem Soc Rev*. 2010. **39**(3240-3255).
8. Mimicking nacre with super-nanotubes for producing optimized super-composites. *Nanotechnology*. 2006. **17**(5480–5484).
9. Bio-mimetic mechanisms of natural hierarchical materials: A review. *Journal of the mechanical behavior of biomedical materials*. 2013. **19**(3–33).
10. Sticking to the story: outstanding challenges in gecko-inspired adhesives. *Journal of Experimental Biology*. 2016. **219**(912-919).
11. Superhydrophobicity in perfection: the outstanding properties of the lotus leaf. *Beilstein J. Nanotechnol*. 2011. **2**(152–161).
12. The shell matrix of the freshwater mussel *Unio pictorum* (Paleoheterodonta, Unionoida). *FEBS J*. 2007. **274**(2933-2945).
13. *Small*. 2005. **1**(172-179).
14. Self-assembling peptides and proteins for nanotechnological applications. *Curr Opin Struct Biol*. 2004. **14**(480-486).
15. More than just bare scaffolds: towards multi-component and decorated fibrous biomaterials. *Chem Soc Rev*. 2010. **39**(3464-3479).
16. Designing peptide based nanomaterials. *Chem Soc Rev*. 2008. **39**(664-675).
17. Supramolecular polymerization. *Chem Rev*. 2009. **109**(5687-5754).
18. From natural to designer self-assembling biopolymers, the structural characterisation of fibrous proteins & peptides using fibre diffraction. *Chem Soc Rev*. 2010. **39**(3445-3453).
19. From short peptides to nanofibers to macromolecular assemblies in biomedicine. *Biotechnol Adv*. 2012. **30**(593-603).
20. Nanotechnology-based electrochemical sensors for biomonitoring chemical exposures. *J Expo Sci Environ Epidemiol*. 2009. **19**(1-18).
21. "Green" electronics: biodegradable and biocompatible materials and devices for sustainable future. *Chem Soc Rev*. 2014. **43**(588-610).
22. Nanostructured films from hierarchical self-assembly of amyloidogenic proteins. *Nat Nano*. 2010. **5**(204-207).
23. Folding DNA to create nanoscale shapes and patterns. *Nature*. 2006. **440**(297-302).
24. Engineering nanoscale order into a designed protein fiber. *Proc Natl Acad Sci USA*. 2007. **104**(10853-8).
25. Nanomaterials based on DNA. *Annu Rev Biochem*. 2010. **79**(65-87).
26. Structure by design: from single proteins and their building blocks to nanostructures. *Trends Biotechnol*. 2006. **24**(449-454).
27. The architectonics of programmable RNA and DNA nanostructured. *Curr Opin Struct Biol* 2006. **16**(531-543).

28. Self-assembled peptide nanostructures: the design of molecular building blocks and their technological utilization. *Chem Soc Rev.* 2007. **36**(1263-1269).
29. Amyloids: not only pathological agents but also ordered nanomaterials. *Angew Chem Int Ed Engl.* 2008. **47**(4062-4069).
30. Peptide self-assembly for crafting functional biological materials. . *Curr Opin Solid State Mater Sci.* 2011. **15**(225-235).
31. Fibrillar peptide gels in biotechnology and biomedicine. *Biopolymers.* 2010. **94**(49-59).
32. Functionalised amyloid fibrils for roles in cell adhesion. *Biomaterials.* 2008. **29**(1553-1562).
33. Designer self-assembling peptide nanofiber biological materials. *Chem Soc Rev.* . 2010. **39**(2780-2790).
34. Rationally engineering natural protein assemblies in nanobiotechnology *Curr Opin Biotechnol.* 2011. **22**(485-491).
35. Self-assembled peptide nanostructures: the design of molecular building blocks and their technological utilization. *Chem Soc Rev.* 2007. **36**(1263-1269).
36. Bionanotechnology: Lessons from Nature. *Wiley.* **2014**(
37. Assembly of the Herpes Simplex Virus Capsid: Characterization of Intermediates Observed During Cell-free Capsid Formation. *Journal of Molecular Biology.* 1996. **263**(432-446).
38. Flagellar rotation and the mechanism of bacterial motility. *Nature.* 1974. **249**(73–74).
39. Structure of a protein superfiber: Spider dragline silk. *Proc Nat Acad Sci USA.* 1990. **87**(7120-7124).
40. Screening Libraries To Identify Proteins with Desired Binding Activities Using a Split-GFP Reassembly Assay. *ACS Chem. Biol.* 2010. **5**(553–562).
41. Characterization of a unique [FeS] cluster in the electron transfer chain of the oxygen tolerant [NiFe] hydrogenase from Aquifex aeolicus. *PNAS.* 2011. **108**(6097–6102).
42. Design, Synthesis and Dynamics of a GFP Fluorophore Mimic with an Ultrafast Switching Function. *J Am Chem Soc.* 2016.
43. In Vitro Chromatin Assembly: Strategies and Quality Control. *Methods Enzymol.* 2016. **573**(3-41).
44. Preparation, Biochemical Analysis, and Structure Determination of SET Domain Histone Methyltransferases. *Methods Enzymol.* 2016. **573**(209-240.).
45. Metallomics and the Cell. *Series editors Sigel, Astrid; Sigel, Helmut; Sigel, Roland K.O. Springer.* 2013. ISBN 978-94-007-5560-4. Electronic-book ISBN 978-94-007-5561-1).
46. Bionanocomposites: A New Concept of Ecological, Bioinspired, and Functional Hybrid Materials. *Adv Mater.* 2007. **19**(1309–1319).
47. Peptide and protein based materials in 2010: from design and structure to function and application. *Chem Soc Rev.* 2010. **39**(3349-50).
48. Protein cages, rings and tubes: useful components of future nanodevices? *Nanotechnology Sci and Appl.* 2008. **1**(67-78).
49. Engineered and designed peptide-based fibrous biomaterials. *Curr Opin Solid State Mater Sci.* 2004. **8**(141-149).
50. Designing supramolecular protein assemblies. *Curr Opin Struct Biol.* 2002. **12**(464-470).
51. Protein engineering and design: from first principles to new technologies. *Curr Opin Struct Biol.* 2010. **20**(480-481).
52. Deformation and failure of protein materials in physiologically extreme conditions and disease. *Nat Mater.* 2009. **8**(175-188).
53. Design of a novel globular protein fold with atomic-level accuracy. *Science.* 2003. **302**(1364-8).

54. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. . *Science*. 2010. **329**(309-13).
55. Self-assembled bionanostructures: proteins following the lead of DNA nanostructures. *J Nanobiotechnology*. 2014. **12**(4).
56. Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science*. 2012. **336**(1171-1174).
57. Protein repeats: structures, functions, and evolution. *Journal of Structural Biology*. 2001. **134**(117-131).
58. Review: proteins with repeated sequence-structural prediction and modeling. *Journal of Structural Biology*. 2001. **134**(132-44).
59. TPR proteins: the versatile helix. *Trends Biochem. Sci*. 2003. **28**(655-662).
60. Designed to be stable: Crystal structure of a consensus ankyrin repeat protein. *Proc Nat Acad Sci USA*. 2003. **100**(1700-1705).
61. Consensus design of a NOD receptor leucine rich repeat domain with binding affinity for a muramyl dipeptide, a bacterial cell wall fragment. *Protein Sci*. 2014. **26**(790-800).
62. Designed armadillo repeat proteins as general peptide-binding scaffolds: consensus design and computational optimization of the hydrophobic core. *J Mol Biol*. 2008. **376**(1282-304).
63. TALEs from a spring--superelasticity of Tal effector protein structures. *PLoS One*. 2014. **9**(e109919).
64. Repeat protein engineering: creating functional nanostructures/biomaterials from modular building blocks. *Biochem Soc Trans*. 2013. **41**(1152-8).
65. Biomolecular templating of functional hybrid nanostructures using repeat protein scaffolds. *Biochem Soc Trans*. 2015. **34**(825-831).
66. A recurring theme in protein engineering: the design, stability and folding of repeat proteins. *Curr Opin Struc Biol*. 2005. **15**(464-471).
67. Exploring the repeat protein universe through computational protein design. *Nature*. 2015. **528**(580-584).
68. Rational design of alpha-helical tandem repeat proteins with closed architectures. *Nature*. 2015. **528**(585-8).
69. Control of the repeat-protein curvature by computational protein design. *Nat Struct Mol Biol*. 2015. **22**(167-74).
70. Design of stable alpha-helical arrays from an idealized TPR motif. *Structure*. 2003. **11**(497-508).
71. Consensus design as a tool for engineering repeat proteins. *Methods Mol Biol*. 2006. **340**(151-170).
72. Beyond consensus: statistical free energies reveal hidden interactions in the design of a TPR motif. *J Mol Biol*. 2004. **343**(731-745).
73. Designed proteins to modulate cellular networks. *ACS Chem. Biol*. 2010. **5**(545-552).
74. Creating novel proteins by combining design and selection. *Protein Eng. Des. Sel*. 2010. **23**(449-455).
75. Structure and stability of a consensus TPR superhelix. *Acta Crystallographica*. 2007. **D63**(800-811).
76. Structure and stability of designed TPR protein superhelices: unusual crystal packing and implications for natural TPR proteins. *Acta Crystallogr D Biol Crystallogr*. 2007. **63**(800-11).
77. A new folding paradigm for repeat proteins. *J. Am. Chem. Soc*. 2005. **127**(10188-10190).
78. Modulating repeat protein stability: The effect of individual helix stability on the collective behavior of the ensemble. *Protein Sci*. 2011. **20**(1042-1047).
79. Designing peptide based nanomaterials. *Chem Soc Rev*. 2008. **37**(664-75).

80. More than just bare scaffolds: towards multi-component and decorated fibrous biomaterials. *Chem Soc Rev.* 2010. **39**(3464-79).
81. Biomimetics: lessons from nature—an overview. *Phil. Trans. R. Soc. A* 2009. **367**(1445–1486).
82. Nanobiotechnology: Protein arrays made to order. *Nat Nanotechnol.* 2011. **6**(541-2).
83. Structure of a 16-nm designed by using protein oligomers. *Science.* 2012. **336**(1129).
84. Toward complex matter: supramolecular chemistry and self-organization. *Proc Natl Acad Sci USA.* 2002. **99**(4763-4768).
85. Molecular self-assembly: bioactive nanostructures branch out. *Nat Nanotechnol.* 2008. **3**(8-9).
86. Nanostructured films from hierarchical self-assembly of amyloidogenic proteins. *Nat Nanotechnol.* 2010. **5**(204-7).
87. Supramolecular chemical biology; bioactive synthetic self-assemblies. *Org Biomol Chem.* 2013. **11**(219-232).
88. Chemically controlled protein assembly: techniques and applications. . *Chem Rev.* 2010. **110**(3315-3336).
89. Principles for designing protein assemblies *Trends Cell Biol.* 2012. **22**(653-661).
90. Engineering and exploiting protein assemblies in synthetic biology. *Mol Biosyst.* 2009. **5**(723-732).
91. Protein-based Engineered Nanostructures. *Springer.* 2016. **book**(ISBN 978-3-319-39196-0).
92. In vitro Self-assembly of tailorable nanotubes from a simple protein building block *Proc Natl Acad Sci USA.* 2008. **105**(3733-3738).
93. A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. . *Science.* 2006. **312**(1526).
94. In vitro self-assembly of tailorable nanotubes from a simple protein building block. *Proc Natl Acad Sci USA.* 2008. **105**(3733-3738).
95. Designed Protein-Protein Association. *Science.* 2008. **319**(206-209).
96. Designed protein-protein association. *Science.* 2008. **319**(206-209).
97. Recent advances in functional supramolecular nanostructures assembled from bioactive building blocks. *Chem Soc Rev.* 2009. **4**(925-934).
98. Designing peptide based nanomaterials. *Chem Soc Rev.* 2008. **4**(664-675).
99. Next-generation peptide nanomaterials: molecular networks, interfaces and supramolecular functionality. *Chem Soc Rev.* 2010. **9**(351-357).
100. Applications of peptide and protein-based materials in bionanotechnology. *Chem Soc Rev.* 2010. **9**(3499-3509).
101. A self-assembly pathway to aligned monodomain gels. *Nat Mater.* 2010. **9**(594-601).
102. Crystal structure of a designed TPR module in complex with its peptide-ligand. *Febs J.* 2010. **277**(1058-1066).
103. Nanostructured functional films from engineered repeat proteins. *Soc. Interface.* 2013. **10**(20130051).
104. Repeat-proteins films exhibit hierarchical anisotropic mechanical properties. *Biomacromolecules.* 2015. **16**(706-14).
105. Nanostructured functional films from engineered repeat proteins. *J. R. Soc. Interface.* 2013. **10**(20130051).
106. Next-generation peptide nanomaterials: molecular networks, interfaces and supramolecular functionality. *Chem Soc Rev.* 2010. **39**(3351-3357).
107. Amyloid-based nanosensors and nanodevices. *Chem Soc Rev.* 2014. **15**(5326-5345).
108. Self-assembling peptides: implications for patenting in drug delivery and tissue engineering. *Recent Pat Drug Deliv Formul.* 2011. **1**(24-51).

109. Molecular self-assembly and applications of designer peptide amphiphiles. *Chem Soc Rev.* 2010. 3480-3498).
110. Designer self-assembling peptide nanofiber biological materials. *Chem Soc Rev.* 2010. **39**(2780-2790).
111. Self-assembly of amphiphilic peptides. *Soft Matter.* 2011. **7**(4122-4138).
112. Self-assembled proteins and peptides for regenerative medicine. *Chem Rev.* 2013. **113**(4837-4861).
113. ECM protein nanofibers and nanostructures engineered using surface-initiated assembly. *J Vis Exp.* 2014. **86**(
114. Short self-assembling peptides as building blocks for modern nanodevices. *Trends Biotechnol.* 2012. **30**(155-165).
115. Self-Assembled Nanoreactors. *Chem Rev* 2005. **105**(1445–1490).
116. Self-assembly of peptide amphiphiles: from molecules to nanostructures to biomaterials. *Biopolymers.* 2010. **94**(1-18).
117. Self-assembly of amphipathic β -sheet peptides: insights and applications. *Biopolymers.* 2012. **98**(169-184).
118. In vitro self-assembly of tailorable nanotubes from a simple protein building block. *Proc Natl Acad Sci USA.* 2008. **105**(3733-8).
119. A self-assembled protein nanotube with high aspect ratio. *Small.* 2009. **5**(2077-2084).
120. Bond Dissociation Energies in Simple Molecules. *NSRDS.* 1970. Washington).
121. Disulfide-Linked Protein Folding Pathways. *Annu Rev Cell Dev Biol.* 2008. **24**(211-235).
122. Collagen fibril formation. *J Biol Chem.* 1978. **253**(6578-6585).
123. Calorimetric study of a series of designed repeat proteins: modular structure and modular folding. *Protein Sci.* 2011. **20**(341-352).
124. Structure and stability of a consensus TPR superhelix. *Acta Crystallogr. D.* 2007. **D63**(800-811).
125. Mapping the energy landscape of repeat proteins using NMR-detected hydrogen exchange *J Mol Biol.* 2008. **379**(617-626).
126. Hyperthin nanochains composed of self-polymerizing protein shackles. *Nat Commun.* 2013. **4**(2211).
127. Extensive non-native polyproline II secondary structure induces compaction of a protein's denatured state. *J Mol Biol.* 2008. **382**(203-212).
128. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys J.* 2011. **101**(892-898).
129. Principles of polymerization. 2004. Wiley-Interscience).
130. Principles of Polymer Chemistry. *Cornell University Press.* 1953. Ithaca, NY.).
131. Step-growth polymerization of inorganic nanoparticles. *science.* 2010. **329**(197-200).
132. Fibrous nanostructures from the self-assembly of designed repeat protein modules. *Angew Chem.* 2012. **51**(13132-5).
133. Cooperative, reversible self-assembly of covalently pre-linked proteins into giant fibrous structures. *Angew Chem* 2014. **31**(8050-8055).
134. Multi-colored fibers by self-assembly of DNA, histone proteins, and cationic conjugated polymers. *Angew Chem.* 2014. **2**(424-428).
135. The role of protein hydrophobicity in conformation change and self-assembly into large amyloid fibers. *Biomacromolecules.* 2014. **4**(1240-1247).
136. The interaction of proteins with solid surfaces. *Opin Struct Biol.* 2004. **14**(110–115).
137. Biological activity of adsorbed proteins. *Surfactant Science Series.* 2003. 393–413).
138. Biomaterials Science: an Introduction to Materials in Medicine. *Academic Press.* **1996**(San Diego).
139. Chemical processing by self-assembly. *AIChE J.* 2001. **47**(1706–1710).

140. Self-assembled organic monolayers: model systems for studying adsorption of proteins at surfaces. *Science*. 1991. **252**(1164-1167.).
141. Infrared study of thiol monolayer assemblies on gold. Preparation, characterization, and functionalization of mixed monolayers. *Langmuir*. . 1993. **9**(141-149).
142. Self-assembled monolayers of thiolates on metals as a form of nanotechnology. *Chem Rev*. 2005. **105**(1103-1169).
143. Self-Assembly and Self-Assembled Monolayers in Micro- and Nanofabrication in Nanotechnology. *Springer*. 1999. 331-369).
144. Synthesis, structure and properties of model organic surfaces. *Annu Rev Phys Chem*. 1992. **43**(437-463).
145. Chemisorption of proteins and their thiol derivatives onto gold surfaces: characterization based on electrochemical nonlinearity. *Biophysical Chemistry*. 1996. **62**(63-72).
146. SPR Studies of the Nonspecific Adsorption Kinetics of Human IgG and BSA on Gold Surfaces Modified by Self-Assembled Monolayers (SAMs). *J Colloid and Interface Science*. 1997. **185**(94-103).
147. Extended and localized surface plasmons in annealed Au films on glass substrates. *J Appl Phys*. 2010. **108**(074303).
148. Effect of photodiode angular response on surface plasmon resonance measurements in the Kretschmann-Raether configuration. *Rev Sci Instrum*. 2012. **83**(093102).
149. DNA assembly on streptavidin modified surface: A study using quartz crystal microbalance with dissipation or resistance measurements. *Sensors and Actuators B: Chemical*. 2008. **131**(371-378).
150. Nanospring behaviour of ankyrin repeats. *Nature*. 2006. **440**(246-249).
151. Probing the molecular origin of native-state flexibility in repeat proteins *J Am Chem Soc*. 2015. **137**(10367-10373).
152. Atomic force microscopic observation of Escherichia coli ribosomes in solution. *Biosci Biotechnol Biochem*. 2006. **70**(300-2).
153. Role of mass accumulation and viscoelastic film properties for the response of acoustic-wave-based chemical sensors. *Anal Chem*. 1999. **71**(2488-2496).
154. Chiral conductive polymers as spin filters. *Adv Mater*. 2015. **27**(1924-1927).
155. Enantiospecific spin polarization of electrons photoemitted through layers of homochiral organic molecules. *Adv Mater*. 2014. **26**(7474-7479).
156. Non-magnetic organic/inorganic spin injector at room temperature. *Nat Mater*. 2010. **8**(638-642).
157. Biological applications of peptides nanotubes: an overview. *Peptides*. 2013. **39**(47).
158. Accessibility, Reactivity, and Selectivity of Side Chains within a Channel of de Novo Peptide Assembly. *J Am Chem Soc*. 2013. **135**(12524-12527).
159. The emerging field of nanotube biotechnology. *Nat Rev Drug Discov*. 2003. **2**(29-37.).
160. Peptide nanotubes. *Angew Chem Int Ed Engl*. 2014. **53**(6866-6881).
161. Casting metal nanowires within discrete self-assembled peptide nanotubes. *Science*. 2003. **300**(625-627).
162. Self-assembling organic nanotubes based on a cyclic peptide architecture. *Nature*. 1993. **366**(324-327).
163. Rational design of helical nanotubes from self-assembly of coiled-coil lock washers. *J Am Chem Soc*. 2013. **135**(15565-15578).
164. Engineered coiled-coil protein microfibers. *Biomacromolecules*. 2014. **15**(3503-3510).
165. De novo design of fibrils made of short alpha-helical coiled coil peptides. *Chem Biol*. 2001. **8**(1025-1032).
166. Structural plasticity of helical nanotubes based on coiled-coil assemblies. 2015.

167. Biomimetic organization: Octapeptide self-assembly into nanotubes of viral capsid-like dimension. *Proc Natl Acad Sci USA*. 2003. **100**(10258-10262).
168. Modular Design of Self-Assembling Peptide-Based Nanotubes. *J Am Chem Soc*. 2015. **137**(10554-10562).
169. A de novo peptide hexamer with a mutable channel. *Nat Chem Biol*. 2011. **7**(935-941).
170. Computational design of water-soluble α -helical barrels. *Science*. 2014. **346**(485-488).
171. Design and properties of functional nanotubes from the self-assembly of cyclic peptide templates. *Chem Soc Rev*. 2012. **41**(6023-6041).
172. Principles of nanostructure design with protein building blocks. *Proteins*. 2007. **68**(1-12).
173. *Nature*. 1982. **300**(341 - 342).
174. Recent progress in understanding hydrophobic interactions. *PNAS*. 2006. **103** (15739–15746).
175. Structure by design: from single proteins and their building blocks to nanostructures. *Trends Biotechnol*. 2006. **24**(449-454).
176. Silane Ligand Exchange to Make Hydrophobic Superparamagnetic Nanoparticles Water-Dispersible. *Chem Mater*. 2007. **19**(1821-1831).
177. Protein-Surfactant Interactions. *Encyclopedia of Surface and Colloid Science (Somasundaran)*. 2006. 5251-5263).
178. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins *Proc Natl Acad Sci USA*. 2009. **106**(1760-1765).
179. Spectroscopic studies on the interaction of bovine (BSA) and human (HSA) serum albumins with ionic surfactants. *Spectrochim Acta A Mol Biomol Spectrosc*. 2000. **56**(2255-2271).
180. Unfolding of β -Sheet Proteins in SDS. *Biophysical Journal* 2007. **92** (3674–3685).
181. Protein Unfolding in Detergents: Effect of Micelle Structure, Ionic Strength, pH, and Temperature. *Biophysical Journal*. 2002. **83** (2219–2230).
182. Structural organization of the pentameric transmembrane R-helices of phospholamban, a cardiac ion channel. *EMBO J*. 1994. **13**(4757–4764).
183. NMR study of the tetrameric KcsA potassium channel in detergent micelles. . *Protein Sci*. 2006. **15**(684–698).
184. Glycophorin A dimerization is driven by specific interactions between transmembrane R-helices. *J. Biol. Chem.* . 1992. **267**(7683–7689).
185. Retention of native-like oligomerization states in transmembrane segment peptides: Application to the Escherichia coli aspartate receptor. *Biochemistry*. 2001. **40**(11106–11113).
186. Oligomerization of a peptide derived from the transmembrane region of the sodium pump gamma subunit: Effect of the pathological mutation G41R. *J Mol Biol*. 2002. **322**(583–590).
187. Sequence dependence of BNIP3 transmembrane domain dimerization implicates side-chain hydrogen bonding and a tandem GxxxG motif in specific helix-helix interactions. *J Mol Biol*. 2006. **364**(974–990).
188. Dimerization of the transmembrane domain of amyloid precursor proteins and familial Alzheimer's disease mutants. *BMC Neurosci*. 2008. **9**(17).
189. A clustering approach to multireference alignment of single-particle projections in electron microscopy. *J Struct Biol*. 2010. **171**(197–206).
190. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol*. 1999. **128**(82–97).
191. How good is automated protein docking?. *Proteins: Structure, Function, and Bioinformatics*. 2013. **81**(2159–2166).

192. PIPER: An FFT-based protein docking program with pairwise potentials. *Proteins: Structure, Function, and Bioinformatics*. 2006. **65**(392–406).
193. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics*. 2004. **20** (45–50).
194. ClusPro: a fully automated algorithm for protein-protein docking. *Nucleic Acids Research*. 2004. **32**(Web Server issue (W96–W99)).
195. Nanocrystals: Building blocks for modern materials design. *Endeavour*. 1997. **21**(56-60).
196. Biological applications of colloidal nanocrystals. *Nanotechnology*. 2003. **14**(R15).
197. Controlling the crystalline three-dimensional order in bulk materials by single-wall carbon nanotubes. *Nature communications*. 2014. **5**(3763).
198. Developmental toxicity in zebrafish (*Danio rerio*) embryos after exposure to manufactured nanomaterials: buckminsterfullerene aggregates (nC60) and fullerol. *Environmental toxicology and chemistry / SETAC*. 2007. **26**(976-9).
199. Bioinspired chemistry: Diversity for self-assembly. *Nat Chem*. 2010. **2**(1010-1).
200. Self-assembling peptides and proteins for nanotechnological applications. *Curr Opin Struct Biol*. 2004. **14**(480-6).
201. Protein engineering approaches to biomaterials design. *Curr Opin Biotechnol*. 2005. **16**(422-426).
202. Designing materials for biology and medicine. *Nature*. 2004. **428**(487-492).
203. Genetically encoded self-assembly of large amyloid fibers. *Biomater Sci*. 2014. **2**(560-566).
204. Materials Assembly and Formation using Engineered Polypeptides. *Annu Rev Mater Res*. 2004. **34**(373-408).
205. Assembly of gold nanoparticles using genetically engineered polypeptides. *Small*. 2005. **1**(698–702).
206. Directed self-assembly of densely packed gold nanoparticles. *Langmuir*. 2012. **28**(16782–16787).
207. DNA discrete modified gold nanoparticles. *Nanoscale*. 2011. **3**(4015–4021).
208. DNA-based self-assembly of chiral plasmonic nanostructures with tailored optical response. *Nature*. 2012. **483**(311-314).
209. Molecular biomimetics: nanotechnology and bionanotechnology using genetically engineered peptides. *Philos Trans Math Phys Eng Sci*. 2009. **367**(1705–1726).
210. Rationally engineering natural protein assemblies in nanobiotechnology *Curr Opin Biotechnol*. 2011. **22**(485-491).
211. Molecular thinking for nanoplasmonic design. *ACS Nano*. 2012. **6**(3655–3662).
212. Optical materials: a new twist for nonlinear optics. *Nat Mater*. 2002. **1**(209–210).
213. A chiral route to negative refraction. *Science*. 2004. **306**(1353–1355).
214. Chiral assembly of gold nanorods with collective plasmonic circular dichroism response. *Soft Matter*. 2011. **7**(8370).
215. Self-alignment of plasmonic gold nanorods in reconfigurable anisotropic fluids for tunable bulk metamaterial applications. *Nano Lett*. 2010. **10**(1347–1353).
216. Designed TPR modules as novel anticancer agents. *ACS Chem. Biol*. 2008. **3**(161-166).
217. Screening libraries to identify proteins with desired binding activities using a split-GFP reassembly assay. *ACS Chem. Biol*. 2010. **5**(553-62).
218. Controlled Nanometric Fibers of Self-Assembled Designed Protein Scaffolds. *Nanoscale* 2014. **6**(10982–10988.).
219. Assembly of designed protein scaffolds into monolayers for nanoparticles patterning. *Colloids & Surfaces B: Biointerfaces*. 2015. **Submitted**(
220. Nanostructured Functional Films from Engineered Repeat Proteins. *J. R. Soc. Interface* 2013. **10**(20130051.).

221. Au-Nanoparticle Nanowires Based on DNA and Polylysine Templates. *Angew Chemie Int Ed*. 2002. **41**(2323–2327).
222. Peptide conjugates for directing the morphology and assembly of 1d nanoparticle superstructures. *Chem-a EurJ*. 2014. **20**(941–945).
223. A new peptide-based method for the design and synthesis of nanoparticle superstructures: nanoparticle double helices construction of highly-ordered gold. *J Am Chem Soc Rev*. 2008. 1-15).
224. Protein- and peptide-directed syntheses of inorganic materials. *Chem Rev*. . 2008. **108**(4935–4978).
225. Biotemplated synthesis of pzt nanowires. *Nano Lett*. 2013. **13**(6197–6202).
226. Characterization of gold nanoparticle binding to microtubule filaments. *Mater Sci Eng C*. 2010. **30**(20-26).
227. Telomerase-generated templates for the growing of metal nanowires. *Nano Lett*. 2004. **4**(787–792).
228. Cowpea mosaic virus as a scaffold for 3-d patterning of gold nanoparticles. *Nano Lett*. 2004. **4**(867–870).
229. An engineered virus as a scaffold for three-dimensional self-assembly on the nanoscale. *Small*. 2005. **1**(702–706).
230. Virus-based toolkit for the directed synthesis of magnetic and semiconducting nanowires. *Science*. 2004. **303**(213–217).
231. Programmable assembly of nanoarchitectures using genetically engineered viruses. *Nano Lett*. 2005. **5**(1429–1434).
232. Fabrication and characterization of gold nano-wires templated on virus-like arrays of tobacco mosaic virus coat proteins. *nanotechnol*. 2013. **24**(025605).
233. DNA-Based Self-Assembly of Chiral Plasmonic Nanostructures with Tailored Optical Response - Supp Infos *Nature* 2012. **483**(311–314).
234. Peptide-Based Methods for the Preparation of Nanostructured Inorganic Materials. *Angew Chem Int Ed Engl* 2010. **49**(1924–1942).
235. Cowpea Mosaic Virus as a Scaffold for 3-D Patterning of Gold Nanoparticles. *Nano Lett*. 2004. **4**(867–870).
236. Oil–water interfacial self-assembly: a novel strategy for nanofilm and nanodevice fabrication. *Chem Soc Rev*. 2012. **41**(1350–1362).
237. Nanoelectrochemistry: metal nanoparticles, nanoelectrodes, and nanopores. *Chem Rev*. 2008. **108**(2688-2720).
238. Repeat-Proteins Films Exhibit Hierarchical Anisotropic Mechanical Properties. *Biomacromolecules* 2015. **16** (706–714).
239. Sub-nanometre sized metal clusters: from synthetic challenges to the unique property discoveries. *Chem Soc Rev*. 2012. **41**(3594-3623).
240. Nanostructured materials for advanced energy conversion and storage devices. *Nat Mater*. 2005. **4**(366-377).
241. Design strategies for organic semiconductors beyond the molecular formula. *Nat Chem*. 2012. **4**(699-704).
242. Controlling the crystalline three-dimensional order in bulk materials by single-wall carbon nanotubes. *Nat Commun*. 2014. **5**(3763).
243. Highly ordered n/p-co-assembled materials with remarkable charge mobilities. *J Am Chem Soc*. 2015. **137**(893-897).
244. DNA-templated assembly of dyes and extended π -conjugated systems. *Chem Commun (Camb)* 2011. **47**(4340-4347).
245. Evidence for charge transfer in doped carbon nanotube bundles from Raman scattering. *Nature*. 1997. **388**(257-259).

246. Stable dispersion of single wall carbon nanotubes in polyimide: the role of noncovalent interactions. *Chem Phys Lett*. 2004. **391**(207-211).
247. From self-assembly to noncovalent synthesis of programmable porphyrins' arrays in aqueous solution. *Chem Commun (Camb)*. 2012. **25**(8165-8176).
248. Self-assembling porphyrin-modified peptides. *Org Lett*. 2005. **7**(2559-2561).
249. Multifunctional porous microspheres based on peptide-porphyrin hierarchical co-assembly. *Angew Chem Int Ed Engl*. 2014. **53**(2366-2370).
250. Porphyrin-peptoid conjugates: face-to-face display of porphyrins on peptoid helices. *Org Lett*. 2013. **15**(1670-1673).
251. Pyrene-stacked nanostructures constructed in the recombinant tobacco mosaic virus rod scaffold. *Chemistry*. 2006. **12**(3735-3740).
252. Porphyrin light-harvesting arrays constructed in the recombinant tobacco mosaic virus scaffold. *Chemistry*. 2007. **13**(8660-8666).
253. Alpha,gamma-peptide nanotube templating of one-dimensional parallel fullerene arrangements. *J Am Chem Soc*. 2009. **131**(11335-11337).
254. Hierarchical self-assembly of semiconductor functionalized peptide α -helices and optoelectronic properties. *J Am Chem Soc*. 2011. **133**(8564-8573).
255. Polymer-peptide templates for controlling electronic interactions of organic chromophores. *J Mater Chem C*. 2013. **1**(4836-4845).
256. Segregated and alternately stacked donor/acceptor nanodomains in tubular morphology tailored with zinc porphyrin-C60 amphiphilic dyads: clear geometrical effects on photoconduction. *J Am Chem Soc*. 2012. **134**(2524-2527).
257. DNA as supramolecular scaffold for porphyrin arrays on the nanometer scale. *J Am Chem Soc*. 2007. **129**(15319-15329).
258. Introducing structural flexibility into porphyrin-DNA zipper arrays. *Org Biomol Chem*. 2011. **9**(777-782).
259. Photoinduced charge transfer in donor-acceptor (DA) copolymer: fullerene bis-adduct polymer solar cells. *ACS Appl Mater Interfaces*. 2013. **5**(861-868).
260. Self-Assembling Porphyrin-Modified Peptides. *Organic Letters*. 2005. **7**(2559-2561).
261. Optically Transparent Recombinant Silk-Elastinlike Protein Polymer Films. *J Phys Chem B* 2011. **115**(1608-1615).
262. Porphyrin Light-Harvesting Arrays Constructed in the Recombinant Tobacco Mosaic Virus Scaffold. *Chem Eur J*. 2007. **13**(8660-8666).
263. Enhancement of Phosphorescence Ability upon Aggregation of Dye Molecules. *J Chem Phys*. 1958. **28**(721).
264. The exciton model in molecular spectroscopy. *Pure Appl Chem Mater*. 1965. **11**(371-392).
265. Study of internal structure of meso-tetrakis (4-sulfonatophenyl) porphine J-aggregates in solution by fluorescence microscope imaging in a magnetic field. *Langmuir*. 2006. **22**(7600-7604).
266. Influence of Supramolecular Organization on Energy Transfer Properties in Chiral Oligo(p-phenylene vinylene) Porphyrin Assemblies. *J Am Chem Soc*. 2007. **129**(9818-9828).
267. J-aggregates: from serendipitous discovery to supramolecular engineering of functional dye materials. *Angew Chem Int Ed Engl*. 2011. **50**(3376-3410).
268. Introducing structural flexibility into porphyrin-DNA zipper arrays *Org Biomol Chem*. 2011. **9**(777-782).
269. Assembling multiporphyrin stacks inside the DNA double helix. *Bioconjugate chemistry*. 2014. **25**(1785-93).
270. Porphyrins: Powerful Chromophores for Structural Studies by Exciton-Coupled Circular Dichroism. *J Am Chem Soc*. 1995. **117**(7021-7022).

271. Theoretical analysis of the porphyrin-porphyrin exciton interaction in circular dichroism spectra of dimeric tetraarylporphyrins. *J Am Chem Soc.* 2003. **125**(7613-7628).
272. The packing of [alpha]-helices: simple coiled-coils. *Acta Crystallogr.* 1953. **6**(689-697).
273. Charge carrier mobility in organic molecular materials probed by electromagnetic waves. *Physical Chemistry Chemical Physics.* 2014. **16**(11093-11113).
274. Dynamics of photogenerated charge carrier and morphology dependence in polythiophene films studied by in situ time-resolved microwave conductivity and transient absorption spectroscopy. *Journal of Photochemistry and Photobiology A: Chemistry.* 2007. **186**(158-165).
275. Pi-Cation-radical formation following visible light photolysis of porphyrins in frozen solution using alkyl chlorides or quinones as electron acceptors. *Inorganic Chemistry.* 1985. **24**(2440-2447).
276. Wide-Range 2D Lattice Correlation Unveiled for Columnarly Assembled Triphenylene Hexacarboxylic Esters. *Angewandte Chemie International Edition.* 2012. **51**(7990-7993).
277. The devil and holy water: protein and carbon nanotube hybrids. *Acc Chem Res.* 2013. **46**(2454-2463).
278. Protein-Conjugated Carbon Nanomaterials for Biomedical Applications. *Carbon* 2015. **95**(767-779).
279. Structure-based carbon nanotube sorting by sequence-dependent DNA assembly. *Science.* 2003. **302**(1545-1548).
280. Ordered DNA wrapping switches on luminescence in single-walled nanotube dispersions. *J Am Chem Soc.* 2008. **130**(12734-1244).
281. Sequence-independent helical wrapping of single-walled carbon nanotubes by long genomic DNA. *Nano Lett.* 2006. **6**(159-164).
282. Bionano donor-acceptor hybrids of porphyrin, ssDNA, and semiconductive single-wall carbon nanotubes for electron transfer via porphyrin excitation. *J Am Chem Soc.* 2011. **133**(19922-19930).
283. Interfacing proteins with graphitic nanomaterials: from spontaneous attraction to tailored assemblies. *Chem Soc Rev.* 2015. **44**(6916-6953).
284. Computational design of virus-like protein assemblies on carbon nanotube surfaces. *Science.* 2011. **332**(1071-1076).
285. Peptides with selective affinity for carbon nanotubes. *Nat Mater.* 2003. **2**(196-200).
286. Peptide-Mediated Formation of Single-Wall Carbon Nanotube Composites. *Nano Lett.* 2006. **6**(40-44).
287. Solubilization of single-walled carbon nanotubes using a peptide aptamer in water below the critical micelle concentration. *Langmuir.* 2015. **31**(3482-3488).
288. Plugging into proteins: poisoning protein function by a hydrophobic nanoparticle. *ACS Nano.* 2010. **4**(7508-7514).
289. Functionalized Fe-Filled Multiwalled Carbon Nanotubes as Multifunctional Scaffolds for Magnetization of Cancer Cells. *Adv Funct Mater.* 2013. **23**(3173-3184).
290. Effect of chirality, length and diameter of carbon nanotubes on the adsorption of 20 amino acids: a molecular dynamics simulation study. *Mol Simul.* 2013. **40**(392).
291. Effect of electron-donating and electron-withdrawing groups on peptide/single-walled carbon nanotube interactions. *J Am Chem Soc.* 2007. **129**(14724-14732).
292. Extraordinary Sensitivity of the Electronic Structure and Properties of Single-Walled Carbon Nanotubes to Molecular Charge-Transfer. *J Phys Chem.* 2008. **112**(13053-13056).
293. Artificial photoactive proteins. *Photoynth Res.* 2008. **98**(677-685).

294. Preparation and characterization of nanomaterials for sustainable energy production. *ACS Nano*. 2010. **4**(5517-26).
295. Artificial photosynthesis for solar fuels. *Faraday Discuss*. 2012. **155**(357-376).
296. The hydrogen economy. *Tarcher, London*. 2002.
297. Semiconductor-based photocatalytic hydrogen generation. *Chem Rev*. 2010. **110**(6503-6570).
298. De novo design of functional proteins: Toward artificial hydrogenases. *Biopolymers*. 2013. **100**(558-571).
299. Hydrogenases. *Chem Rev*. 2014. **114**(4081-4148).
300. X-ray crystal structure of the Fe-only hydrogenase (Cpl) from *Clostridium pasteurianum* to 1.8 angstrom resolution. *Science*. 1998. **282**(1853-1858).
301. Classification and phylogeny of hydrogenases. *FEMS Microbiol Rev*. 2001. **25**(455-501).
302. Structure and function of photosystems I and II. *Annu Rev Plant Biol*. 2006. **57**(521-565).
303. Biotechnological applications of hydrogenases. *Curr Opin Biotechnol*. 2004. **15**(343-348).
304. Biochemistry. A natural choice for activating hydrogen. *Science*. 2008. **321**(498-499).
305. Structure/function relationships of [NiFe]- and [FeFe]-hydrogenases. *Chem Rev*. 2007. **107**(4273-4303).
306. Optimizing the specificity of nucleic acid hybridization. *Nat Chem*. 2012. **22**(208-214).
307. The organometallic active site of [Fe] hydrogenase: models and entatic states. *Proc Natl Acad Sci USA*. 2003. **1**(3683-2688).
308. A synthetic nickel electrocatalyst with a turnover frequency above 100,000 s⁻¹ for H₂ production. *Science*. 2011. **12**(863-866).
309. Structural and functional analogues of the active sites of the [Fe]-, [NiFe]-, and [FeFe]-hydrogenases. *Chem Rev*. 2009. **109**(2245-2274).
310. Modulating repeat protein stability: the effect of individual helix stability on the collective behavior of the ensemble. *Protein Sci*. 2011. **20**(1042-1047).
311. Photo-induced hydrogen production in a helical peptide incorporating a [FeFe] hydrogenase active site mimic. *Chem Commun (Camb)*. 2012. **48**(9816-9818).
312. Evolution and unique bioenergetic mechanisms in oxygenic photosynthesis. *Curr Opin Chem Biol*. 2006. **10**(91-100).
313. A De Novo Designed 2[4Fe-4S] Ferredoxin Mimic Mediates Electron Transfer. *J Am Chem Soc*. 2014. **136**(17343-17349).
314. Protein Control of Redox Potentials of Iron-sulfur Proteins. *Chem Rev*. 1996. **96**(2491-2514).
315. POVME: an algorithm for measuring binding-pocket volumes. *J Mol Graph Model*. 2011. **29**(773-776).
316. Chemical rescue of a site-modified ligand to a [4Fe-4S] cluster in PsaC, a bacterial-like dicluster ferredoxin bound to Photosystem I. *Biochim Biophys Acta*. 2007. **1767**(712-724).
317. Structure, function, and formation of biological iron-sulfur clusters. *Annu Rev Biochem*. 2005. **74**(247-281).
318. Biogenesis of iron-sulfur clusters in photosystem I: holo-NfuA from the cyanobacterium *Synechococcus* sp. PCC 7002 rapidly and efficiently transfers [4Fe-4S] clusters to apo-PsaC in vitro. *J Biol Chem*. 2008. **17**(28426-28435).
319. Flexibility in anaerobic metabolism as revealed in a mutant of *Chlamydomonas reinhardtii* lacking hydrogenase activity. *J Biol Chem*. 2009. **284**(7201-7213).
320. Insights into [FeFe]-hydrogenase structure, mechanism, and maturation. *Structure*. 2011. **10**(1038-1052).

321. Iron-sulfur clusters: nature's modular, multipurpose structures. *Science*. 1997. **277**(653-659).
322. De novo design of an artificial bis[4Fe-4S] binding protein. *Biochemistry*. 2013. **52**(7586-7594).
323. Proton coupling to [4Fe-4S](2+/+) and [4Fe-4Se](2+/+) oxidation and reduction in a designed protein. *J Am Chem Soc*. 2002. **124**(6826-6827).
324. De novo design of a non-natural fold for an iron-sulfur protein: alpha-helical coiled-coil with a four-iron four-sulfur cluster binding site in its central core. *Biochim Biophys Acta*. 2010. **1797**(406-413).
325. Empirical and computational design of iron-sulfur cluster proteins. *Biochim Biophys Acta*. 2012. **1817**(1256-1262).
326. *The PyMOL Molecular Graphics System*. 2002. DeLano Scientific: San Carlos, CA, (USA)
327. Proteins containing 4Fe-4S clusters: an overview. *Annu Rev Biochem*. 1980. **49**(139-161).
328. Introduction of a [4Fe-4S (S-cys)₄]^{+1,+2} iron-sulfur center into a four-alpha helix protein using design parameters from the domain of the Fx cluster in the Photosystem I reaction center. *Protein Sci*. 1997. **6**(340-346).
329. Synthesis and characterization of de novo designed peptides modelling the binding sites of [4Fe-4S] clusters in photosystem I. *Biochim Biophys Acta*. 2009. **1787**(995-1008).
330. Proton coupling to [4Fe-4S](2+/+) and [4Fe-4Se](2+/+) oxidation and reduction in a designed protein. *J Am Chem Soc*. 2002. **124**(6826-6827).
331. The rational design and construction of a cuboidal iron-sulfur protein. *Proc Natl Acad Sci USA*. 1997. **94**(6635-6640).
332. Helix-loop-helix peptides as scaffolds for the construction of bridged metal assemblies in proteins: the spectroscopic A-cluster structure in carbon monoxide dehydrogenase. *J Am Chem Soc*. 2001. **123**(10255-110264).
333. Ferredoxin and ferredoxin-heme maquettes. *Proc Natl Acad Sci USA*. 1996. **93**(15041-15046).
334. Conversion of light to chemical free energy. I. Porphyrin-sensitized photoreduction of ferredoxin by glutathione. *J Biol Chem*. 1969. **10**(1720-1728).
335. On the nature of the spin coupling between the iron-sulfur clusters in the eight-iron ferredoxins. *J Biol Chem*. 1974. **249**(4326-4328).
336. Function and biogenesis of iron-sulphur proteins. *Nature*. 2009. **460**(831-838).
337. A porphyrin-stabilized iridium oxide water oxidation catalyst. *Can J Chem*. 2011. **89**(152-157).
338. 26.5–40 GHz Ka-band pulsed EPR spectrometer. *Concepts Magn Reson*. 2006. **29B**(125–136).
339. [Fe₄S₄]⁻ and [Fe₃S₄]-cluster formation in synthetic peptides. *Biochim Biophys Acta*. 2011. **1807**(1414–1422).
340. Spectrophotometric determination of serum iron at the submicrogram level with a new reagent (ferrozine). *Anal Biochemistry*. 1971. **40**(450–458).
341. Multicomponent redox gradients on photoactive electrode surfaces. *Chem Commun (Camb)*. 2004. 726-727).
342. Redox properties of several bacterial ferredoxins using square wave voltammetry. *J Biol Chem*. 1990. **265**(14371-14376).
343. Expanding the genetic code of Escherichia coli. *Science*. 2001. **292**(498-500).
344. An enhanced system for unnatural amino acid mutagenesis in E. coli. *J Mol Biol*. 2010. **395**(361-374).
345. How to Measure and Predict the Molar Absorption Coefficient of a Protein. *Protein Sci*. 1995. **4**(2411–2423).

346. Immobilization of peroxidase glycoprotein on gold electrodes modified with mixed epoxy-boronic Acid monolayers. *J Am Chem Soc.* 2002. **124**(12845-53).
347. Concentration Dependence of Aggregate Formation upon Adsorption of 5-(Octyldithio)-2-nitrobenzoic Acid on Gold Electrodes. *Langmuir.* 2001. **16**(9804-9811).
348. Modulation of Electroenzymatic NADPH Oxidation through Oriented Immobilization of Ferredoxin:NADP⁺ Reductase onto Modified Gold Electrodes. *J Am Chem Soc.* 2000. **122**(9808-9817).
349. Interfaces and thin films as seen by bound electromagnetic waves. *Annu Rev Phys Chem.* 1998. **49**(569-638).
350. Surface Plasmons on Smooth and Rough Surfaces and on Gratings. *Springer-Verlag.* 1998. Berlin).
351. Quantitative determination of surface concentration of protein with surface plasmon resonance using radiolabeled proteins. *Journal of Colloid and Interface Science.* 1991. **143**(513-526).
352. Jumping mode scanning force microscopy. *Appl Phys Lett* 1998. **73**(3300-3302).
353. WSXM: A software for scanning probe microscopy and a tool for nanotechnology. *Rev Sci Instrum.* 2007. **78**(13705-13713).
354. Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol.* 2003. **142**(334-347).
355. Xmipp: An Image Processing Package for Electron Microscopy. *J Struct Biol.* 1996. **116**(237-240).
356. Maximum-likelihood multi-reference refinement for electron microscopy images. *J Mol Biol.* 2005. **348**(139-149).
357. A clustering approach to multireference alignment of single-particle projections in electron microscopy. *J Struct Biol.* 2010. **171**(197-206).
358. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol.* 1999. **128**(82-97).
359. A study of the nucleation and growth processes in the synthesis of colloidal gold. *Discuss Faraday Soc.* 1951. **11**(55-75).
360. DNA and aptamer stabilized gold nanoparticles for targeted delivery of anticancer therapeutics. *Nanoscale.* 2014. **6**(7436-7442).
361. Synthesis of Poly(β -amino ester)s with Thiol-Reactive Side Chains for DNA Delivery. *J Am Chem Soc.* 2006. **128**(12726-12734).
362. Stability of single- and few-molecule junctions of conjugated diamines. *J Am Chem Soc.* 2013. **135**(5420-5426).
363. De novo design of an artificial bis[4Fe-4S] binding protein. *Biochemistry.* 2013. **52**(7586-7594).
364. Synthesis and characterization of de novo designed peptides modelling the binding sites of [4Fe-4S] clusters in photosystem I. *Biochim Biophys Acta.* 2009. **1787**(995-1008).
365. Design of three-dimensional domain-swapped dimers and fibrous oligomers. *Proc Natl Acad Sci U S A.* 2001. **98**(1404-1409).
366. On the nature of the spin coupling between the iron-sulfur clusters in the eight-iron ferredoxins. *J Biol Chem.* 1974. **249**(4326-4328).
367. Spectrophotometric determination of serum iron at the submicrogram level with a new reagent (ferrozine). *Anal Biochem.* 1971. **40**(450-458).

Appendix 1.

**Controlled nanometric fibers of self-assembled
designed protein scaffolds**

CrossMark
click for updates

Controlled nanometric fibers of self-assembled designed protein scaffolds†

Sara H. Mejias,^{ab} Begoña Sot,^{ab} Raul Guantes^c and Aitziber L. Cortajarena^{*ab}Cite this: *Nanoscale*, 2014, 6, 10982Received 5th March 2014
Accepted 21st May 2014

DOI: 10.1039/c4nr01210k

www.rsc.org/nanoscale

The use of biological molecules as platforms for templating and nanofabrication is an emerging field. Here, we use designed protein building blocks based on small repetitive units (consensus tetratricopeptide repeat – CTPR) to generate fibrillar linear nanostructures by controlling the self-assembly properties of the units. We fully characterize the kinetics and thermodynamics of the assembly and describe the polymerization process by a simple model that captures the features of the structures formed under defined conditions. This work, together with previously established functionalization potential, sets up the basis for the application of these blocks in the fabrication and templating of complex hybrid nanostructures.

Advances in nanotechnology require methods to control the structure of materials from the molecular to the macroscopic scales. Self-assembly driven by specific biomolecular interactions is a powerful approach for nanostructural templating, and bottom-up design and fabrication of biomaterials.^{1–3} Molecular self-assembly is a process in which complex structures with specific functions are constructed from simple molecular building blocks. Self-assembly of materials based on peptides, proteins and nucleic acids has been recently explored.^{4–13} In particular, the functional versatility that protein templates can encode is a clear advantage with respect to nucleic acid-based nanostructures.¹⁴ Understanding the structural and reactive properties of the building blocks and controlling the interactions that drive their self-assembly are key steps to successfully generate protein-based functional nanostructures.

In this work we focus on repeat proteins as our building block units for molecular self-assembly. These blocks possess properties that make them suitable for templating, such as defined structure and stability. Repeat proteins are composed of tandem arrays of the same small structural motif. Their simple architecture is based on well-defined short-range local interactions between the repeated domains, therefore these modules are ideal for biomolecular-engineering approaches.¹⁵ In particular, we work with a designed consensus tetratricopeptide repeat (CTPR) (Fig. 1A).^{16,17}

Each CTPR unit is a 34 amino acid sequence that folds into a helix-turn-helix motif defined by only 8 conserved residues.¹⁷ This unit can be considered as a building block whose properties, such as stability¹⁸ and function,^{19–21} can be engineered separately as previously shown. Identical CTPRs can be expressed in arrays to generate proteins with different numbers of repeats (from 2 to 20) (Fig. 1) that are much more stable than TPR domains existing in nature. Additionally, the stability of these modules can be finely tuned by the number of repeats²² and by the sequence of the repeated units.¹⁸ These modules have been proven useful in the fabrication of hydrogels^{23,24} and ordered protein films,²⁵ revealing their inherent self-assembly properties. In order to utilize CTPR proteins as templates for nanofabrication and patterning we need to characterize their assembly and be able to control and predict it under different experimental conditions.

In the present study we used the CTPR20 protein, the longest array constructed with 20 TPR consensus units, as our molecular unit (Fig. 1B).²⁶ CTPR long arrays display an ordered superhelical structure with regular geometry and nanometric periodicity in which 8 repeats exactly complete a full superhelical turn of about 7.2 nanometers (Fig. 1B). Individual building blocks can then be assembled using a bottom up approach to generate more complex structures (Fig. 1C). It has been previously shown that the TPR modules have a tendency to self-assemble through “head-to-tail” interactions in crystal forms,^{26,27} in solid films²⁵ and also in solution.²⁸ In the crystal structures, individual molecules stack “head-to-tail” to form

^aIMDEA-Nanociencia, Cantoblanco, 28049 Madrid, Spain. E-mail: aitziber.lopezcortajarena@imdea.org

^bCentro Nacional de Biotecnología (CNB-CSIC) – IMDEA Nanociencia Associated Unit “Unidad de Nanobiotechnología”, Cantoblanco, 28049 Madrid, Spain

^cDepartment of Condensed Matter Physics, Material Science Institute “Nicolás Cabrera” and Institute for Condensed Matter Physics (IFIMAC), Facultad de Ciencias, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain

† Electronic supplementary information (ESI) available: Detailed Materials and methods section. See DOI: 10.1039/c4nr01210k

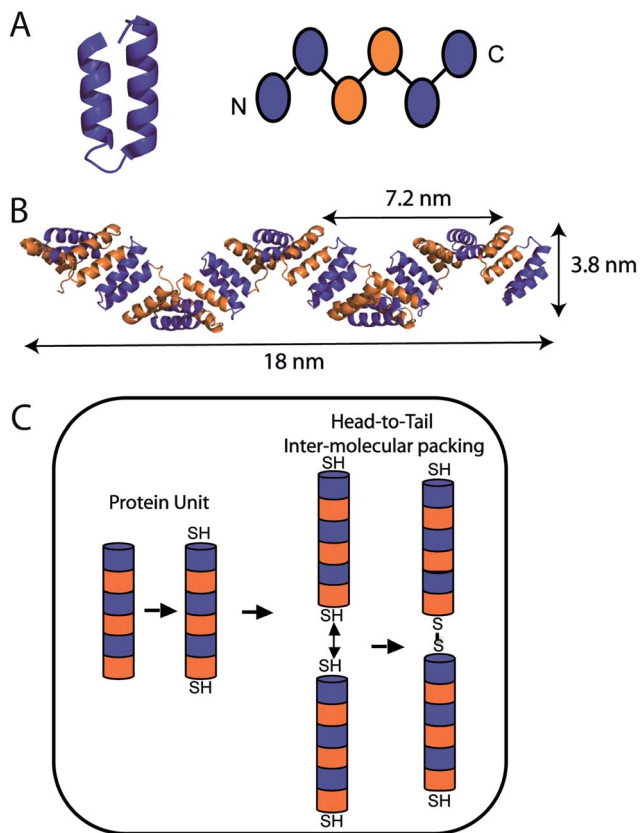


Fig. 1 CTPR proteins as building blocks for controlled polymerization. (A) Structure of a single helix-turn-helix TPR domain (34 amino acids). Schematic representation of TPR packing interactions in which the first repeat composed of two helices is highlighted in blue, and alternating repeats are coloured in orange and blue. (B) Crystal structure of CTPR20 in which each repeat is coloured either orange or blue, alternatively (PDB ID: 2AVP).²⁶ The superhelical repeated structure of the 20 repeats is clearly shown in this lateral view of the protein. The protein dimensions and the dimension of one superhelical turn are indicated. (C) Schematics showing the bottom-up strategy to generate protein-based polymeric nanofibers. CTPR20 protein units are formed by 20 identical CTPRs (not all shown here), and the units are modified to present two unique cysteine residues at the N- and C-terminal ends (left). As a result of the “head-to-tail” inter-molecular interactions the protein units are hypothesized to polymerize with cysteine mediated di-sulfide bonds acting as staples of the interaction (right). The inter-molecular packing interfaces in the polymers are the same as the intra-molecular interfaces between two repeats in the same molecule, as shown in panel A.

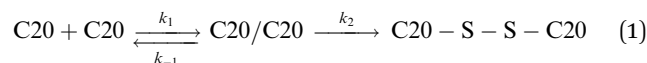
virtually continuous superhelices along the crystal, in which the inter-molecular packing interface is identical to the intra-molecular repeat-repeat interface (Fig. 1).²⁶ According to the schematics in Fig. 1, linear higher order structures could assemble by specific contacts between superhelices, similarly to the ones present in crystalline forms. This interface is inherent to the repeat units so the unit sequence does not need to be modified ensuring the preservation of the structure, stability and function within the arrays.

Here we generate and fully characterize long linear polymers from CTPR20 building blocks through a defined “head-to-tail” interaction. This approach allows us to construct the simplest

higher order structures based on CTPR blocks and thin linear nanofibers with defined features. CTPR proteins can encode binding functionalities within their concave face,^{29,30} with 3-TPRs being the most prevalent functional unit. Therefore polymeric nanostructures could potentially include many functional sites. The combination of the controlled assembly with the specific design of the stability and functional properties of the building blocks provides a versatile platform for the nanofabrication of bio-inspired protein based functional materials and devices. For example, a next step will be to use these platforms to organize different molecular elements in nanoscale devices.

The ability of the designed CTPR20 protein units to form polymers through “head-to-tail” inter-molecular interactions was monitored using a designed CTPR20 protein without the solvating helix and one cysteine at each terminal end of the protein (Fig. 1). A solvating helix was added to the C-terminal of CTPR20 in order to increase the protein solubility,^{16,17} and its removal will permit inter-molecular repeat packing analogous to the intra-molecular packing and cysteines will staple these interactions. Cysteine-modified interfaces have been already applied to generate nano-scale assemblies.^{31,32} Our designed variant is made to encode directional “head-to-tail” packing according to the schematic picture shown in Fig. 1C without perturbing the structure, stability and functional capabilities of the CTPR domains.

The formation of CTPR polymers is facilitated by the kinetics of association between the N-terminal and C-terminal subunits (*i.e.*, the “head-to-tail” interaction). The polymerization includes the recognition of two molecules through their packing interfaces and the formation of a disulfide bond that acts as a staple of the interaction to generate longer polymers. Therefore, the polymerization is expected to follow the mechanism described in eqn (1):



where k_1 and k_{-1} are the rate constants of the “head-to-tail” interaction and dissociation respectively, and k_2 is the rate constant of the formation of the disulfide bond.

The polymerization in solution of Cys-CTPR20-Cys proteins is monitored by the increase in the average hydrodynamic size by dynamic light scattering (DLS). When a freshly reduced protein sample is placed in the DLS instrument at defined protein concentration and temperature the polymerization kinetics of the CTPR20 proteins can be clearly observed by the increase in the hydrodynamic diameter with time (Fig. 2A).

For CTPR20 proteins with no cysteines even though the “head-to-tail” interfacial recognition should take place, no significant polymer growth can be detected by DLS (Fig. 2A, circles) or fiber formation in the TEM image (Fig. 3A). The hydrodynamic diameter in Fig. 2A increases linearly with time as expected from the reaction scheme given by eqn (1), and can be fitted to estimate the polymerization rate (ESI^\dagger). These results show that the CTPR scaffold proteins in solution at relatively diluted protein concentrations (low μM) are able to polymerize and form larger structures. The CTPR

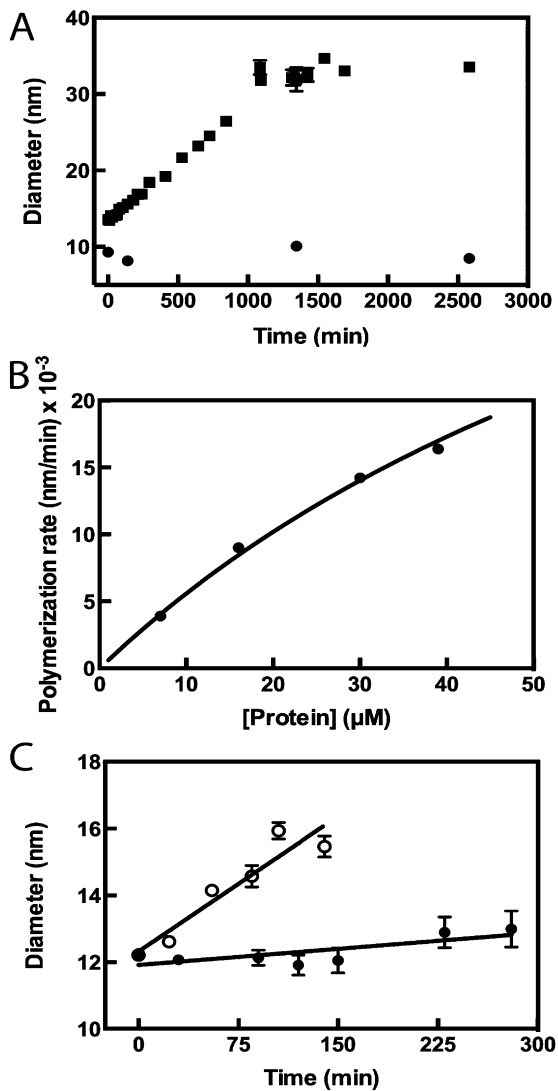


Fig. 2 CTPR20 protein polymerization kinetics. (A) Polymerization growth monitored by the increase in the size of the Cys-CTPR20-Cys polymers as a function of time at 40 μM protein concentration (solid squares). The control showing the size of CTPR20 without cysteines measured over time is shown in solid circles. The error bars show the standard deviation of three measurements, in some data points the error bars are not visible because they are smaller than the size of the symbols. (B) Effect of the protein concentration on the polymerization kinetics. Polymerization growth rates at different protein concentrations (7 μM, 16 μM, 30 μM and 40 μM) (filled circles) can be well fitted to the saturation eqn (S2) (ESI†) (solid line). (C) Effect of the temperature on the polymerization kinetics. The increase in the size of CTPR20 at 7 μM protein concentration is measured by DLS and plotted *versus* the time at two polymerization temperatures, 25 °C (filled circles) and 42 °C (empty circles). At 42 °C, the polymerization is about 10 times faster than at 25 °C. Error bars show the standard deviation of three measurements.

polymerization can be reversed by adding a reducing agent to the sample after polymerization saturation, and the hydrodynamic size of the sample returns to the initial value (13.94 ± 0.63 nm).

To quantitatively characterize in detail the polymerization process, we monitored the effect of the protein concentration in

the polymerization rate at 25 °C. First, we observed that the growth rates are dependent on the initial protein concentration, as expected (Fig. 2B). Since there is no measurable increase in the hydrodynamic size without cysteines (Fig. 2A, circles) we assume that the equilibrium in eqn (1) is shifted towards the monomeric forms, with the dissociation of the complex being faster than the association under our experimental conditions ($k_1 \lll k_{-1}$). The disulfide bond equilibrium under our experimental conditions can be assumed fully shifted to the disulfide bond formation, considering the described bond dissociation energy.³³ Additionally, the formation of cysteine bonds *in vitro* has been described to be relatively slow and the limiting step of many folding processes.³⁴ Therefore, we can assume that dissociation of the CTPR20 interface is much faster than the staple through the disulfide bond ($k_2 \lll k_{-1}$). Under these conditions, we can assume that the CTPR20 interface (complex C20/C20 in eqn (1)) is in the steady-state, and we fit the velocity of growth as a function of the protein concentration to a Michaelis-Menten-like function (ESI, eqn (S2)†), to obtain the effective maximum velocity, V_{\max} and the dissociation constant $K_D = k_{-1}/k_1$ (since the Michaelis constant $K = (k_{-1} + k_2)/k_1 \sim K_D$). These values are shown in Table 1.

We next analyzed the effect of temperature on the polymerization kinetics, repeating identical polymerization experiments at 42 °C. Fig. 2C compares the growth rates at 25 °C and 42 °C for an initial protein concentration of 7 μM. At 42 °C, the functional dependence of the growth velocity is also of Michaelis-Menten form, with the fitted maximum velocity about ten times faster than at 25 °C (Table 1).

The increase in the polymerization rate with the temperature is described by eqn (S3) (ESI†) from which the activation energy (E_a) of the polymerization is estimated to be around 25 kcal mol⁻¹, in the same range of spontaneous natural fiber growth such as collagen formation.³⁵ From the equilibrium constant (K_D) we can calculate the change in free energy of the “head-to-tail” interaction, ΔG_{exp} (Table 1), using eqn (S4) (ESI†). These equilibrium parameters are important for the use of these scaffolds under experimental conditions that ensure the thermodynamic stability of the nanostructures.

We can validate the values obtained for the inter-molecular “head-to-tail” interaction with those expected from previous detailed structural and stability data on CTPR proteins.^{18,22,26} As mentioned before, the new inter-molecular interface is identical to the intra-molecular repeat-repeat interface (Fig. 1) and therefore, the expected ΔG for “head-to-tail” packing can be calculated in different ways. First, we estimated ΔG based on a calculation of the free energy associated with the hydrophobic surface area burial upon binding, using the transfer free energy scales from hydrophobic to polar media³⁶ for each amino acid on the interface. Taking into account the change in the free energy of the amino acids in the interface between two monomers, this ΔG is -5.59 kcal mol⁻¹. Additionally, we estimated the inter-repeat coupling interaction from the previously published linear 1D-Ising model description of the folding and stability of CTPR proteins.^{18,37} This model dissects the experimental thermodynamic stability of the repeated proteins in two parameters, the intrinsic repeat stability (H) and the coupling

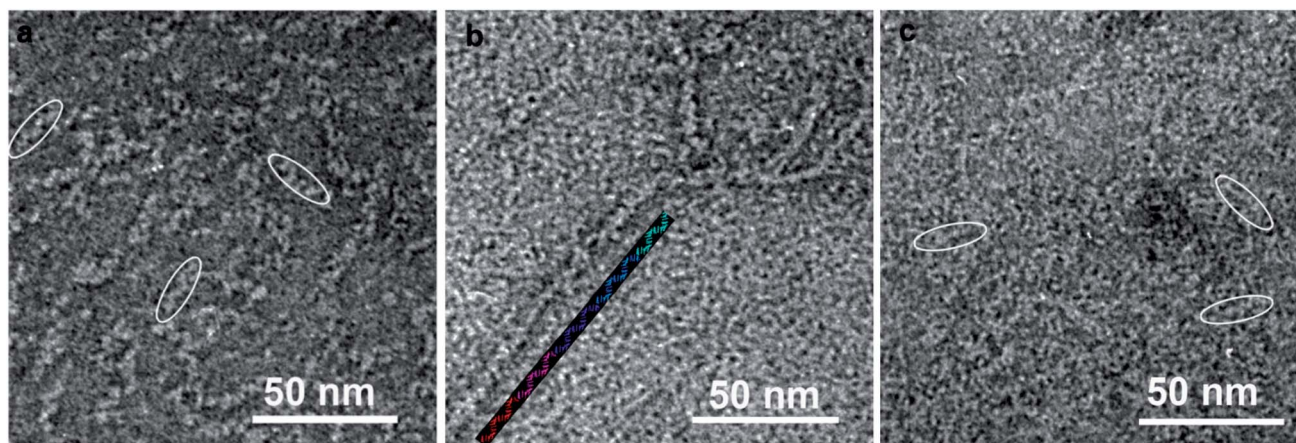


Fig. 3 Negative stain TEM imaging of the protein polymerization process. (A) Transmission electron micrograph of the CTPR20 monomer deposited on TEM grids. Individual CTPR20 monomers are circled. The particle size measured was the average of 5 molecules. (B) CTPR20 sample after polymerization saturation, where the linear polymeric structures are observed. The schematic arrangement of CTPR20 units in the linear fibers is shown with each CTPR20 unit displayed in different colors on approximately the scale of the fiber in the TEM image. (C) CTPR20 after polymerization treated with a reducing agent (1 mM DTT), where the polymeric structures dissociate, and monomers appear. Individual CTPR20 monomers are circled.

Table 1 Kinetic and thermodynamic parameters of CTPR building block polymerization^a

Temperature (°C)	V_{\max} (nm min ⁻¹)	K_D (μM)	ΔG_{exp} (kcal mol ⁻¹)
25	0.056	91.81	-5.50
42	0.435	90.49	-5.51

^a The maximum velocity of the polymerization process (V_{\max}), the dissociation constant for the inter-molecular interaction (K_D), and the difference in Gibbs energy (ΔG_{exp}) obtained from the CTPR polymerization experiments at two different temperatures: 25 °C and 42 °C.

energy between two adjacent repeats (J).^{18,37,38} The repeat-repeat interfacial free energy is $\Delta G = -5.45$ kcal mol⁻¹ considering a J coupling value of 2.3.¹⁸ These results show that the ΔG calculated from the polymerization experiments is in close agreement with the interfacial interaction energy calculated by two independent methods, and confirm that the driving force of the polymerization is the “head-to-tail” interaction between the C- and N-terminal repeats of two proteins.

In order to obtain structural information about the morphology of the higher order species observed by DLS we imaged the polymerized protein samples using transmission electron microscopy (TEM). First, as a control, CTPR20 protein monomers without cysteines were imaged. Fig. 3A shows the presence of individual rod-shaped CTPR20 monomers, of 18.8 ± 1.33 nm long and 3.6 ± 0.4 nm width, in close agreement with the dimensions of the CTPR20 protein calculated from the crystal structure.²⁶ The superhelical structure of the CTPR20 monomers is also observed. Then, protein samples after saturation of the polymerization growth were imaged. Fig. 3B clearly shows the presence of linear fibers. Their experimentally measured width was 3.6 ± 0.6 nm, in agreement with fibers formed by the linear assembly of CTPR20 monomers. These

fibers are significantly thinner than other protein-based amyloid and collagen fibers previously studied and comparable to hyperthin nanochains of 2–3 nm recently reported.¹²

To determine the specificity of the polymerization process through the disulfide bond staple of the “head-to-tail” interaction, the polymerized sample is reduced (Methods section, ESI†). The disruption of the polymeric fibers to monomers by the reducing agent is observed in Fig. 3C, indicating the reversibility of the assembly and that the polymers observed in Fig. 3B are the result of the designed controlled linear interaction.

During the polymerization experiments the average hydrodynamic size of the sample is measured over the time by DLS. To have a quantitative description of the polymerization process it is necessary to correlate the average size with the number of building block units. We have previously estimated the relationship between the number of repeats and the hydrodynamic size for CTPR proteins with different numbers of repeats (Fig. 4A).³⁹ The hydrodynamic size of CTPR proteins with different repeats was experimentally measured by fluorescence correlation spectroscopy, and protein sizes were also estimated from the crystal structures using the program Hydropro.⁴⁰ This relationship, shown in Fig. 4A, can be fitted to the scaling equation $R_h = 1.48N^{0.376}$ that is used to calculate the average number of repeats per polymer chain (N) from the values of average hydrodynamic radius (R_h). Additionally, these calculations allow us to estimate the average polymer size that would be expected from a defined polymerization process.

To quantify the self-assembly process and predict both the kinetics of chain growth and the chain length under different experimental conditions, we tested a simple model based on 1D polymer-like structure growth.⁴¹ Each protein fiber can grow through two N- and C-terminal ends and the probability of adding an extra molecule at each end would be the same because of the identical interaction interface. Therefore, we

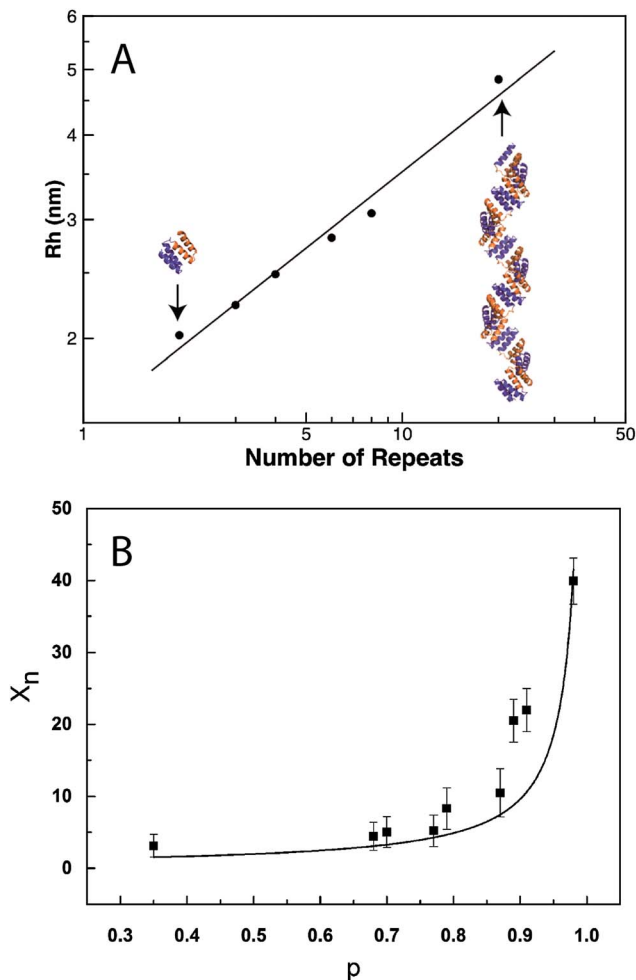


Fig. 4 (A) Hydrodynamic radius (R_h) of CTPR proteins as a function of the number of repeated units. The log–log plot is fitted to the relationship $R_h = 1.486N^{0.376}$. We also show the structures of CTPR proteins with 2 and 20 repeats with each repeat unit coloured blue or orange, alternatively. (B) Step growth polymerization model. The relationship between the degree of polymerization quantified as the average number of monomers in the polymer chains (X_n) and the reaction probability (or extent of reaction) quantified as the fraction of Cys ends bonded (p). Black squares represent the experimental data from DLS measurements. The solid line shows the step-growth model prediction ($X_n = 1/(1 - p)$). Error bars indicate the standard error of three measurements.

hypothesized that the process may occur following a simple statistical step-growth polymerization process. The step-growth model is a random process that proceeds *via* a step-by-step succession of elementary reactions between reactive sites and assumes equal reactivity of the functional groups.^{41,42} This model has been successfully used recently to describe the polymerization of inorganic nanoparticles.⁴³

CTPR modules can be considered bifunctional monomers with two identical functionalities one at each end. At different times of the polymerization reaction several parameters that define the polymer growth are calculated: (1) the average degree of polymerization (X_n) defined as the average number of CTPR20 monomers per polymer chain, calculated from the

average hydrodynamic size using the relationship described in Fig. 4A; (2) the fraction of functional groups bounded, or extent of reaction (p), estimated measuring the free cysteine groups as a function of polymerization time. For an ideal step growth polymerization these parameters are related by defined equations.⁴¹

Eqn (S5) (ESI[†]) gives the relationship between X_n and p shown in Fig. 4B: as p increases, X_n becomes larger until the saturation of the polymerization process at $p = 1$. The experimental values at different polymerization stages closely follow the behaviour of the ideal model.

Conclusions

In conclusion, we have shown that we can generate thin long linear protein nanochains with designed CTPR protein building blocks in solution at relatively low protein concentrations. In this work we took advantage of the intrinsic “head-to-tail” interactions encoded in the repeated sequence, and therefore no modification of the scaffold is required thus the structure and stability of the building blocks are preserved. Additionally, a simple disulfide bond staple is used to generate linear polymers. This approach is simpler than others described for polymerization of repeat protein scaffolds in which chemical ligation was used.²⁸ Our structures can reversibly disassemble into the monomeric units under reducing conditions while maintaining the structure of the building blocks.

We have described in detail the kinetics and thermodynamics of the interaction. This quantitative description provides the tools to rationally control the polymerization process by tuning experimental conditions such as protein concentration, temperature and polymerization time to achieve the desired size distributions of the polymeric chains and to ensure the thermodynamical stability of the higher order structures.

This work is an example of the molecular understanding and detailed biophysical characterization at the nanoscale of the assembly process of designed protein building blocks. Here, we set up the basis for fabrication and templating based on these protein scaffolds. We demonstrate that nanometric thin fibers can be built from simple building blocks with minimal modification of the units. The CTPR system provides several advantages compared to other protein based fibers reported: the design is simple without mutational or chemical modification of the blocks; the fiber formation is covalent but reversible; the stability of the CTPR units, and therefore the stability of the final structures, can be rationally designed; the building blocks can encode not only structural features but also diverse binding capabilities, avoiding extra functionalization steps of the nanostructures; the thin diameter and the superhelical conformation of the arrays grant a large exposed surface area and therefore potential functional binding sites; the structural periodicity of the CTPR superhelix can be used to define periodicity on the functional decoration; the modular structure permits modular combination of repeated blocks with different characteristics, constituting an unique versatile platform for nanofabrication. Finally, a predictive simple polymerization

model can be used to describe the self-assembly of the building blocks.

The full characterization of the assembly process and the possibility of introducing specific functionalities to these protein scaffolds enable the rational design and production of protein-based higher order functional assemblies, opening new doors to the use of these designed biomolecules as building blocks for patterning and nanofabrication. These nanostructures can be functionalized by introducing unique groups, with atomic resolution, at any position of the protein. Therefore, one direct application that we are exploring is their use as platforms to precisely organize at the nanoscale different components for applications in energy, electron transfer and nanoelectronics.

Acknowledgements

We thank M.P. Morales (ICMM-CSIC) for the use of the Zetasizer NanoZS instrument. Financial support for this research was obtained from the Spanish Ministerio de Economía y Competitividad (BIO2012-34835 A.L.C.), the European Commission International Reintegration Grant (IRG-246688) (A.L.C.) and Marie Curie COFUND “AMAROUT-Europe” Programme (A.L.C.). S.H.M. thanks IMDEA-Nanociencia for financial support through an “Ayuda de Iniciación a la Investigación” fellowship, Universidad Autónoma de Madrid for a Master’s fellowship, and the Basque Government for a PhD fellowship. B.S. was supported by a Ramón y Cajal contract RYC-2011-08746.”

Notes and references

- 1 E. Gazit, *Nat. Nanotechnol.*, 2008, **3**, 8–9.
- 2 R. V. Uljijn and D. N. Woolfson, *Chem. Soc. Rev.*, 2010, **39**, 3349–3350.
- 3 K. Rajagopal and J. P. Schneider, *Curr. Opin. Struct. Biol.*, 2004, **14**, 480–486.
- 4 K. B. Vargoa, R. Parthasarathya and D. A. Hammer, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 11657–11662.
- 5 T. P. Knowles, T. W. Oppenheim, A. K. Buell, D. Y. Chirgadze and M. E. Welland, *Nat. Nanotechnol.*, 2010, **5**, 204–207.
- 6 S. Zhang, M. A. Greenfield, A. Mata, L. C. Palmer, R. Bitton, J. R. Mantei, C. Aparicio, M. O. de la Cruz and S. I. Stupp, *Nat. Mater.*, 2010, **9**, 594–601.
- 7 S. Gajria, T. Neumann and M. Tirrell, *Wiley Interdiscip. Rev.: Nanomed. Nanobiotechnol.*, 2011, **3**, 479–500.
- 8 N. P. King, W. Sheffler, M. R. Sawaya, B. S. Sollmar, J. P. Sumida, I. André, T. Gonen, T. O. Yeates and D. Baker, *Science*, 2012, **336**, 1171–1174.
- 9 J. C. Sinclair, K. M. Davies, C. Venien-Bryan and M. E. Noble, *Nat. Nanotechnol.*, 2011, **6**, 558–562.
- 10 P. K. Lo, K. L. Metera and H. F. Sleiman, *Curr. Opin. Chem. Biol.*, 2010, **14**, 597–607.
- 11 I. Medalsy, O. Dgany, M. Sowwan, H. Cohen, A. Yukashevskaya, S. G. Wolf, A. Wolf, A. Koster, O. Almog, I. Marton, Y. Pouny, A. Altman, O. Shoseyov and D. Porath, *Nano Lett.*, 2008, **8**, 473–477.
- 12 R. Matsunaga, S. Yanaka, S. Nagatoishi and K. Tsumoto, *Nat. Commun.*, 2013, **4**, 2211.
- 13 H. Gradisar, S. Bozic, T. Doles, D. Vengust, I. Hafner-Bratkovic, A. Mertelj, B. Webb, A. Sali, S. Klavzar and R. Jerala, *Nat. Chem. Biol.*, 2013, **9**, 362–366.
- 14 H. Gradisar and R. Jerala, *J. Nanobiotechnol.*, 2014, **12**, 4.
- 15 E. Main, A. Lowe, S. Mochrie, S. Jackson and L. Regan, *Curr. Opin. Struct. Biol.*, 2005, **15**, 464–471.
- 16 E. R. G. Main, Y. Xiong, M. J. Cocco, L. D’Andrea and L. Regan, *Structure*, 2003, **11**, 497–508.
- 17 L. D’Andrea and L. Regan, *Trends Biochem. Sci.*, 2003, **28**, 655–662.
- 18 A. L. Cortajarena, S. G. Mochrie and L. Regan, *Protein Sci.*, 2011, **20**, 1042–1047.
- 19 A. L. Cortajarena, T. Y. Liu, M. Hochstrasser and L. Regan, *ACS Chem. Biol.*, 2010, **5**, 545–552.
- 20 A. L. Cortajarena, F. Yi and L. Regan, *ACS Chem. Biol.*, 2008, **3**, 161–166.
- 21 M. E. Jackrel, A. L. Cortajarena, T. Y. Liu and L. Regan, *ACS Chem. Biol.*, 2010, **5**, 553–562.
- 22 A. L. Cortajarena and L. Regan, *Protein Sci.*, 2011, **20**, 341–352.
- 23 T. Z. Grove, J. Forster, G. Pimienta, E. Dufresne and L. Regan, *Biopolymers*, 2012, **97**, 508–517.
- 24 T. Z. Grove, C. O. Osuji, J. D. Forster, E. R. Dufresne and L. Regan, *J. Am. Chem. Soc.*, 2010, **132**, 14024–14026.
- 25 T. Z. Grove, L. Regan and A. L. Cortajarena, *J. R. Soc., Interface*, 2013, **10**, 20130051.
- 26 T. Kajander, A. L. Cortajarena, S. G. Mochrie and L. Regan, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2007, **63**, 800–811.
- 27 A. L. Cortajarena, J. Wang and L. Regan, *FEBS J.*, 2010, **277**, 1058–1066.
- 28 J. J. Phillips, C. Millership and E. R. Main, *Angew. Chem.*, 2012, **51**, 13132–13135.
- 29 A. L. Cortajarena, T. Kajander, W. Pan, M. J. Cocco and L. Regan, *Protein Eng., Des. Sel.*, 2004, **17**, 399–409.
- 30 A. L. Cortajarena and L. Regan, *Protein Sci.*, 2006, **15**, 1193–1198.
- 31 E. R. Ballister, A. H. Lai, R. N. Zuckermann, Y. Cheng and J. D. Mougous, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 3733–3738.
- 32 F. F. Miranda, K. Iwasaki, S. Akashi, K. Sumitomo, M. Kobayashi, I. Yamashita, J. R. H. Tame and J. G. Heddl, *Small*, 2009, **5**, 2077–2084.
- 33 B. B. Darwent, *Bond Dissociation Energies in Simple Molecules*, Washington, 1970.
- 34 B. S. Mamathambika and J. C. Bardwell, *Annu. Rev. Cell Dev. Biol.*, 2008, **24**, 211–235.
- 35 B. R. Willians, R. A. Gelman, D. C. Poppe and K. A. Piez, *J. Biol. Chem.*, 1978, **253**, 6578–6585.
- 36 A. Radzicka and R. Wolfenden, *Biochemistry*, 1988, **27**, 1664–1670.
- 37 T. Kajander, A. L. Cortajarena, E. R. Main, S. G. Mochrie and L. Regan, *J. Am. Chem. Soc.*, 2005, **127**, 10188–10190.
- 38 A. L. Cortajarena, S. G. Mochrie and L. Regan, *J. Mol. Biol.*, 2008, **379**, 617–626.

- 39 A. L. Cortajarena, G. Lois, E. Sherman, C. S. O'Hern, L. Regan and G. Haran, *J. Mol. Biol.*, 2008, **382**, 203–212.
- 40 A. Ortega, D. Amoros and J. Garcia de la Torre, *Biophys. J.*, 2011, **101**, 892–898.
- 41 G. Odian, *Principles of polymerization*, Wiley-Interscience, 2004.
- 42 P. J. Flory, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, NY, 1953.
- 43 K. Liu, Z. Nie, N. Zhao, W. Li, M. Rubinstein and E. Kumacheva, *science*, 2010, **329**, 197–200.

Appendix 2.

Assembly of designed protein scaffolds into monolayers for nanoparticle patterning



Assembly of designed protein scaffolds into monolayers for nanoparticle patterning



Sara H. Mejias^{a,1}, Pierre Couleaud^{a,1}, Santiago Casado^a, Daniel Granados^a, Miguel Angel Garcia^{b,c}, Jose M. Abad^{a,d,*}, Aitziber L. Cortajarena^{a,e,*}

^a IMDEA-Nanociencia and Centro Nacional de Biotecnología (CNB-CSIC)—IMDEA Nanociencia Associated Unit, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain

^b Instituto de Cerámica y Vidrio (ICV-CSIC), Cantoblanco, 28049 Madrid, Spain

^c Instituto de Magnetismo Aplicado "Salvador Velayos", UCM-ADIF, 28230 Madrid, Spain

^d Departamento de Química Analítica y Análisis Instrumental, Facultad de Ciencias, Universidad Autónoma de Madrid, 28049 Madrid, Spain

^e CIC BiomaGUNE, Parque Tecnológico de San Sebastián, Paseo Miramón 182, Donostia-San Sebastián 20009, Spain

ARTICLE INFO

Article history:

Received 15 August 2015

Received in revised form 1 December 2015

Accepted 20 January 2016

Available online 22 January 2016

Keywords:

Repeat protein

Protein design

Self-assembly

Monolayer

Gold nanoparticles

Biomolecular templating

Building blocks

Nanoparticle 2D arrays

Nanopatterning

ABSTRACT

The controlled assembly of building blocks to achieve new nanostructured materials with defined properties at different length scales through rational design is the basis and future of bottom-up nanofabrication. This work describes the assembly of the idealized protein building block, the consensus tetratricopeptide repeat (CTPR), into monolayers by oriented immobilization of the blocks. The selectivity of thiol–gold interaction for an oriented immobilization has been verified by comparing a non-thiolated protein building block. The physical properties of the CTPR protein thin biomolecular films including topography, thickness, and viscoelasticity, are characterized. Finally, the ability of these scaffolds to act as templates for inorganic nanostructures has been demonstrated by the formation of well-packed gold nanoparticles (GNPs) monolayer patterned by the CTPR monolayer.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

The use of building blocks and self-assembly is acknowledged as a versatile strategy for the fabrication of multifunctional nanostructures. Inorganic building blocks and nanoparticles have already shown great potential in the fabrication of new materials and multifunctional structures [1–4]. Additionally, the synthesis of hybrid organic–inorganic multifunctional nanostructures has also been developed [5,6]. At the contrary, the use of purely organic building blocks, such as biomolecules, has not yet been fully exploited in nanotechnology. Self-assembly of biological molecules into defined functional structures has tremendous potential in nanopatterning and in the design of novel materials and functional devices

[7–15]. Moreover, templating inorganic nanostructures, such as gold nanoparticles, using biomolecules is a challenging way to generate ordered inorganic materials [16,17] for example to produce memory devices, non-linear optical and sensing devices, and nanowires [18,19]. These strategies permit to have a high control at the nanometer scale, that is the major limitation of the conventional lithographic top-down processes used actually.

Molecular self-assembly is a process by which complex three-dimensional structures with specified functions are constructed from simple molecular building blocks. These methodologies enable the precise control over the structure but also over the function that can be encoded in the biomolecules. In order to take advantage of the great potential of using biomolecules to generate functional templates, it is key to have a molecular understanding of the building blocks and a control over interactions that drive their assembly. Additionally, green, easy and innovative strategies to have a precise control at the atomic scale for the development of scaffold and patterned structures are still to be found.

* Corresponding author.

E-mail addresses: josemaria.abad@uam.es (J.M. Abad), aitziber.lopezcortajarena@imdea.org, alcortajarena@cicbiomagune.es (A.L. Cortajarena).

¹ These authors contributed equally to this work.

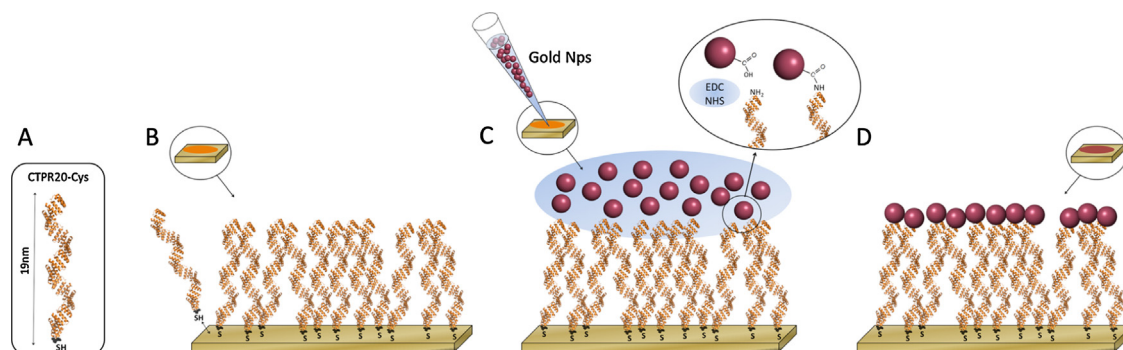


Fig. 1. (A) Representation of the crystal structure of CTPR20-Cys protein. (B) Schematic representation of CTPR20-Cys adsorption onto gold surface through thiol-gold interactions. (C) Overview of the functionalization process of gold-adsorbed CTPR20-Cys with gold nanoparticles (GNPs) through amide bond formation in the presence of EDC and NHS. (D) Schematic representation of the GNPs monolayer patterned by the previously adsorbed monolayer of CTPR20-Cys onto gold surface.

In that sense the use of proteins is of great interest as a new strategy for bottom-up approaches in nanotechnology [20]. For example, compared to other approaches used for patterning nanoparticles, such as nanolithography, any bottom up approach is more cost effective and easier to implement. In this sense, interfacial bottom up 2D assembly of nanoparticles has become a novel and widely used strategy, however it presents some limitations, including the difficulty of assembling nanoparticles with different coatings and mostly the low adhesion of the film when transferred to the substrate [21]. Nanoparticle templating by proteins combines the cost efficiency of bottom-up approaches with the capability of covalent bioconjugation with sub-nanometric resolution. The complexity and sophistication of protein-based structures in nature hints to the great potential of designed protein-based templating. In particular, the use of designed repeat proteins as the building blocks for fabrication and patterning provides several advantages [22]. Repeat proteins present a modular structure defined by local repeated interactions and are composed of tandem arrays of the same small structural motif [23,24]. Their simple architecture makes easier to understand the basic rules that relate sequence to structure for these repeated modules and make them ideal molecular building blocks [25,26].

Consensus tetratricopeptide repeat (CTPR) [23] is an idealized 34 amino acid helix-turn-helix repeat module that can be combined in tandem to form CTPR proteins with different numbers of repeats from 2 to 20 so called CTPR2 to CTPR20 proteins [27]. CTPR proteins display superhelical structures, in which eight repeats comprise one full turn of the superhelix (Fig. 1A) [27,28]. Additionally, the stability and ligand binding properties of CTPRs can be modulated in a predictable manner, which opens the opportunity to tune the building block properties for specific applications [29–32]. We have recently used these modules to show the specific functionalization of patterned polymeric surfaces [33,34]. Moreover, it has been shown that under certain conditions CTPR proteins self-assemble into ordered structures including ordered protein films comprised [35] and linear nanofibers in solution [36,37], mimicking the packing observed in the crystal forms of CTPRs [27,28,38]. These results illustrate the potential of these protein modules as self-assembling building blocks.

One key component for the assembly is the side-to-side lateral interactions between CTPR superhelices [35,39]. Fig. 1A schematically shows the potential alignment of CTPR superhelices. In this work, we focus on the self-assembly of the CTPR20-Cys protein composed of 20 identical CTPR repeats with a single C-terminal cysteine. CTPR20 is chosen since it is the longest CTPR that can be produced recombinantly with high yield. The longer the CTPR protein the higher the aspect ratio [26], thus CTPR20 is expected to have higher propensity to be oriented on the surface and form more sta-

ble side-to-side interactions. In addition, longer CTPRs show higher stability [40] which is important for future applications. CTPR20-Cys is immobilized on a gold surface through gold-sulfur bond and, driven by the lateral inter-molecular interactions, can generate ordered and oriented protein self-assembled monolayers (SAMs) (Fig. 1B). To characterize the protein's specific adsorption on gold surface and its self-packing, we monitored the immobilization by different techniques including quartz crystal microbalance (QCM), surface plasmon resonance (SPR) and atomic force microscopy (AFM).

Finally, to demonstrate the potential of the repeat protein scaffolds to act as templates for nanoscale patterning, gold nanoparticles (GNPs) were grafted onto immobilized CTPR20-Cys through the formation of a covalent amide bond between the terminal amine of the protein and the carboxylate groups of the GNPs (Fig. 1C). The resulting functionalized surface with a monolayer of GNPs was characterized by scanning electron microscopy (SEM) providing a final evidence of the oriented monolayer formation of CTPR20-Cys protein and a potential application in the controlled patterning of GNPs (Fig. 1D). The use of GNPs in an ordered monolayer organization is of great interest in field of plasmonics [41] and organic solar cells progress [42] or for the development of localized surface plasmon resonance-based sensor [43].

2. Results and discussion

The main goal of this work is to demonstrate that repeat proteins can be used as scaffolds for the patterning of inorganic nanostructures, such as gold nanoparticles.

The first step of the study is the adsorption of CTPR20-Cys onto gold surface. Thus, we first compared the immobilization on gold surface of two proteins: CTPR20-Cys for oriented thiol mediated immobilization and CTPR20 without Cys as a non-oriented deposition control. The process was monitored by quartz crystal microbalance (QCM), which is a powerful tool to study interfacial process at surfaces. After stabilization of the temperature, frequency and resistance, purified proteins solutions prepared as previously reported [13,44] were flowed through the cell. Fig. 2A and B presents the resulting frequency and resistance changes as a function of time after protein injection. As it can be seen in Fig. 2A, upon flowing CTPR20-Cys protein, a rapid decrease in the resonant frequency was observed during the first 2 min as a result of the protein binding to the surface by chemisorption of cysteine on gold. Afterward, the frequency decreased more slowly until a steady state was reached. Simultaneously with the frequency decrease, an increase in motional resistance was also observed (Fig. 2B). It indicates that the binding of the protein does not result in a rigid load

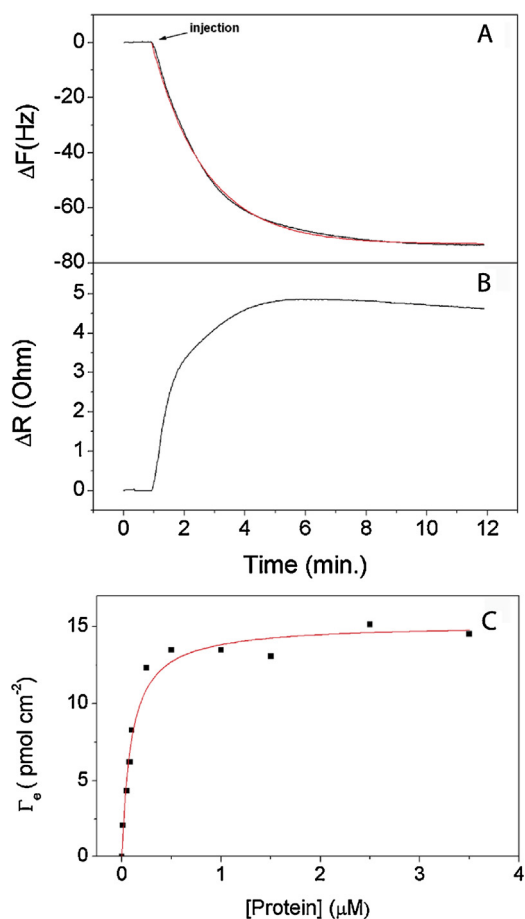


Fig. 2. CTPR20-Cys adsorption on gold surface monitored by quartz crystal microbalance. (A) Adsorption kinetics of CTPR20-Cys onto quartz crystals with gold electrodes at 2.5 μM protein concentration. The frequency changes of the quartz crystal resonator are recorded upon protein injection. The curve was fit to a first-order kinetics equation (solid line), obtaining -73.2 Hz and 0.57 min^{-1} values for $-\Delta F_{\text{max}}$ and k , respectively. (B) Simultaneous resistance increase recorded by QCM upon CTPR20-Cys protein injection. (C) Equilibrium surface coverage by QCM. Equilibrium surface coverage of CTPR20-Cys adsorbed on gold surface represented versus the protein concentration and fitted using a Langmuir adsorption isotherm. The calculated value for Γ_s was $15.2 \times 10^{-12} \text{ mol cm}^{-2}$ for CTPR20-Cys.

but viscoelastic causing an energy loss by dissipation from the QCM sensor.

The ratio between the motional resistance change and the frequency shift can be used to evaluate the relative influence of viscoelastic and mass loading effects (see below). The shape of the frequency–time profile can be employed to study the kinetics of adsorption by fitting to a first-order kinetics equation (Fig. 2A and Section 4). The values of ΔF_{max} and k obtained from the data fit were -73.2 Hz and 0.57 min^{-1} , respectively. From the ΔF_{max} at the equilibrium, assuming that the frequency decrease is only due to the change in mass arising from the adsorption of the protein, the mass of CTPR20-Cys immobilized on the gold electrode surface was estimated to be 1285 ng cm^{-2} using Eq. (1). This value corresponds to a surface coverage of about $15.2 \times 10^{-12} \text{ mol cm}^{-2}$ considering a molecular mass of 85242 Da for the CTPR20-Cys (9.0×10^{12} CTPR20-Cys molecules cm^{-2}). This value is in agreement with the number of molecules for a compact monolayer of oriented CTPR20-Cys proteins, considering a projected area of about 11 nm^2 per molecule based on the dimensions from the crystal structure [27]. Additional support to this fact is provided by the comparison with the extensively studied monolayers of *n*-alkanethiolates onto gold surface [45,46] which result in a maximum coverage of $\sim 4.2 \times 10^{14}$ molecules cm^{-2} for molecules that cover 0.217 nm^2

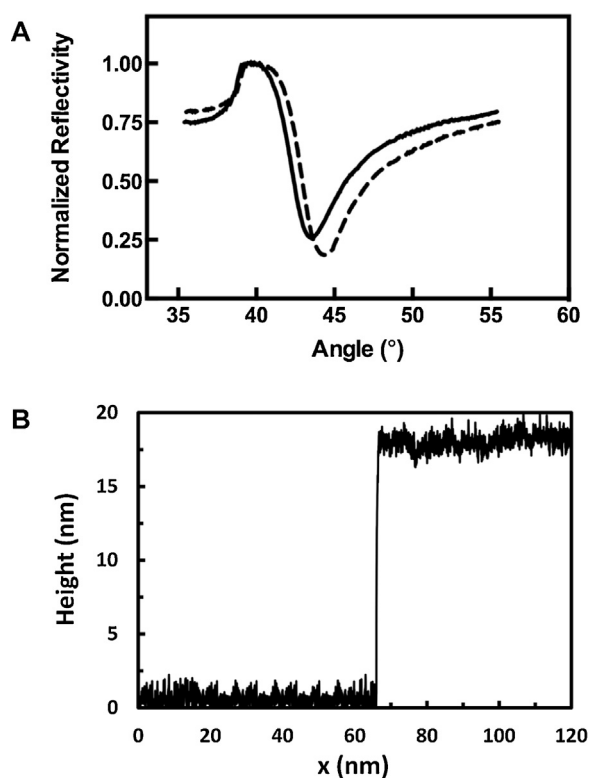


Fig. 3. (A) Surface protein coverage by surface plasmon resonance. SPR spectra of bared gold surface (solid line) and CTPR20-Cys deposited on the gold surface (dashed line). The protein spectrum fits with a gold film thickness of $32 \pm 2 \text{ nm}$, a protein coverage of approximately 1000 ng cm^{-2} . (B) Height profile of the CTPR20-Cys monolayer deposited on the gold surface acquired by a stylus profilometer. The thickness of the protein layer was $17.93 \pm 1.05 \text{ nm}$.

[47]. For an equally packed CTPR20-Cys molecule (11 nm^2) the theoretical maximum coverage would be reached with 50 times less molecules 8.4×10^{12} molecules cm^{-2} , which is in the same range of the experimental value of 9.0×10^{12} molecules cm^{-2} .

These results show that the protein units assemble in a compact manner on the surface by thiol chemisorption of cysteine residues onto the gold surface. As it is the case for more simple thiol derivatives, CTPR20-Cys proteins are not randomly deposited since the amount of protein per surface area would be otherwise smaller. To serve as comparison, QCM adsorption measurements of CTPR20 without cysteine were also carried out (Electronic supplementary material). As shown in Fig. S1 flowing CTPR20 gives a final surface coverage of $6.5 \times 10^{-12} \text{ mol cm}^{-2}$, significantly lower than for CTPR20-Cys. This result can be associated to the fact that direct adsorption of CTPR20 proceeds through unspecific multipoint attachment of the protein to the surface contrary to the oriented thiol-derivative protein that is tightly adsorbed on the gold surface by a unique single-point interaction [48,49]. This fact was also put in evidence by studying the thermodynamics of adsorption. The calculated equilibrium proteins coverages from ΔF_{max} for different protein concentrations are fitted to a Langmuir adsorption isotherm (Materials and methods section, Eq. (4)) to calculate the saturation protein coverage ($\Gamma_s = 15.2 \times 10^{-12} \text{ mol cm}^{-2}$) (Fig. 2C). In contrast, for CTPR20 without cysteine the Γ_s was $7.3 \times 10^{-12} \text{ mol cm}^{-2}$ (Fig. S2). The coverage at saturation for CTPR20 is half of the one obtained for CTPR20-Cys, presumably due to a random immobilization of CTPR20 in different orientations.

In order to confirm the validity of conversion the change in frequency to molecular mass from QCM data using Sauerbrey's equation, surface plasmon resonance (SPR) experiments were conducted, as detailed in the Materials and methods section [50,51], to

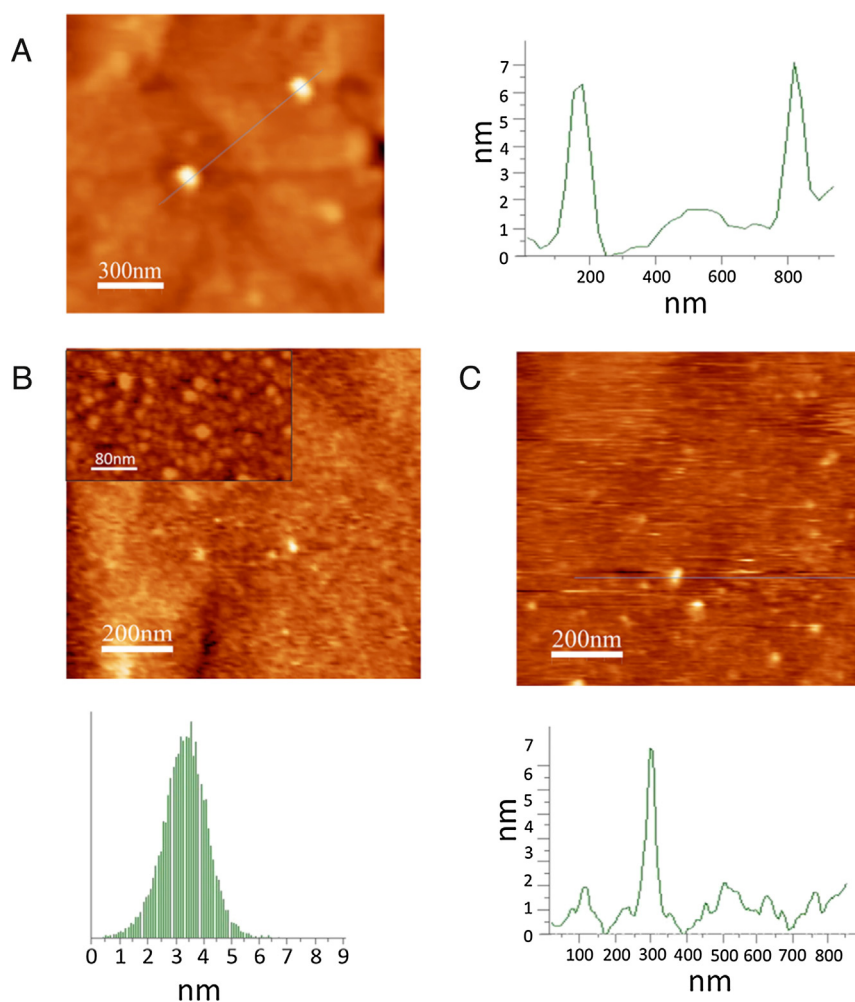


Fig. 4. AFM image of CTPR20-Cys protein immobilized on gold surfaces. (A) AFM image of individual CTPR20-Cys molecules deposited on gold surface at 10 nM protein concentration after 1 hour incubation at room temperature. Each bright spot corresponds to a single protein. The right panel shows the height profile of two single protein molecules. Note that the nanometer-scale widths obtained by AFM are not reliable due to the size and shape of the contacting probe. (B) AFM image of a large area of completely saturated CTPR20-Cys monolayer deposited on gold surface at 2.5 μ M protein concentration. The inset shows a zoom-in image in which individual protein units can be observed. The panel below shows the global roughness analysis of the image that showed an RMS roughness of 0.86 nm and an average size of 3.42 nm. (C) AFM image of the same region as image B after the addition of DTT to the sample in order to remove the CTPR20-Cys. The panel below shows the height profile along the line to show a single protein molecule.

quantify the amount of protein bound. Fig. 3A shows a SPR experiment carried out employing a CTPR20-Cys protein concentration of 2.5 μ M to saturate the surface. The mass measured by SPR from the resonance angle shift upon protein binding to the gold surface was approximately 1000 ng cm^{-2} . This mass is slightly lower than the value obtained from QCM (1285 ng cm^{-2}), probably due to the uncertainty of the refractive index of the adsorbed protein and to the effect of the hydration water in the QCM measurements. The SPR mass can be considered comparable to the QCM mass confirming that the frequency changes observed in the QCM experiments are mainly ascribed to mass changes as consequence of the protein adsorption since viscoelastic effects should induce a bigger difference between Δm_{QCM} and Δm_{SPR} , as previously reported [52].

We used AFM to image the CTPR20-Cys proteins immobilized onto gold surfaces. Single CTPR20-Cys molecules were imaged first at a low protein concentration (10 nM) to characterize the features of the individual monomers (Fig. 4A). A single size distribution peak was obtained for individual CTPR20-Cys molecules with an average height of approximately 8 nm, indicating that each bright dot on the image corresponds to a protein monomer (Fig. 4A). The width measurements are not reliable due to the tip convolution [53]. The end to end distance of the CTPR20 superhelix calculated from the

crystal structure is approximately 19 nm (Fig. 1A) [27]. There are several reasons that will account for the smaller size measured by AFM: (1) CTPR20-Cys adsorption onto gold surface through its thiol group might be tilted with an angle relative to an orthogonal orientation to the surface; (2) vertical elastic tip compression of the elongated superhelical structure, as it has been described for other repeat proteins [54]; (3) AFM characterizes single protein molecules in solution, which might have some flexibility [55] compared to the dimensions determined from the rigid crystal structure of the protein [56].

Following, CTPR20-Cys monolayer assembly on gold surface was studied. CTPR20-Cys at 2.5 μ M concentration, which guarantees full coverage of the surface according to the equilibrium immobilization studies by QCM (Fig. 2C), was deposited onto annealed gold surfaces. The AFM image shows a densely packed array of protein particles that completely covers the gold surface (Fig. 4B). Under the solution deposition conditions and in the range of protein concentration used (10 nM–2.5 μ M), CTPR20-Cys protein is stable, folded, and fully monomeric [27,40], thus is not expected the deposition of multimeric particles. The AFM images at saturation provide insights about the tight packing of the protein molecules on the monolayer and clearly show a full coverage of the surface (Fig. 4B), corrob-

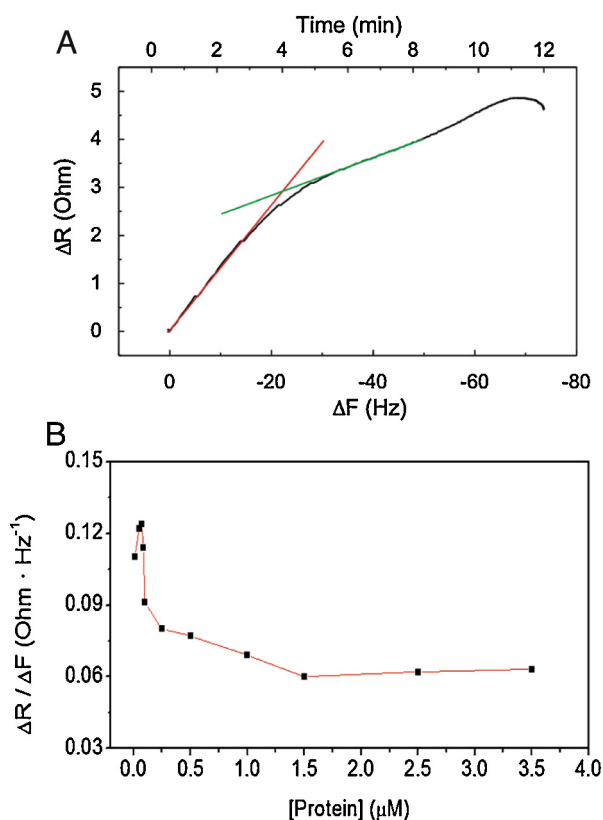


Fig. 5. Viscoelastic properties of deposited protein material measured by QCM. (A) The change in resistance during the course of protein immobilization and assembly kinetics vs. the change in frequency for CTPR20-Cys at 2.5 μM protein concentration. For CTPR20-Cys two regions are observed. The fit to linear equations with two different slopes are shown. (B) The change in resistance divided by the change in frequency ($\Delta R/\Delta F$) for different protein concentrations of CTPR20-Cys measured after steady state was reached.

orating a higher occupancy than expected from simple random distribution of individual CTPR20 monomers. The overall roughness analysis of the surface resulted in a RMS roughness of 0.86 nm with an average height of 3.42 nm, smaller than the one observed for individual CTPR20-Cys proteins (RMS roughness of 1.88 nm and average size of 6.36 nm), in agreement with a compaction of the molecules. The CTPR20-Cys covered surface shown in Fig. 4B was incubated with an excess of dithiothreitol (DTT) that will displace the previously adsorbed CTPR20-Cys, in order to study the specificity of the interaction between CTPR20-Cys and the gold surface. After the incubation and wash of the surface the AFM image (Fig. 4C) shows some dots corresponding to individual protein molecules, while most of the protein has been detached from the surface. This result confirms the specificity of the thiol–gold immobilization of the CTPR20-Cys.

In order to determine the thickness of the monolayer imaged by AFM, profilometry experiments were performed on a CTPR20-Cys layer deposited on one half of a gold surface using 2.5 μM protein concentration. Three protein layers were measured to give a thickness value of 17.93 ± 1.05 nm (Fig. 3B), consistent with the formation of a single CTPR20-Cys protein monolayer.

In addition, the ratio between the motional resistance (ΔR) and the frequency (ΔF) provides information about the viscoelastic properties of the deposited protein layer [57]. In Fig. 5A, the ΔR during the course of CTPR20-Cys protein assembly is plotted against ΔF . Two separate regions are observed with different $\Delta R/\Delta F$ slopes indicating that surface coverage has a strong influence on the viscoelastic properties of the protein film. The first region (ΔF from 0 Hz to -20 Hz) shows a linear relationship between $\Delta R/\Delta F$

presenting a steeper slope. It corresponds to the beginning of the adsorption process associated to lower protein densities and therefore less lateral interactions between protein molecules. The deposit tends to be more flexible and more dissipative at this stage. When higher protein densities are reached (ΔF from -20 Hz to -60 Hz), protein molecules can establish lateral interactions with neighbouring molecules, yielding finally to a more rigid and less dissipative film. The change in slope between two regions indicates the transition from individual proteins to a compact state. The intersection of the two straight segments from the linear fit occurs at 22.2 Hz and at a $\Delta R/\Delta F$ ratio $0.130 \Omega \text{ Hz}^{-1}$.

To confirm that the observed difference in $\Delta R/\Delta F$ between high and low protein density on the surface could be attributed to the protein compact state, experiments using different protein concentrations and recording ΔF and ΔR simultaneously were performed (Fig. 5B). It is observed that at low protein concentrations the $\Delta R/\Delta F$ ratio increases as the protein concentration increases, associated to a higher viscoelasticity of the protein film, reaching a maximum of $0.124 \Omega \text{ Hz}^{-1}$ at 0.075 μM of protein. It is noteworthy that this $\Delta R/\Delta F$ value is comparable to that obtained from the intersection in Fig. 5A. This result shows that at low concentrations there are not enough neighbouring CTPR20-Cys molecules immobilized for them to assemble, as previously suggested. When the protein concentration is increased, a subsequent $\Delta R/\Delta F$ decrease is obtained until reaching a stable value of $0.06 \Omega \text{ Hz}^{-1}$ at surface-saturating protein concentrations. The surface becomes more densely packed with CTPR20-Cys molecules, promoting lateral interactions and leading to an increase in the rigidity of the film. These results clearly suggest that the mode of attachment through the C-terminal Cys residue leads to a spatial arrangement, with parallel alignment between protein molecules, favoring the establishment of lateral protein interactions. In the case of CTPR20 protein without cysteine this analysis shows a different behavior with a linear increase in the viscoelasticity of the deposit as the protein is adsorbed (Electronic Supplementary material Fig. S3).

The final step of this study is the functionalization of the protein monolayer with gold nanoparticles (GNPs) in order to monitor the templating capabilities of the CTPR monolayer. In addition, the attachment of GNPs will allow us to achieve a higher resolution characterization of the surface using SEM. GNPs stabilized with thioctic acid (TA-GNPs) displaying free carboxylic acid functions, required for the immobilization step, were used. As the CTPR20-Cys was selectively immobilized onto the gold surface through its single C-terminal cysteine, the N-terminal moiety of the protein was selected as the anchoring point. The anchoring reaction is achieved through the formation of an amide bond between amine of the N-terminal of the CTPR20-Cys (previously immobilized on gold surface) and the carboxylic acid functions of TA-GNPs, which were activated *in situ* by EDC/NHS. After the reaction of the TA-GNPs with the gold surface coated with CTPR20-Cys, scanning electron microscope (SEM) was used to characterize the functional surfaces obtained. The AFM results showed the evidences of a monolayer of compacted CTPR-20-Cys (Fig. 4) onto gold surface and a homogeneous coverage of protein over surfaces in the order of μm^2 . Each CTPR20-Cys molecule should be able to interact only with one GNP through an amide bond formation, and the area of the protein section is about 11 nm^2 , in the same range of the GNPs diameter. Therefore it is expected to observe a homogenous coverage of a single layer of GNPs over the CTPR20 surface.

Fig. 6A shows a significant GNPs surface coverage yield (clearer areas) in a highly packed distribution. The entire surface was not fully covered by GNPs but CTPR20-Cys can be observed on almost the whole surface (as black/grey background). An important observation is that when the functionalization with GNPs occurred, it formed preferentially a monolayer of GNPs with a high level of

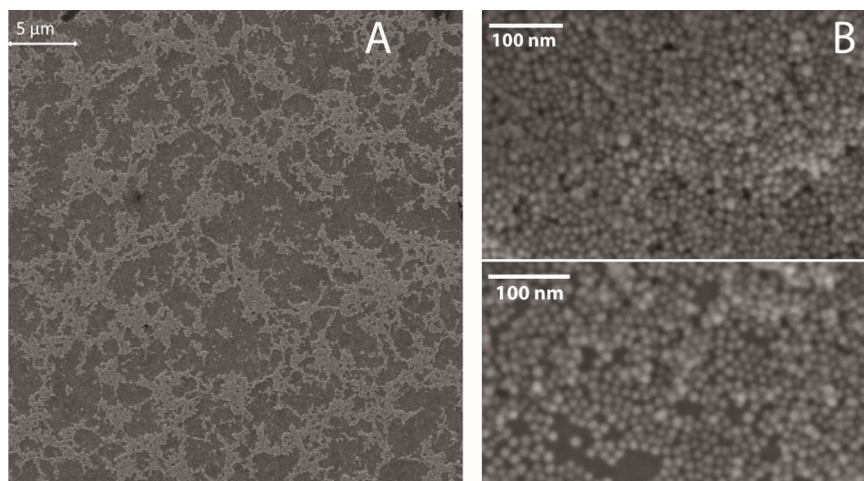


Fig. 6. (A) SEM image of a large area ($30\ \mu\text{m} \times 30\ \mu\text{m}$) of gold surface covered by CTPR20 and functionalized with GNPs. (B) Representative SEM images of GNPs immobilized onto CTPR20 covering gold surface. Two areas in which is clear the close packed GNPs monolayer are displayed. SEM images are in the same scale as the AFM images of the protein monolayer displayed in Fig. 4.

compaction, leading to islands of close-packed GNPs monolayers (Fig. 6B). At a higher magnification, the monolayer organization of the GNPs functionalized onto CTPR20 immobilized onto gold surface can be clearly confirmed (Fig. 6B). The controls of GNPs onto gold surface and GNPs onto CTPR20 without coupling agents gave very different results in SEM, with a majority of isolated GNPs aggregates and/or randomly dispersed (Electronic supplementary material). The comparison of both SEM (Fig. 6B) and AFM (Fig. 4B) at the same scale confirms the monolayer organization of the GNPs and that the same kind of organization is observed for the protein onto gold surface, presenting in both cases a high compaction. This was not observed when the GNPs were deposited on gold surfaces without prior CTPR20-Cys protein layer formation (Figs. S5 and S6), which corroborates that the CTPR20-Cys monolayer is templating the GNPs.

3. Conclusions

In conclusion, we have demonstrated that we can immobilize oriented long designed CTPR proteins on gold surface through thiol–gold specific interaction and that the protein modules present intrinsic lateral assembly properties. By controlled immobilization we generated a tightly packed protein SAM and characterized the packing by QCM, SPR, and AFM. We monitored the assembly transition by the change on the viscoelastic properties of the immobilized materials. It is clear that the oriented immobilization of the long protein arrays allows the arrangement of the modules and the formation of a rigid protein film on the surface. Regarding the mechanism of interaction between the proteins and the forces that drive the assembly, we can speculate based on the previously observed lateral packing of superhelices in solid films [35] and in crystal forms [27]. Long CTPR proteins are structurally rigid with anisotropic shape (Fig. 1), that allows for preferential alignment along one direction and lateral packing when the proteins are immobilized directionally oriented on a surface. The crystal packing provides hints that CTPR protein superhelices have a tendency to interact through salt bridges between positively and negatively charged residues on the protein surface [58]. Similar interactions can mediate side-to-side packing within monolayer. The functionalization with GNPs additionally confirms that most of the protein molecules are oriented as expected (N-terminal free) and in a monolayer disposition. This work is an example where the molecular understanding of the biomolecular blocks and the control of the immobilization and assembly process on surfaces can lead to

the generation of materials with defined macroscopic properties. For example, by changing the protein packing density, the rigidity of the protein film can be tuned at will. Additionally, these ordered and structured biocompatible surfaces can be functionalized by introducing unique groups with atomic resolution at defined positions in the protein. The results of this work show the potential of these designed scaffolds as platforms for patterning GNPs. The assembly of GNPs into closely packed thin films has great interest because of their potential applications in electronics, nanophotonics, data storage, and plasmonics [59]. The templating strategy described here is versatile and can be easily applied to a variety of nanoparticles and other molecules for the generation of closely packed arrays. The application of these biomolecular templates can result in novel types of hybrid materials within the bottom-up approaches.

4. Materials and methods

4.1. Protein design and purification

CTPR20 protein with a single cysteine at the C-terminal was expressed as His-tagged fusion and purified using standard affinity chromatography methods as previously described [13,27,44]. The protein was dialyzed into PBS buffer (150 mM NaCl, 50 mM phosphate buffer pH 7.4) and stored frozen at $-20\ ^\circ\text{C}$. The protein concentration was determined by UV absorption using a Cary 50 UV–vis spectrophotometer (Varian) at 280 nm using the extinction coefficient calculated from the amino acid composition [60].

4.2. Protein reduction for gold surface immobilization experiments

Prior to any immobilization experiment on gold, purified CTPR20-Cys at a protein concentration of about $23\ \mu\text{M}$ was freshly reduced with 1 mM dithiothreitol (DTT) during 20 min to ensure full reduction of the sample. DTT was removed by buffer exchange over a NAP-5 column (GE Healthcare Life Science) in PBS-2 buffer with lower salt and buffer concentration (10 mM NaCl, 10 mM phosphate buffer pH 7.4) or water. Protein fractions without DTT are collected at $4\ ^\circ\text{C}$ and their concentration measured by UV absorption spectroscopy.

4.3. Quartz crystal microbalance (QCM)

QCM is an extremely sensitive mass sensor capable of providing real-time monitoring of immobilization events in the nanogram range and determination of kinetic rate constant. QCM measures frequency changes in the resonant frequency of the quartz crystal according to Sauerbrey's equation (Eq. (1)), where a decrease in mass corresponds to an increase in frequency and vice versa. Moreover, QCM with motional resistance monitoring also allows obtaining information related to the structure, conformation, and interactions of the molecules on the surface through the acoustic dissipation. In situ mass changes were measured with a SRS QCM200 Quartz Crystal Microbalance from SRS Instruments (Sunnyvale, CA., USA) with AT-cut quartz crystals (5 MHz) of 25 mm diameter with gold electrodes deposited over a Chromium adhesion layer. An asymmetric electrode format was used with the side having the larger gold area facing the solution. The electroactive working area (front side) was 1.370 cm² and the piezoelectric area (backside) was 0.317 cm².

Prior to use, the quartz crystals were cleaned by exposure to "piranha" solution (3:1 concentrated H₂SO₄/30% H₂O₂) followed by exhaustive rinsing with distilled water and a final rinse with ethanol/water (2:1). *Caution: piranha solution reacts violently with most organic materials and must be handled with extreme care.* The quartz crystal resonator was set in a probe made of Teflon and immersed in water-jacketed beaker thermostated at 25 °C. The crystals sensitivity was 17.68 10⁻⁹ g Hz⁻¹ cm⁻² [61–63].

Under constant flow protein samples in PBS-2 buffer at concentrations from 5 nM to 2.5 μM are injected in the system and both the changes in motional frequency and in resistance were recorded until saturation is reached.

Sauerbrey's equation describes the correlation between frequency changes in the resonant frequency of the quartz crystal and the immobilize mass for QCM experiments:

$$\Delta m = -C_f F \quad (1)$$

where Δm is the mass change (ng cm⁻²), C_f (17.7 ng Hz⁻¹ cm⁻²) the proportionality constant for the 5.0 MHz crystals used in this study, and ΔF the frequency change (in hertz). Although this equation rigorously applies to the system in air and mass additions forming an evenly distributed rigid layer on the active sensor area [64]. It is widely accepted to estimate the adsorbed mass of protein in liquid environments [65–67]. Sauerbrey equation can therefore be reliably used to measure the mass of thin, evenly distributed, rigid and non-porous overlayers even in the liquid phase [64]. In this way, it has been reported that the linearity between adsorbed mass and frequency response may persist under certain conditions [68].

Assuming that the immobilization process is kinetically controlled the frequency curves were fit to a first-order kinetics equation:

$$\Delta F = -\Delta F_{\max} (1 - e^{-kt}) \quad (2)$$

where ΔF is the frequency change (in hertz), ΔF_{\max} the frequency change between the initial and the steady-state frequencies, and k the first-order rate constant (min⁻¹).

The frequency-time data for CTPR20 can be well fitted to a first order double- exponential decay with two adsorption processes, one fast decay rate and a subsequent slower adsorption:

$$\Delta F = \Delta F_{\text{fast}} \exp(-k_{\text{fast}} t) + \Delta F_{\text{slow}} \exp(-k_{\text{slow}} t) + \Delta F_{\max} \quad (3)$$

From the ΔF_{\max} at the equilibrium and using Sauerbrey's equation (Eq. (1)) the total mass of protein can be estimated assuming that the frequency decrease is only due to the change in mass arising from the adsorption of the protein immobilized on the gold electrode surface.

From the frequency change at the equilibrium (ΔF_{\max}), the total amount of protein immobilized per surface area can be calculated at the different protein injections. The equilibrium protein coverage values were represented versus protein concentrations and fitted to a Langmuir adsorption isotherm:

$$\Gamma_e = \frac{\Gamma_s K_a C_s}{1 + K_a C_s} \quad (4)$$

where K_a is the thermodynamic binding constant, C_s is the bulk protein concentration, and Γ_e and Γ_s are the equilibrium and saturation protein coverage.

4.4. Surface plasmon resonance (SPR) measurements

A gold film of approximately 30 nm was first deposited on sodalime glass substrate (1 mm thick) by electron beam evaporation in a home-based deposition chamber with a base pressure of $P_b = 1 \times 10^{-8}$ Torr. The SPR spectrum of the gold surface was measured to calculate the exact thickness of the film. Then, a CTPR20-Cys protein solution at 2.5 μM concentration was deposited on the gold surface. After incubation for 30 min at room temperature the SPR spectrum of the protein film in solution was acquired.

SPR spectra were acquired using the Kretschmann–Raether configuration, [69,70] with a home-made device described elsewhere [50]. SPR was excited with a 632.8 nm laser. The angular response of the photodiode was corrected as described before [51]. For each sample 4 scans were recorded from different spatial locations. The spectra shown here correspond to the average (with the line thickness being the standard deviation). Any possible drift in the motor position was corrected by fixing the position of the critical angle to 39.5°; this value depends only on the glass and the surrounded air, and it is independent of the gold and protein films. To calculate the total mass of protein immobilized on the gold surface, a shift of 0.1° on the resonance angle is related to a mass of 1 ng mm⁻², as described before for proteins [71]. SPR curve simulation was carried out using Winspall freeware by RES-TEC (<http://www.res-tec.de/applications.html>) including the correction of refraction for triangular prisms.

4.5. Atomic force microscopy (AFM)

For AFM measurements CTPR20-Cys protein was immobilized on gold surfaces. The gold substrates were commercial gold-coated borosilicate glass substrates (Gold Arrandee™) annealed using a Bunsen burner prior to use. First, we monitored gold surfaces under different annealing conditions without protein to achieve sub-nanometric roughness and ensure that the surface roughness is small enough to clearly image the nanometric-scale protein molecules. CTPR20-Cys freshly reduced protein as described above was incubated on the gold surface during 1 h. After incubation the surface is thoroughly washed with water.

The protein coated surface is imaged in solution using a Nanotec Cervantes system in jumping mode [72]. Olympus standard silicon nitride probes of 0.05 N m⁻¹ and 18 kHz (OMCL-RC800PSA) were employed. Images were processed using the WSxM software (Nanotech Electronica, Madrid, Spain <http://www.nanotec.es/>) [73].

4.6. Profilometry measurements

CTPR20-Cys protein was immobilized on gold surfaces using the same protocol used for the generation of the AFM samples. Three different layers were obtained by incubation of 2.5 μM protein. The layer thickness was measured using a "Dektak XT" mechanical profilometer with a 2.5 μm radio stylus and 1 mg force (the minimal force).

4.7. Gold nanoparticle (GNP) synthesis

Gold nanoparticles were synthesized following the well-established Turkevich protocol [74] with slight adjustments as described previously [75]. Briefly, 100 mL of an aqueous solution of HAuCl₄ (Sigma–Aldrich) at 1 mM was refluxed at 140 °C and reduced for 15 min with 10 mL of an aqueous solution of sodium citrate (55 mM). The reaction led rapidly to a red-wine solution of citrate based GNPs. The sample is then cooled to room temperature and filter through 0.2 μm pore-size filter to eliminate large aggregates. The final concentration of GNPs (9 nM) was determined by UV–vis spectrometry using an extinction coefficient at 520 nm, $\epsilon = 2.7 \times 10^8 \text{ M}^{-1} \text{ cm}^{-1}$. The as-modified Turkevich protocol for the synthesis of GNPs leads to a well-dispersed colloidal suspension of citrate-stabilized GNPs with an average diameter of $12.4 \pm 1.2 \text{ nm}$ (measured by TEM, Fig. S6). A second step of ligand exchange with thioctic acid was carried out to introduce a more stable surfactant with carboxylic groups. To 10 mL of GNPs at 9 nM was added 1.15 mL of thioctic acid at 10 mM in ethanol to react overnight at room temperature. Then, sample was centrifuged 40 min at $10,000 \times g$ and supernatant discarded. The red-wine pellet was redispersed in 7 mL milliQ water, resulting in a light purple suspension. Then, drops of NaOH 10 mM were added until the solution turned to red-wine color, indicating a good colloidal dispersion of thioctic acid stabilized GNPs (TA-GNPs) due to negative charges of the carboxylic acids.

4.8. GNPs conjugation with CTPR20-Cys immobilized onto gold surface

The conjugation between TA-GNPs and CTPR20-Cys, previously immobilized onto gold surface by the C-terminal Cys residue, was carried out through the reaction between carboxylic acid functions of the GNPs ligands and the free exposed terminal amine of the protein. First, 9 μL of ethyl(dimethylaminopropyl) carbodiimide (EDC, 1 mM) and 6 μL of *N*-hydroxysuccinimide (NHS, 1 mM) were added to 1 mL of TA-GNPs at 30 nM for 10 min to perform the activation of the carboxylic groups. Then, 200 μL of activated TA-GNPs were incubated onto the gold surface covered by CTPR20-Cys during 3 days at room temperature. Controls of the same experiment without adding EDC/NHS coupling agents and the immobilization of TA-GNPs onto protein-free gold surface were also carried out under the same experimental conditions.

After the incubation time, the liquid phase was removed from the gold surface and the surface was washed several times with MilliQ water and dried under nitrogen flux.

4.9. Scanning electron microscopy (SEM)

SEM measurements of the surfaces with gold nanoparticles were carried out at low acceleration voltages with an ultra high resolution Scanning Electron Microscope, Auriga Cross Beam FIB–SEM from Carl Zeiss equipped with a Gemini electron beam column and in-lens detector technology, at IMDEA Nanociencia. The samples were deposited on gold substrates and imaged using a 2–3 kV beam acceleration voltages.

Acknowledgements

Financial support for this research was obtained from the Spanish Ministerio de Economía y Competitividad (BIO2012-34835 A.L.C. and J.M.A.), the European Commission International Reintegration Grant (IRG-246688) (A.L.C.), Marie Curie COFUND “AMAROUT-Europe” Programme (A.L.C.) and ERC-CoG-648071 (A.L.C.). S.H.M. thanks IMDEA-Nanociencia for financial support through an “Ayuda de Iniciación a la Investigación” fellowship and

the Basque Government for financial support (PhD Scholarship). JMA acknowledges research funding by a “Ramon y Cajal” contract from the Spanish Ministry of Science and Innovation. We thank Prof. Encarnación Lorenzo for providing QCM measurement facilities.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.colsurfb.2016.01.039>.

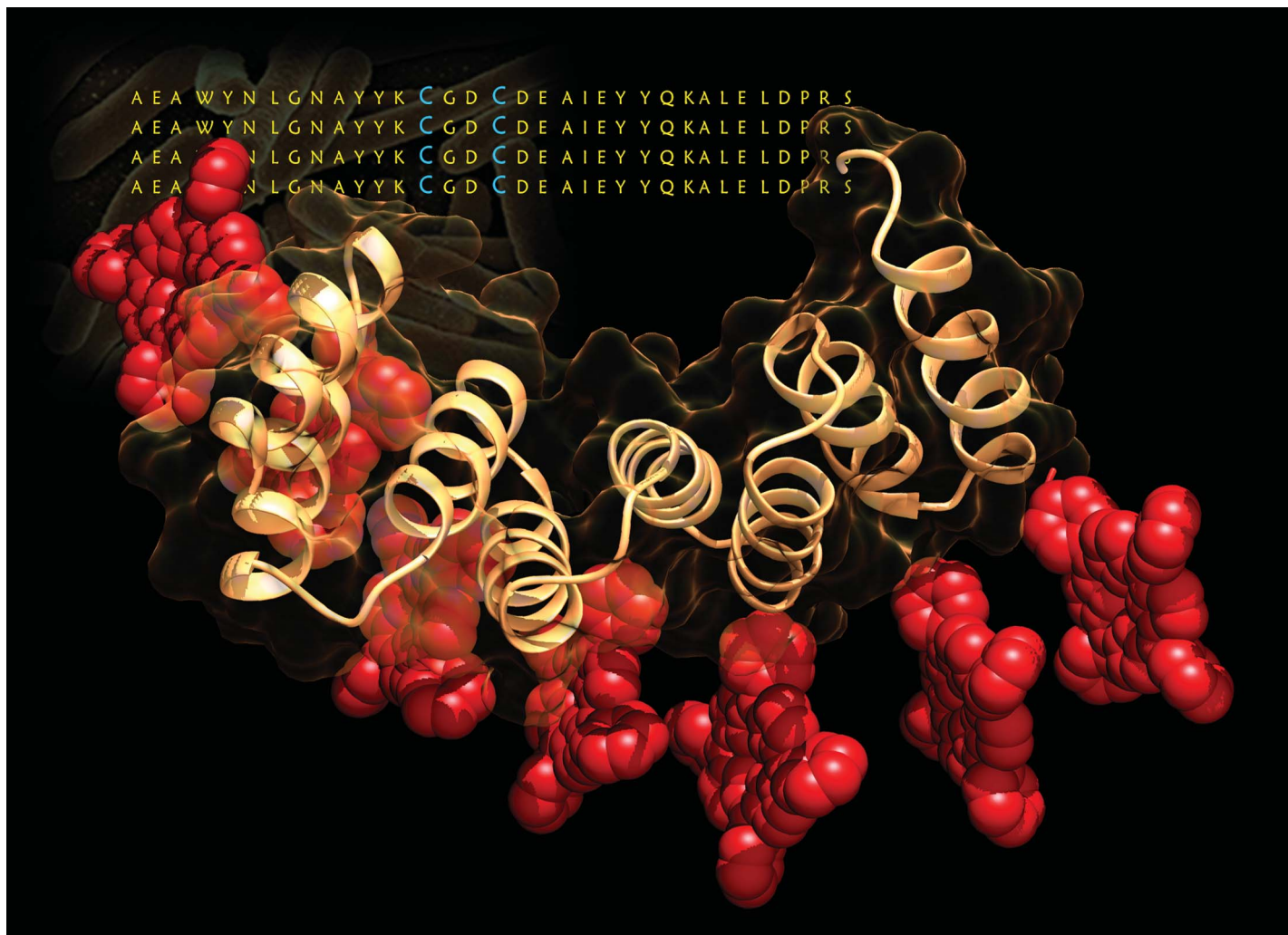
References

- [1] A.P. Alivisatos, Nanocrystals: building blocks for modern materials design, *Endeavour* 21 (1997) 56–60.
- [2] T. Pellegrino, S. Kudera, T. Liedl, A. Muñoz Javier, L. Manna, W.J. Parak, On the development of colloidal nanoparticles towards multifunctional structures and their possible use for biological applications, *Small* 1 (2005) 48–63.
- [3] V.F. Puentes, K.M. Krishnan, A.P. Alivisatos, Colloidal nanocrystal shape and size control: the case of cobalt, *Science* 291 (2001) 2115–2117.
- [4] A.M. Jackson, J.W. Myerson, F. Stellacci, Spontaneous assembly of subnanometre-ordered domains in the ligand shell of monolayer-protected nanoparticles, *Nat. Mater.* 3 (2004) 330–336.
- [5] W.J. Parak, D. Gerion, T. Pellegrino, D. Zanchet, C. Micheel, S.C. Williams, R. Boudreau, M.A. Le Gros, C.A. Larabell, A.P. Alivisatos, Biological applications of colloidal nanocrystals, *Nanotechnology* 14 (2003) R15.
- [6] G.A. Devries, M. Brunnbauer, Y. Hu, A.M. Jackson, B. Long, B.T. Neltner, O. Uzun, B.H. Wunsch, F. Stellacci, Divalent metal nanoparticles, *Science* 315 (2007) 358–361.
- [7] E. Gazit, Bioinspired chemistry: diversity for self-assembly, *Nat. Chem.* 2 (2010) 1010–1011.
- [8] D.N. Woolfson, Z.N. Mahmoud, More than just bare scaffolds: towards multi-component and decorated fibrous biomaterials, *Chem. Soc. Rev.* 39 (2010) 3464–3479.
- [9] R.V. Ulijn, D.N. Woolfson, Peptide and protein based materials in 2010: from design and structure to function and application, *Chem. Soc. Rev.* 39 (2010) 3349–3350.
- [10] K. Rajagopal, J.P. Schneider, Self-assembling peptides and proteins for nanotechnological applications, *Curr. Opin. Struct. Biol.* 14 (2004) 480–486.
- [11] S.A. Maskarinec, D.A. Tirrell, Protein engineering approaches to biomaterials design, *Curr. Opin. Biotechnol.* 16 (2005) 422–426.
- [12] R. Langer, T. Da, Designing materials for biology and medicine, *Nature* 428 (2004) 487–492.
- [13] T. Kajander, A.L. Cortajarena, L. Regan, Consensus design as a tool for engineering repeat proteins, *Methods Mol. Biol.* 340 (2006) 151–170.
- [14] D.M. Ridgley, B.G. Freedman, P.W. Lee, J.R. Barone, Genetically encoded self-assembly of large amyloid fibers, *Biomater. Sci.* 2 (2014) 560–566.
- [15] M. Sarikaya, C. Tamerler, D. Schwartz, F.N. Baneyx, Materials assembly and formation using engineered polypeptides, *Annu. Rev. Mater. Res.* 34 (2004) 373–408.
- [16] P.F. McMillan, New materials from high-pressure experiments, *Nat. Mater.* 1 (2002) 19–25.
- [17] M. Zin, H. Ma, M. J.A.K. Sarikaya, Assembly of gold nanoparticles using genetically engineered polypeptides, *Small* 1 (2005) 698–702.
- [18] V.V. Zhirnov, D.J.C. Herr, New frontiers: self-assembly in nanoelectronics, *IEEE Comput.* 34 (2001) 33–43.
- [19] M. Asbahi, K.T.P. Lim, F. Wang, H. Duan, N. Thiagarajah, V. Ng, J.K.W. Yang, Directed self-assembly of densely packed gold nanoparticles, *Langmuir* 28 (2012) 16782–16787.
- [20] C. Tamerler, M. Sarikaya, Molecular biomimetics: nanotechnology and bionanotechnology using genetically engineered peptides, *Philos. Trans. A Math. Phys. Eng. Sci.* 367 (2009) 1705–1726.
- [21] L. Hu, M. Chen, X. Fang, L. Wu, Oil–water interfacial self-assembly: a novel strategy for nanofilm and nanodevice fabrication, *Chem. Soc. Rev.* 41 (2012) 1350–1362.
- [22] D. Romera, P. Couleaud, S.H. Mejias, A. Aires, A.L. Cortajarena, Biomolecular templating of functional hybrid nanostructures using repeat protein scaffolds, *Biochem. Soc. Trans.* 34 (2015) 825–831.
- [23] L. D’Andrea, L. Regan, TPR proteins: the versatile helix, *Trends Biochem. Sci.* 28 (2003) 655–662.
- [24] E.R.G. Main, A.R. Lowe, S.G.J. Mochrie, S.E. Jackson, L. Regan, A recurring theme in protein engineering: the design, stability and folding of repeat proteins, *Curr. Opin. Struct. Biol.* 15 (2005) 464–471.
- [25] T.Z. Grove, L. Regan, New materials from proteins and peptides, *Curr. Opin. Struct. Biol.* 22 (2012) 451–456.
- [26] E.R. Main, J.J. Phillips, C. Millership, Repeat protein engineering: creating functional nanostructures/biomaterials from modular building blocks, *Biochem. Soc. Trans.* 41 (2013) 1152–1158.
- [27] T. Kajander, A.L. Cortajarena, S.G. Mochrie, L. Regan, Structure and stability of a consensus TPR superhelix, *Acta Crystallogr. D* D63 (2007) 800–811.

- [28] T. Kajander, A.L. Cortajarena, E.R. Main, S.G. Mochrie, L. Regan, A new folding paradigm for repeat proteins, *J. Am. Chem. Soc.* 127 (2005) 10188–10190.
- [29] T.Z. Grove, M. Hands, L. Regan, Creating novel proteins by combining design and selection, *Protein Eng. Des. Sel.* 23 (2010) 449–455.
- [30] A.L. Cortajarena, L. Regan, Ligand binding by TPR domains, *Protein Sci.* 15 (2006) 1193–1198.
- [31] A.L. Cortajarena, S.G. Mochrie, L. Regan, Modulating repeat protein stability: the effect of individual helix stability on the collective behavior of the ensemble, *Protein Sci.* 20 (2011) 1042–1047.
- [32] T.Z. Grove, C.O. Osuji, J.D. Forster, E.R. Dufresne, L. Regan, Stimuli-responsive smart gels realized via modular protein design, *J. Am. Chem. Soc.* 132 (2010) 14024–14026.
- [33] A. Sanz de Leon, J. Rodriguez-Hernandez, A.L. Cortajarena, Honeycomb patterned surfaces functionalized with polypeptide sequences for recognition and selective bacterial adhesion, *Biomaterials* 34 (2013) 1453–1460.
- [34] M. Palacios-Cuesta, A.L. Cortajarena, O. García, J. Rodríguez-Hernández, Versatile functional microstructured polystyrene-based platforms for protein patterning and recognition, *Biomacromolecules* 14 (2013) 3147–3154.
- [35] T.Z. Grove, L. Regan, A.L. Cortajarena, Nanostructured functional films from engineered repeat proteins, *J. R. Soc. Interface* 10 (2013) 20130051.
- [36] J.J. Phillips, C. Millership, E.R. Main, Fibrous nanostructures from the self-assembly of designed repeat protein modules, *Angew. Chem.* 51 (2012) 13132–13135.
- [37] S.H. Mejias, B. Sot R. Guantes A.L., Cortajarena, Controlled nanometric fibers of self-assembled designed protein scaffolds, *Nanoscale* (2014). (in press).
- [38] A.L. Cortajarena, J. Wang, L. Regan, Crystal structure of a designed TPR module in complex with its peptide-ligand, *FEBS J.* 277 (2010) 1058–1066.
- [39] N.A. Carter, T.Z. Grove, Repeat-proteins films exhibit hierarchical anisotropic mechanical properties, *Biomacromolecules* 16 (2015) 706–714.
- [40] A.L. Cortajarena, L. Regan, Calorimetric study of a series of designed repeat proteins: modular structure and modular folding, *Protein Sci.* 20 (2011) 341–352.
- [41] Y.-S. Shon, H.Y. Choi, M.S. Guerrero, C. Kwon, Preparation of nanostructured film arrays for Transmission localized surface plasmon sensing, *Plasmonics* 4 (2009) 95–105.
- [42] M.-C. Chen, Y.-L. Yang, S.-W. Chen, J.-H. Li, M. Aklilu, Y. Tai, Self-assembled monolayer immobilized gold nanoparticles for plasmonic effects in small molecule organic photovoltaic, *ACS Appl. Mater. Interfaces* 5 (2013) 511–517.
- [43] X. Zhang, J. Zhang, H. Wang, Y. Hao, X. Zhang, T. Wang, Y. Wang, R. Zhao, H. Zhang, B. Yang, Thermal-induced surface plasmon band shift of gold nanoparticle monolayer: morphology and refractive index sensitivity, *Nanotechnology* 21 (2010) 465702.
- [44] A.L. Cortajarena, G. Lois, E. Sherman, C.S. O'Hern, L. Regan, G. Haran, Extensive non-native polyproline II secondary structure induces compaction of a protein's denatured state, *J. Mol. Biol.* 382 (2008) 203–212.
- [45] J.C. Love, L.A. Estroff, J.K. Kriebel, R.G. Nuzzo, G.M. Whitesides, Self-assembled monolayers of thiolates on metals as a form of nanotechnology, *Chem. Rev.* 105 (2005) 1103–1169.
- [46] J.L. Wilbur, G.M. Whitesides, Self-assembly and self-assembled monolayers in micro- and nanofabrication, in: G. Timp (Ed.), *Nanotechnology*, Springer, 1999, pp. 331–369.
- [47] L.H. Dubois, R.G. Nuzzo, Synthesis, structure and properties of model organic surfaces, *Annu. Rev. Phys. Chem.* 43 (1992) 437–463.
- [48] S. Nakata, N. Kido, M. Hayashi, M. Hara, H. Sasabe, T. Sugawara, T. Matsuda, Chemisorption of proteins and their thiol derivatives onto gold surfaces: characterization based on electrochemical nonlinearity, *Biophys. Chem.* 62 (1996) 63–72.
- [49] V. Silin, H. Weetall, D.J. Vanderah, SPR studies of the nonspecific adsorption kinetics of human IgG and BSA on gold surfaces modified by self-assembled monolayers (SAMs), *J. Colloid Interface Sci.* 185 (1997) 94–103.
- [50] A. Serrano, O. Rodríguez de la Fuente, M.A. García, Extended and localized surface plasmons in annealed Au films on glass substrates, *J. Appl. Phys* 108 (2010) 074303.
- [51] F. Galvez, C. Monton, A. Serrano, I. Valmianski, J. de la Venta, I.K. Schuller, M.A. García, Effect of photodiode angular response on surface plasmon resonance measurements in the Kretschmann–Raether configuration, *Rev. Sci. Instrum.* 83 (2012) 093102.
- [52] K.M.M. Aung, X. Ho, X. Su, DNA assembly on streptavidin modified surface: a study using quartz crystal microbalance with dissipation or resistance measurements, *Sens. Actuators B: Chem.* 131 (2008) 371–378.
- [53] C. Bustamante, J. Vesenka, C.L. Tang, W. Rees, M. Guthold, R. Keller, Circular DNA molecules imaged in air by scanning force microscopy, *Biochemistry* 31 (1992) 22–26.
- [54] G. Lee, K. Abdi, Y. Jiang, P. Michaely, V. Bennett, P.E. Marszalek, Nanospring behaviour of ankyrin repeats, *Nature* 440 (2006) 246–249.
- [55] S.S. Cohen, I. Riven, A.L. Cortajarena, L. De Rosa, L.D. D'Andrea, L. Regan, G. Haran, Probing the Molecular origin of native-state flexibility in repeat proteins, *J. Am. Chem. Soc.* 137 (2015) 10367–10373.
- [56] T. Matsuura, H. Tanaka, T. Matsumoto, T. Kawai, Atomic force microscopic observation of *Escherichia coli* ribosomes in solution, *Biosci. Biotechnol. Biochem.* 70 (2006) 300–302.
- [57] R. Lucklum, C. Behling, P. Hauptmann, Role of mass accumulation and viscoelastic film properties for the response of acoustic-wave-based chemical sensors, *Anal. Chem.* 71 (1999) 2488–2496.
- [58] T. Kajander, A.L. Cortajarena, S.G. Mochrie, L. Regan, Structure and stability of a consensus TPR superhelix, *Acta Crystallogr. D63* (2007) 800–811.
- [59] D. Talapin, J.-S. Lee, M. Kovalenko, E. Shevchenko, Prospects of colloidal nanocrystals for electronic and optoelectronic applications, *Chem. Rev.* 110 (2010) 389–458.
- [60] C.N. Pace, F. Vajdos, L. Fee, G. Grimsley, T. Gray, How to measure and predict the molar absorption coefficient of a protein, *Protein Sci.* 4 (1995) 2411–2423.
- [61] J.M. Abad, M. Velez, C. Santamaria, J.M. Guisan, P.R. Matheus, L. Vazquez, I. Gazaryan, L. Gorton, T. Gibson, V.M. Fernandez, Immobilization of peroxidase glycoprotein on gold electrodes modified with mixed epoxy-boronic acid monolayers, *J. Am. Chem. Soc.* 124 (2002) 12845–12853.
- [62] M. Darder, E. Casero, D.J. Díaz, H.D. Abuña, F. Pariente, E. Lorenzo, Concentration dependence of aggregate formation upon adsorption of 5-(Octylthio)-2-nitrobenzoic acid on gold electrodes, *Langmuir* 16 (2001) 9804–9811.
- [63] J. Madoz-Gurpide, J.M. Abad, J. Fernandez-Recio, M. Velez, L. Vazquez, C. Gomez-Moreno, V.M. Fernandez, Modulation of electroenzymatic NADPH oxidation through oriented immobilization of ferredoxin: NADP+ reductase onto modified gold electrodes, *J. Am. Chem. Soc.* 122 (2000) 9808–9817.
- [64] M. Rodahl, F. Höök, C. Fredriksson, C.A. Keller, A. Krozer, P. Brzezinski, M. Voinova, B. Kasemo, Simultaneous frequency and dissipation factor QCM measurements of biomolecular adsorption and cell adhesion, *Faraday Discuss.* 107 (1997) 229–246.
- [65] S. Bruckenstein, M. Shay, Experimental aspects of use of the quartz crystal microbalance in solution, *Electrochim. Acta* 30 (1985) 1295–1300.
- [66] M.D. Ward, D.A. Buttry, In situ interfacial mass detection with piezoelectric transducers, *Science* 249 (1990) 1000–1007.
- [67] M. Rodahl, B. Kasemo, On the measurement of thin liquid overlayers with the quartz crystal microbalance, *Sens. Actuators A* 54 (1996) 448–462.
- [68] J. Rickert, A. Brecht, W. Göpel, QCM operation in liquids: constant sensitivity during formation of extended protein multilayers by affinity, *Anal. Chem.* 69 (1997) 1441–1448.
- [69] W. Knoll, Interfaces and thin films as seen by bound electromagnetic waves, *Annu. Rev. Phys. Chem.* 49 (1998) 569–638.
- [70] H. Raether, *Surface Plasmons on Smooth and Rough Surfaces and on Gratings*, Berlin, (1998).
- [71] E. Stenberg, B. Persson, H. Roos, C. Urbaniczky, Quantitative determination of surface concentration of protein with surface plasmon resonance using radiolabeled proteins, *J. Colloid Interface Sci.* 143 (1991) 513–526.
- [72] P.J. de Pablo, J. Colchero, J. Gomez-Herrero, A.M. Baro, Jumping mode scanning force microscopy, *Appl. Phys. Lett.* 73 (1998) 3300–3302.
- [73] I. Horcas, R. Fernández, J.M. Gómez-Rodríguez, J. Colchero, J. Gómez-Herrero, A.M. Baro, WsXM: A software for scanning probe microscopy and a tool for nanotechnology, *Rev. Sci. Instrum.* 78 (2007) 13705–13713.
- [74] J. Turkevich, P.C. Stevenson, J. Hillier, A study of the nucleation and growth processes in the synthesis of colloidal gold, *Discuss. Faraday Soc.* 11 (1951) 55–75.
- [75] A. Latorre, C. Posch, Y. Garcimart, M. Sanlorenzo, I. Vujic, M. Zekhtser, K. Rappersberger, S. Ortiz-urda, DNA and aptamer stabilized gold nanoparticles for targeted delivery of anticancer therapeutics, *Nanoscale* 6 (2014) 7436–7442.

Appendix 3.

**Repeat protein scaffolds: ordering photo-
and electroactive molecules in solution and
solid state**



Showcasing research from Nazario Martín's laboratory, Organic Chemistry Department, Complutense University of Madrid, Spain.

Repeat protein scaffolds: ordering photo- and electroactive molecules in solution and solid state

Protein engineering for the construction of new macroscopically ordered electroactive materials. A biomolecular scaffold based on engineered repeat proteins has been used to organize porphyrin molecules at the required distances along the protein structure. Moreover, an ordered bio-organic thin film has been achieved exhibiting anisotropic photoconductivity.

As featured in:



See Carmen Atienza, Nazario Martín, Aitziber L. Cortajarena *et al.*, *Chem. Sci.*, 2016, 7, 4842.



www.rsc.org/chemicalscience

Registered charity number: 207890

Cite this: *Chem. Sci.*, 2016, 7, 4842

Repeat protein scaffolds: ordering photo- and electroactive molecules in solution and solid state†

Sara H. Mejías,^{‡a} Javier López-Andarias,^{‡c} Tsuneaki Sakurai,^d Satoru Yoneda,^d Kevin P. Erazo,^a Shu Seki,^d Carmen Atienza,^{*c} Nazario Martín^{*ac} and Aitziber L. Cortajarena^{*abe}

The precise control over the organization of photoactive components at the nanoscale is one of the main challenges for the generation of new and sophisticated macroscopically ordered materials with enhanced properties. In this work we present a novel bioinspired approach using protein-based building blocks for the arrangement of photo- and electroactive porphyrin derivatives. We used a designed repeat protein scaffold with demonstrated unique features that allow for the control of their structure, functionality, and assembly. Our designed domains act as exact biomolecular templates to organize porphyrin molecules at the required distance. The hybrid conjugates retain the structure and assembly properties of the protein scaffold and display the spectroscopic features of orderly aggregated porphyrins along the protein structure. Finally, we achieved a solid ordered bio-organic hybrid thin film with anisotropic photoconductivity.

Received 23rd March 2016
Accepted 23rd May 2016

DOI: 10.1039/c6sc01306f

www.rsc.org/chemicalscience

Introduction

One of the main challenges in current science is the construction of highly ordered materials bearing photonic and/or electronic active units.^{1,2} Nowadays, many bottom-up approaches have been undertaken to control the organization and morphology of electro- and photoactive components at the nanometer scale, enabling the formation of supramolecular architectures that typically result in a significant improvement of their properties.^{3–5} Nonetheless, the majority of these methodologies lack the selective positioning and orientation of the different functional components, and the control at different length scales is still challenging. In addition, poor monodispersity of the aggregates is another drawback. However, templating these components with bioinspired scaffolds can provide the control over the location, monodispersity and chirality needed for technological purposes. Furthermore, bioinspired self-assembling based on biological building blocks including peptides, proteins and DNA has recently been explored to construct sophisticated macroscopic materials that are hierarchically organized through self-assembly.^{6–14}

Promising examples show the use of biomolecules for ordering organic materials. For example, DNA^{15,16} and self-assembling peptide nanotubes¹⁷ have been used to arrange photo- and electroactive molecules such as fullerenes. Similarly, helical peptides and polymer-peptide have been applied to pattern semiconductors and organic chromophores into ordered structures.^{18,19} Porphyrins represent an important class of photo- and electroactive molecules which have been supramolecularly organized using a variety of biomolecules. Thus, as representative examples, the tobacco mosaic virus,²⁰ peptides,^{21–23} and nucleic acids^{24,25} have been used as scaffolds for precisely controlling the position and orientation of porphyrins. These works represent the proof of concept that bioinspired assembly can be successfully used to generate more efficient organo-electronic devices. However, an important limitation of these approaches is how to transfer hierarchically this organization to the macroscopic level in the solid state. The combination of both the generation of bio-organic conjugates with photonic and/or electronic properties and the macroscopic organization of these building blocks could lead to the construction of functional bulk biomaterials with applications in fields such as organic electronics and photovoltaics to name a few. For this end, the design of simple biological building blocks with defined functionalities able to assemble into structured materials is a key issue.

Results and discussion

In this work, we focus on the design and application of a particular type of biomolecular scaffold based on engineered repeat proteins which encompass the structural simplicity of

^aIMDEA-Nanoscience, Campus de Cantoblanco, E-28049 Madrid, Spain^bCIC biomAGUNE, Paseo de Miramón 182, E-20009 Donostia-San Sebastian, Spain^cDepartamento de Química Orgánica I, Facultad de Ciencias Químicas, Universidad Complutense de Madrid, E-28040 Madrid, Spain. E-mail: nazmar@ucm.es^dDepartment of Applied Chemistry, Graduate School of Engineering, Osaka University, Japan^eIkerbasque, Basque Foundation for Science, Ma Díaz de Haro 3, E-48013 Bilbao, Spain

† Electronic supplementary information (ESI) available: Supplementary Fig. 1–7, experimental details, characterization and NMR data for all new compounds. See DOI: 10.1039/c6sc01306f

‡ S. H. Mejías and J. López-Andarias contributed equally to this work.



DNA and short peptides and the functional versatility of proteins.²⁶ Specifically, the repeat module of choice is a designed consensus tetratricopeptide repeat protein (CTPR).^{27,28} The CTPR module is a 34 amino acid sequence that folds in a helix-turn-helix structure. Few conserved residues define the TPR fold, which permits the rational manipulation of the protein scaffold while retaining the structure. CTPR repeats combined in tandem form superhelical arrays with different number of repeats from 2 to 20, so-called CTPR2 to CTPR20 proteins.^{29,30} Previous works on this protein system showed that their stability,^{31,32} and ligand binding properties^{33,34} can be tuned in a predictable manner, which opens the opportunity to tailor the properties of the individual building block in a modular manner for specific applications.

In line with the creation of protein-based nanostructured materials, we have recently shown that CTPR modules have intrinsic self-assembling properties.^{35–37} CTPR proteins retain their structure and functionality in the solid state, creating a nanostructured macroscopic film through head-to-tail and side-to-side interactions similar to those presented in their crystalline form.³⁶ This was the first example in which a solid-state protein film displayed both structural and functional properties.

Taking into account these two unique features of the CTPR scaffold: (i) the structural understanding of the CTPR module and the small number of conserved residues that define the CTPR fold to accurately modify selected positions of the amino acid sequence without perturbing its structure and, (ii) the extraordinary capability of CTPR scaffolds to form nanostructured solid films, these proteins offer an unique opportunity to fabricate well-ordered hybrid bio-organic materials. To the best of our knowledge, this combination still remains unexplored.

Hence, we have designed a modified CTPR protein as scaffold for assembling a specific number of porphyrin chromophores in a defined distance and orientation. Based on the crystal structure of the CTPR protein,³⁴ different amino acid positions were selected to arrange the chromophores along the CTPR superhelix that allow for efficient π - π interactions between the molecules. The selected positions are not conserved between CTPR modules, thus it is expected that their modification will not affect the structure of the protein template.

Specifically, two cysteine residues were introduced in each repeated unit to provide unique reactivity for the immobilization of the porphyrin derivatives. A CTPR protein with four identical repeats was generated resulting in a protein with eight quasi-equally spaced cysteine residues in four loops of the protein (Fig. 1a and b). The distance between the Cys side chains is between 7 and 8 Å, in the order of the distance required to establish π - π interactions between the porphyrin rings. As expected, the mutations did not significantly affect the helical structure of the protein scaffold (Fig. 1c).

Besides, in the design of the porphyrin moiety two important requirements should be considered: (i) porphyrin molecules must exhibit high solubility in water but avoiding charged functional groups, since undesired interactions with charged

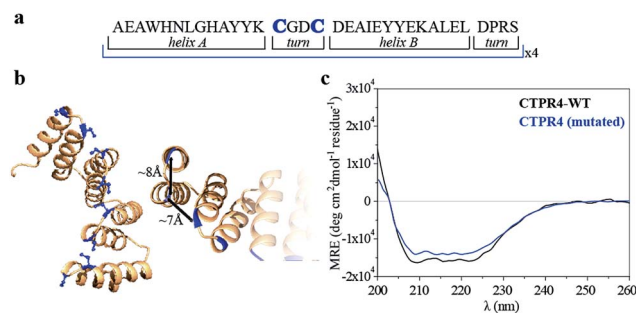


Fig. 1 (a) Sequence of amino acids of the repeat module, highlighting the positions 14 and 17 that have been mutated to cysteine. (b) Ribbon representation of mutated CTPR4 protein structural model based on the structure of CTPR8 (PDB ID:2HYZ). The mutated cysteine residues for the porphyrin attachment are highlighted in blue. (c) CD spectra of mutated CTPR4 comparing with the original CTPR4.

side-chains of the amino acids in the protein should be eliminated and, (ii) an efficient cross-linker must be placed in the porphyrin structure in order to promote quantitative or quasi-quantitative conjugation with the eight cysteine residues of the designed protein. With this in mind, two porphyrin derivatives were synthesized, free-base and zinc-metalloporphyrin (**1** and **2**, respectively), decorated with twelve triethylene glycol water-soluble tails and a maleimide reactive group as efficient cross-linker for the conjugation reaction (Fig. 2a and see ESI Scheme S1†). For further information about the design, synthesis and characterization of porphyrin derivatives, see ESI.†^{38,39}

The designed protein (CTPR4) and porphyrin moieties (**1** and **2**) (Fig. 2a and b) were conjugated using the maleimide–cysteine

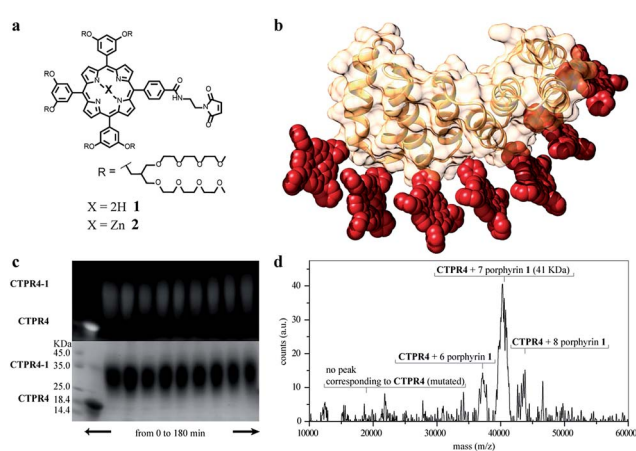


Fig. 2 (a) Molecular structure of porphyrin derivatives **1** and **2**. (b) 3D model structure of the CTPR4-1 conjugate (triethylene glycol chains are omitted for clarity). (c) SDS-PAGE gel electrophoresis of the CTPR4-1 conjugation at different reaction times (20–180 every 20 min, lanes 3–11). Molecular weight marker (lane 1) and CTPR4 control with fluorescent loading buffer (lane 2). The gel is imaged using UV-light to monitor the fluorescence of the porphyrins (upper panel) and after Coomassie Blue staining (lower panel). (d) MALDI-TOF spectrum of the final conjugation reaction in CTPR4-1. A main peak appears at 41 kDa that corresponds to the mass of CTPR4 protein (19 163 Da) with 7 porphyrins (MW of **1**: 3160 Da).



chemistry. At different times during the conjugation reaction, the analysis of the samples by gel electrophoresis showed that a conjugate of higher molecular weight than **CTPR4** protein (19 163 Da, see ESI Fig. S1†) and porphyrins ($MW_1 = 3160$ Da, $MW_2 = 3223$ Da) was obtained even at the shortest reaction time (Fig. 2c and see ESI Fig. S2†). When the gel was imaged without staining, the fluorescence signal of the porphyrin could be detected showing a new band between 31 and 45 kDa, corresponding to the molecular weight expected for the **CTPR4-1** or **CTPR4-2** conjugates (Fig. 2c – top). Moreover, the staining of the protein with Coomassie Blue confirmed that the higher molecular weight band was composed of both protein and porphyrin (Fig. 2c – bottom).

To quantify the number of porphyrin molecules bound per protein in the conjugation reaction, mass spectrometry was used (Fig. 2d and see ESI Fig. S2†). After 4 h reaction time, the peak corresponding to the **CTPR4** protein disappeared and a most abundant peak at 41 kDa appeared, which corresponds to a **CTPR4** protein covalently linked to seven porphyrin units, together with two smaller peaks both at higher and lower molecular weights, corresponding to the conjugate with eight and six porphyrins, respectively. The purification of the protein-porphyrin conjugates from the excess of free porphyrin is an essential step for further use and characterization of homogeneous hybrid structures. **CTPR4-1** and **CTPR4-2** purification was successfully carried out using size exclusion chromatography (see ESI Fig. S3†). All the characterization experiments were performed using the purified conjugates.

As has previously been mentioned, the distinctive properties of the CTPR scaffold allowed for the design of a **CTPR4** protein with an appropriate distance between the reactive moieties that, considering the rotational freedom of both the protein side chains and the linked porphyrins, will enable an efficient π - π intermolecular interaction between the porphyrin rings. It is well established that porphyrins have the ability to self-aggregate noncovalently to form H-aggregates (face-to-face) or J-aggregates (side-to-side), these states are characterized by a shift on the Soret absorption band towards the blue or red, respectively, compared to the monomeric absorption band.^{40–45} Thus, the UV-vis spectra of both porphyrins as references and **CTPR4-1** and **CTPR4-2** conjugates were recorded in order to obtain more information about the conformation of the porphyrins in the conjugates. The absorption spectrum of porphyrin **1** in a PBS buffer solution presents an intense Soret band at 406 nm and four Q-bands in the region between 500 and 700 nm. Meanwhile, porphyrin **2** presents the Soret band at 425 nm and two Q-bands at 555 and 595 nm. In contrast, the UV-vis spectra for **CTPR4-2** and **CTPR4-1** conjugates showed a red shift of 5 and 17 nm in the Soret band of the porphyrin, respectively, compared with the corresponding free porphyrins (Fig. 3a and see ESI Fig. S4†). These shifts suggest the formation of a J-type aggregate of porphyrins in both conjugates, certainly induced by the geometry of the protein and the position of the cysteine residues in the framework. To further confirm the π - π interaction between porphyrin moieties within the conjugates, UV-vis spectra were measured at different H₂O : MeOH ratio leading to an increased intensity and a blue-shift of the Soret

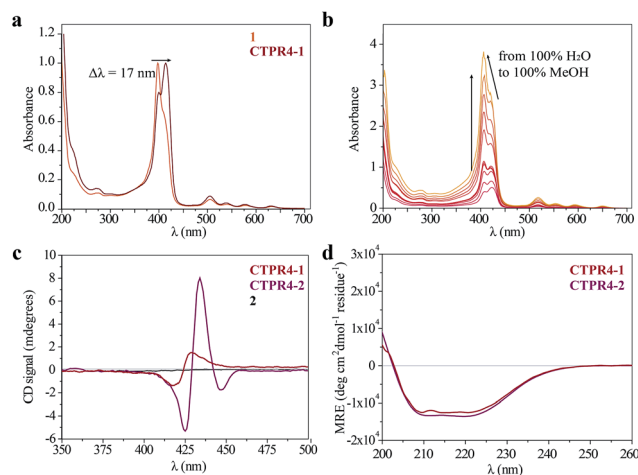


Fig. 3 (a) Normalized UV-vis spectrum of **CTPR4-1** conjugate compared to normalized UV-vis spectrum of **1** in PBS buffer. (b) UV-vis spectra of **CTPR4-1** (0.56 μ M) at different methanol concentrations (from 0 to 100%). (c) CD spectra of porphyrin **2** and **CTPR4-1** and **CTPR4-2** conjugates (1.70 μ M) in PBS buffer in the spectral region of the porphyrin absorption. (d) CD spectra of **CTPR4-1** and **CTPR4-2** conjugates in PBS buffer in the spectral region of the protein's secondary structure absorption.

band while the percentage of methanol increases (Fig. 3b and see ESI Fig. S4†). This fact is in agreement with the disruption of the π - π interaction when protein denaturation is achieved with an organic polar solvent.

Circular dichroism (CD) studies in solution were performed on the **CTPR4-1** and **CTPR4-2** conjugates. In the first case, the **CTPR4-1** conjugate showed a bisignate signal with a negative Cotton effect at 417 nm, a positive Cotton effect at 428 nm and a zero cross point at 423 nm (Fig. 3c). The presence of these Cotton effects, that is, exciton-coupled circular dichroism, is indicative of a close proximity between the chromophores, as previously reported.^{46–48} In the second case, a strong trisignate signal with a negative Cotton effect at 425 nm, positive Cotton effect at 434 nm and another negative Cotton effect at 446 nm with zero cross points at 429 and 442 nm was shown for the **CTPR4-2** conjugate (Fig. 3c). Multisignates in the Soret band have previously been ascribed in the literature to strong π - π stacking interactions between close porphyrins.^{25,49} Thus, these dichroic signals in the porphyrin absorption region demonstrated the transfer of chirality from the CTPR protein scaffold to the porphyrin molecules, since free porphyrins **1** and **2** lacked optical activity in their CD spectra (Fig. 3c). Furthermore, the feature CD signal for alpha-helical secondary structure of the **CTPR4** protein was totally retained in both **CTPR4-1** and **CTPR4-2** conjugates in PBS buffer (Fig. 3d). This result confirmed the successful ability of CTPR proteins to be used as a robust scaffold for ordering organic chromophores, preserving its secondary and tertiary structure even when six to eight porphyrin molecules per protein were incorporated. It is important to note that the final molecular weight of both conjugates represent more than twice the initial one corresponding to the protein alone; however, the structural integrity



of the biological framework is remarkably conserved. Moreover, a denaturalization was also carried out for both conjugates through CD measurements. No appreciable dichroic signal in the range of Soret band of the porphyrins was observed after adding methanol, which confirms the rupture of π - π interactions when the protein scaffold is unstructured (See ESI Fig. S5†).

The intrinsic capability of CTPR proteins to assemble into highly ordered thin films makes them good candidates to organize porphyrin arrays in the solid state.³⁶ However, the formation of protein-based solid ordered materials has been achieved in only a very few protein systems. Furthermore, the combination with the building-block tunability is unique of the CTPR scaffold. In this sense, ordered thin films with both conjugates were effectively generated under similar experimental conditions to those described for CTPR proteins alone. Importantly, the structural integrity of the protein block and the porphyrin arrangement was tested by CD experiments (Fig. 4a and b). Fig. 4a and b show how the secondary structure of CTPR4 scaffold remained alpha helical with no evidence of any significant structure other than alpha helix, and how the chiral environment of the porphyrins was maintained in both conjugates. Moreover, as has previously been reported, CTPR proteins are macroscopically aligned in the solid film form,^{36,37} thus it was expected that our conjugates would also be aligned. To shed light on this fact, fluorescence anisotropy measurements and X-ray powder diffraction (XRD) were carried out. On the one hand, Fig. 4c shows the change in the fluorescence intensity of the porphyrin moieties when the emission polarizer was rotated from 0 to 360°. The signal showed clear maximum and minimum values, which indicates the anisotropy of the sample. The data can be well fitted to a sine wave function with

maximum to minimum peak distance of 107°, in agreement with the alignment phase determined for the CTPR films. This result indicates that the macroscopic ordered pattern of the film is indeed imposed on the porphyrins that are, otherwise, isotropic.

On the other hand, a deep understanding on the organization of the films was reached by XRD experiments. The XRD pattern of a film obtained from CTPR4 protein showed a set of three clear and intense sharp peaks at $2\theta = 10.82$, 21.64 and 32.05° that could correspond to a lamellar packing with a periodical d -spacing of 8.18 Å (Fig. 4d). A similar d -spacing has been previously observed in other films based on repeat proteins and α -helical coiled-coil proteins, being related to the meridional spacing for an α -helix structure, dependent on the angle of inclination.^{14,37,50} Overall, these data demonstrate the directional orientation of the protein on a surface when a film is formed. Moreover, it is remarkable that the diffraction patterns observed for the films based on the conjugates CTPR4-1 and CTPR4-2 present exactly the same reflections peaks as CTPR4. These experimental findings corroborate that the long-range order of the protein is preserved even with these number of porphyrins introduced in its structure. This structural feature highlights the robustness of this biological scaffold to be used as an efficient template for ordering organic chromophores not only in solution but in the technologically relevant solid state.

Organized porphyrin arrays on the protein scaffold are interesting as charge carrier transporting and photoconductive motifs. The photoconductivity along the arrays in CTPR4-1 and CTPR4-2 conjugate films was examined by using the flash-photolysis time-resolved microwave conductivity (FP-TRMC) technique.^{51,52} In this non-contact method, charge carriers are photo-generated upon exposure to 355 nm laser pulses to the films. Then, local-scale motions of the generated charge carriers can be probed by monitoring the reflected microwave (~ 9.1 GHz) power from the film sample set at the microwave cavity. Upon laser flash, a drop-cast film of CTPR4-1 showed a conductivity transient with prompt rise and slow decay features (Fig. 5a). The conductivity ($\varphi \sum \mu$) indicates the product of charge carrier generation efficiency (φ) upon photoexcitation and sum of charge carrier mobilities ($\sum \mu = \mu_h + \mu_e$). Furthermore, the same film yielded a clear transient absorption spectra (TAS), where photo-bleach of the neutral and generation of radical cations were observed at the Q-band region of the free-base porphyrins (Fig. 5b).⁵³ In fact, the normalized profiles of FP-TRMC and TAS at 530 nm gave almost identical kinetic traces (Fig. 5a), indicating that hole transport is the dominant factor for the observed local-scale photoconductivity of CTPR4-1 under air. By using a typical absorption coefficient of the radical cation of tetraphenyl free-base porphyrin,²¹ φ was calculated to be 5.0×10^{-2} , followed by the evaluation of one-dimensional mobility (μ_h) of $1.5 \times 10^{-3} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. Then, through the same measurement processes, μ_h of CTPR4-2 was determined as $1.3 \times 10^{-3} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ (See ESI Fig. S6†). As a control experiment, a non-conjugated protein-based film, that is, a CTPR4 film resulted in charge carriers by irreversible photo-damage upon laser exposure (see ESI Fig. S7†). Thus, the higher conductivity values for CTPR4-1 and CTPR4-2 were due to the

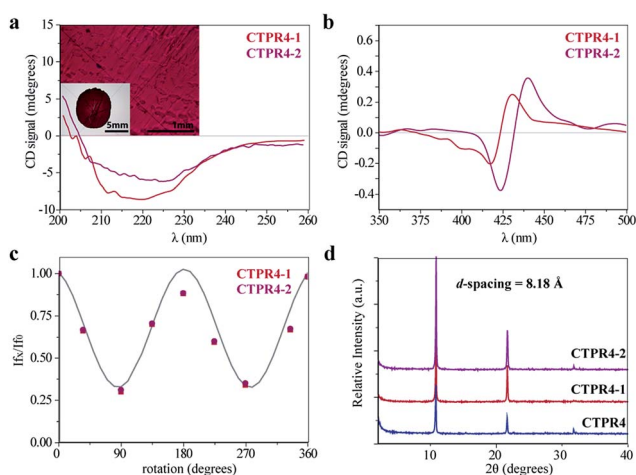


Fig. 4 (a) CD spectra of the CTPR4-1 and CTPR4-2 conjugates in solid thin films in the spectral region of the protein's secondary structure absorption. Inset: photographs of the CTPR4-2 film at different magnifications (scale bars: 5 and 1 mm). (b) CD spectra of the CTPR4-1 and CTPR4-2 conjugates in solid thin films in the spectral region of the porphyrin absorption. (c) Fluorescence anisotropy of the CTPR4-1 and CTPR4-2 films. (d) XRD diffractogram of CTPR4, CTPR4-1 and CTPR4-2 thin films.



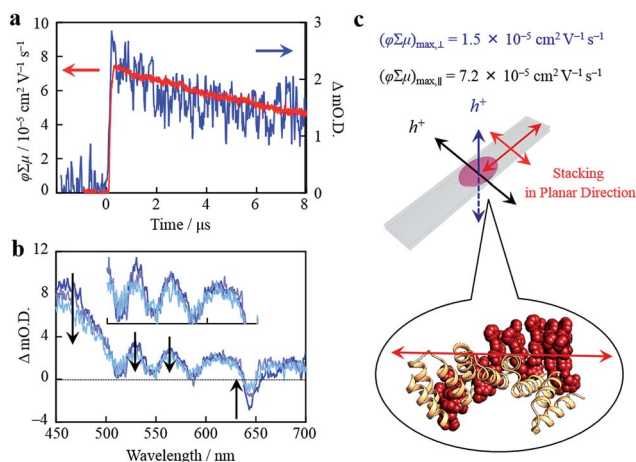


Fig. 5 (a) Kinetic traces of transient absorption spectra at 530 nm (blue) and kinetic traces of FP-TRMC (red) of a film of CTPR4-1. (b) Snapshot of transient absorption spectra of a film of CTPR4-1 at ca. 0.2 (blue), 3 (light purple), and 8 (turquoise) μs after an excitation. (c) Schematic illustration of a CTPR4-1 film on quartz and stacking direction of porphyrin arrays together with observed values of conductivity maxima in perpendicular and planar direction to the substrate surface.

local motion of positive charge carriers as a consequence of the porphyrin arrays. Of further interest, we revealed that structural alignment of the arrays in a macroscopic scale gave rise to the anisotropic electrical conductivity. For example, by changing the direction of the CTPR4-1 sample in the cavity,⁵⁴ $(\varphi \sum \mu)_{\text{max}}$ along the perpendicular and parallel directions to the substrate surface were evaluated as 1.5×10^{-5} and $7.2 \times 10^{-5} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, respectively (Fig. 5c). This increment of the electrical conductivity along the parallel direction is in agreement with the anisotropy of the film observed by fluorescence measurements (Fig. 4c).

Conclusions

This work describes an innovative approach in which a protein building block is designed to organize porphyrin molecules. In contrast to recent works in which existing biomolecular structures have been used to arrange similar molecules, here we have developed a strategy based on a modular protein unit with tunable properties including stability, function and self-assembly. The scaffold is based on a simple unit, which allows for a controlled engineering and introduction of reactive functionalities at defined positions for the conjugation of organic molecules. By combination of simple units we can generate longer proteins while preserving the structural understanding. Therefore, we can pattern the reactive moieties on the 3D structure of the scaffold. Here, we evidence this potential by precisely organizing porphyrin molecules on the CTPR scaffold at the distance required to achieve efficient π - π interactions between the rings. In the final conjugates the protein retains its signature helical structure and imposes order and chirality into the porphyrin molecules that show efficient stacking interactions. These results show the potential of this versatile scaffold

that could be applied for control grafting of a variety of functional molecules and nanostructural elements.

Another feature that makes the applied system superior to other protein templates is its assembly properties. Prior works on the CTPR scaffold have shown how ordered nanostructures and nanostructured materials can be fabricated from these simple repeated units. Remarkably, in the ordered assemblies the proteins maintain their structure and function. These observations pave the way to their application for the fabrication of solid functional devices. In this work, as a proof of concept, we have shown the formation of ordered films using hybrid protein-porphyrin conjugates. The films obtained displayed the described nanostructured directional order both in the protein and in the photoactive components. Finally, the photoconductivity of the hybrid thin films showed a remarkable anisotropy in agreement with the directional order of the photoactive molecules. The developed approach is simple and should be easily translatable to other systems that require precise order at different length scales to achieve materials and devices with enhanced properties.

Acknowledgements

This work has been supported by the European Commission IRG-246688 Bionanotools (ALC), the Spanish Ministry of Economy and Competitiveness (MINECO) BIO2012-34835 (ALC) and CTQ2014-520456-R (NM) and the European Research Council ERC-320441-Chirallcarbon (NM), and ERC-2014-CoG-648071 (ProNANO) (ALC). C. A. thanks to the Ramón y Cajal granted and J. L.-A. thanks to Spanish Ministry of Education for FPU granted. SHM thanks the Basque Government for financial support through a PhD fellowship. KPE gives thanks for financial support from the MIT-Spain internship program.

Notes and references

- 1 A. S. Arico, P. Bruce, B. Scrosati, J.-M. Tarascon and W. van Schalkwijk, *Nat. Mater.*, 2005, **4**, 366–377.
- 2 Z. B. Henson, K. Mullen and G. C. Bazan, *Nat. Chem.*, 2012, **4**, 699–704.
- 3 J. López-Andarias, J. L. López, C. Atienza, F. G. Brunetti, C. Romero-Nieto, D. M. Guldi and N. Martín, *Nat. Commun.*, 2014, **5**, 3763.
- 4 J. López-Andarias, M. J. Rodríguez, C. Atienza, J. L. López, T. Mikie, S. Casado, S. Seki, J. L. Carrascosa and N. Martín, *J. Am. Chem. Soc.*, 2015, **137**, 893–897.
- 5 A. Ruiz-Carretero, P. G. A. Janssen, A. Kaeser and A. P. H. J. Schenning, *Chem. Commun.*, 2011, **47**, 4340–4347.
- 6 S. Zhang, M. A. Greenfield, A. Mata, L. C. Palmer, R. Bitton, J. R. Mantei, C. Aparicio, M. O. de la Cruz and S. I. Stupp, *Nat. Mater.*, 2010, **9**, 594–601.
- 7 T. O. Yeates, *Nat. Nanotechnol.*, 2011, **6**, 541–542.
- 8 J. C. Sinclair, K. M. Davies, C. Venien-Bryan and M. E. M. Noble, *Nat. Nanotechnol.*, 2011, **6**, 558–562.
- 9 J. Sun and R. N. Zuckermann, *ACS Nano*, 2013, **7**, 4715–4732.
- 10 O. D. Krishna and K. L. Kiick, *Biopolymers*, 2010, **94**, 32–48.



- 11 A. Kuzyk, R. Schreiber, Z. Fan, G. Pardatscher, E.-M. Roller, A. Hoge, F. C. Simmel, A. O. Govorov and T. Liedl, *Nature*, 2012, **483**, 311–314.
- 12 J. Liu, Y. Geng, E. Pound, S. Gyawali, J. R. Ashton, J. Hickey, A. T. Woolley and J. N. Harb, *ACS Nano*, 2011, **5**, 2240–2247.
- 13 T. P. J. Knowles, T. W. Oppenheim, A. K. Buell, D. Y. Chirgadze and M. E. Welland, *Nat. Nanotechnol.*, 2010, **5**, 204–207.
- 14 D. Papapostolou, A. M. Smith, E. D. T. Atkins, S. J. Oliver, M. G. Ryadnov, L. C. Serpell and D. N. Woolfson, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 10853–10858.
- 15 Q. Ying, J. Zhang, D. Liang, W. Nakanishi, H. Isobe, E. Nakamura and B. Chu, *Langmuir*, 2005, **21**, 9824–9831.
- 16 S. F. M. van Dongen, J. Clerx, K. Norgaard, T. G. Bloemberg, J. J. L. M. Cornelissen, M. A. Trakselis, S. W. Nelson, S. J. Benkovic, A. E. Rowan and J. M. Nolte, *Nat. Chem.*, 2013, **5**, 945–951.
- 17 C. Reiriz, R. J. Brea, R. Arranz, J. L. Carrascosa, A. Garibotti, B. Manning, J. M. Valpuesta, R. Eritja, L. Castedo and J. R. Granja, *J. Am. Chem. Soc.*, 2009, **131**, 11335–11337.
- 18 R. J. Kumar, J. M. MacDonald, T. B. Singh, L. J. Waddington and A. B. Holmes, *J. Am. Chem. Soc.*, 2011, **133**, 8564–8573.
- 19 N. Bhagwai and K. L. Kiick, *J. Mater. Chem. C*, 2013, **1**, 4836–4845.
- 20 M. Endo, M. Fujitsuka and T. Majima, *Chem.–Eur. J.*, 2007, **13**, 8660–8666.
- 21 T. E. Kang, H.-H. Cho, C.-H. Cho, K.-H. Kim, H. Kang, M. Lee, S. Lee, B. Kim, C. Im and B. J. Kim, *ACS Appl. Mater. Interfaces*, 2013, **5**, 861–868.
- 22 J. R. Dunetz, C. Sandstrom, E. R. Young, P. Baker, S. A. Van Name, T. Cathopolous, R. Fairman, J. C. de Paula and K. S. Åkerfeldt, *Org. Lett.*, 2005, **7**, 2559–2561.
- 23 Y.-S. Huang, X. Yang, E. Schwartz, L. P. Lu, S. Albert-Seifried, C. E. Finlayson, M. Koepf, H. J. Kitto, B. Ulgut, M. B. J. Otten, J. J. L. M. Cornelissen, R. J. M. Nolte, A. E. Rowan and R. H. Friend, *J. Phys. Chem. B*, 2011, **115**, 1590–1600.
- 24 L.-A. Fendt, I. Bouamaied, S. Thöni, N. Amiot and E. Stulz, *J. Am. Chem. Soc.*, 2007, **129**, 15319–15329.
- 25 A. Brewer, G. Siligardi, C. Neylon and E. Stulz, *Org. Biomol. Chem.*, 2011, **9**, 777–782.
- 26 H. Gradisar and R. Jerala R., *J. Nanobiotechnol.*, 2014, **12**, 4.
- 27 L. D. D'Andrea and L. Regan, *Trends Biochem. Sci.*, 2003, **28**, 655–662.
- 28 E. R. Main, Y. Xiong, M. J. Cocco, L. D'Andrea and L. Regan, *Structure*, 2003, **11**, 497–508.
- 29 T. Kajander, A. L. Cortajarena, E. R. G. Main, S. G. J. Mochrie and L. Regan, *J. Am. Chem. Soc.*, 2005, **127**, 10188–10190.
- 30 T. Kajander, A. L. Cortajarena, S. Mochrie and L. Regan, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2007, **63**, 800–811.
- 31 A. L. Cortajarena, S. G. J. Mochrie and L. Regan, *Protein Sci.*, 2011, **20**, 1042–1047.
- 32 A. L. Cortajarena and L. Regan, *Protein Sci.*, 2011, **20**, 336–340.
- 33 A. L. Cortajarena, T. Kajander, W. Pan, M. J. Cocco and L. Regan, *Protein Eng., Des. Sel.*, 2004, **17**, 399–409.
- 34 A. L. Cortajarena, T. Y. Liu, M. Hochstrasser and L. Regan, *ACS Chem. Biol.*, 2010, **5**, 545–552.
- 35 S. H. Mejias, B. Sot, R. Guantes and A. L. Cortajarena, *Nanoscale*, 2014, **6**, 10982–10988.
- 36 T. Z. Grove, L. Regan and A. L. Cortajarena, *J. R. Soc. Interface*, 2013, **10**, 20130051.
- 37 N. A. Carter and T. Z. Grove, *Biomacromolecules*, 2015, **16**, 706–714.
- 38 D. Paul, H. Miyake, S. Shinoda and H. Tsukube, *Chem.–Eur. J.*, 2006, **12**, 1328–1338.
- 39 J. S. Lindsey, I. C. Schreiman, H. C. Hsu, P. C. Kearney and A. M. Marguerettaz, *J. Org. Chem.*, 1987, **52**, 827–836.
- 40 M. Kasha and E. G. McRae, *J. Chem. Phys.*, 1958, **28**, 721.
- 41 M. Kasha, H. R. Rawls and M. A. El-Bayoumi, *Pure Appl. Chem.*, 1965, **11**, 371–392.
- 42 Y. Kitahama, Y. Kimura and K. J. Takazawa, *Langmuir*, 2006, **22**, 7600.
- 43 F. J. M. Hoeben, M. Wolff, J. Zhang, S. Feyter, P. Leclère, A. P. H. J. Schenning and E. W. Meijer, *J. Am. Chem. Soc.*, 2007, **129**, 9818–9828.
- 44 F. Würthner, T. E. Kaiser and R. Saha-Möller, *Angew. Chem., Int. Ed.*, 2011, **50**, 3376–3410.
- 45 G. D. Sharma, P. A. Angaridis, S. Pipou, G. E. Zervaki, V. Nikolaou, R. Misra and A. G. Coutsolelos, *Org. Electron.*, 2015, **25**, 295–307.
- 46 A. Brewer, G. Siligardi, C. Neylon and E. Stulz, *Org. Biomol. Chem.*, 2011, **9**, 777–782.
- 47 M. Vybornyi, A. L. Nussbaumer, S. M. Langenegger and R. Häner, *Bioconjugate Chem.*, 2014, **25**, 1785–1793.
- 48 S. Matile, N. Berova, K. Nakanishi, S. Novkova, I. Philipova and B. Blagojev, *J. Am. Chem. Soc.*, 1995, **117**, 7021–7022.
- 49 G. Pescitelli, S. Gabriel, Y. Wang, J. Fleischhauer, R. W. Woody and N. Berava, *J. Am. Chem. Soc.*, 2003, **125**, 7613–7628.
- 50 F. H. C. Crick, *Acta Crystallogr.*, 1953, **6**, 689–697.
- 51 S. Seki, A. Saeki, T. Sakurai and D. Sakamaki, *Phys. Chem. Chem. Phys.*, 2014, **16**, 11093–11113.
- 52 A. Saeki, S. Seki, Y. Koizumi and S. Tagawa, *J. Photochem. Photobiol., A*, 2007, **186**, 158–165.
- 53 Z. Gasyana, W. R. Browett and M. J. Stillman, *Inorg. Chem.*, 1985, **24**, 2440–2447.
- 54 T. Osawa, T. Kajitani, D. Hashizume, H. Ohsumi, S. Sasaki, M. Takata, Y. Koizumi, A. Saeki, S. Seki, T. Fukushima and T. Aida, *Angew. Chem., Int. Ed.*, 2012, **51**, 7990–7993.

