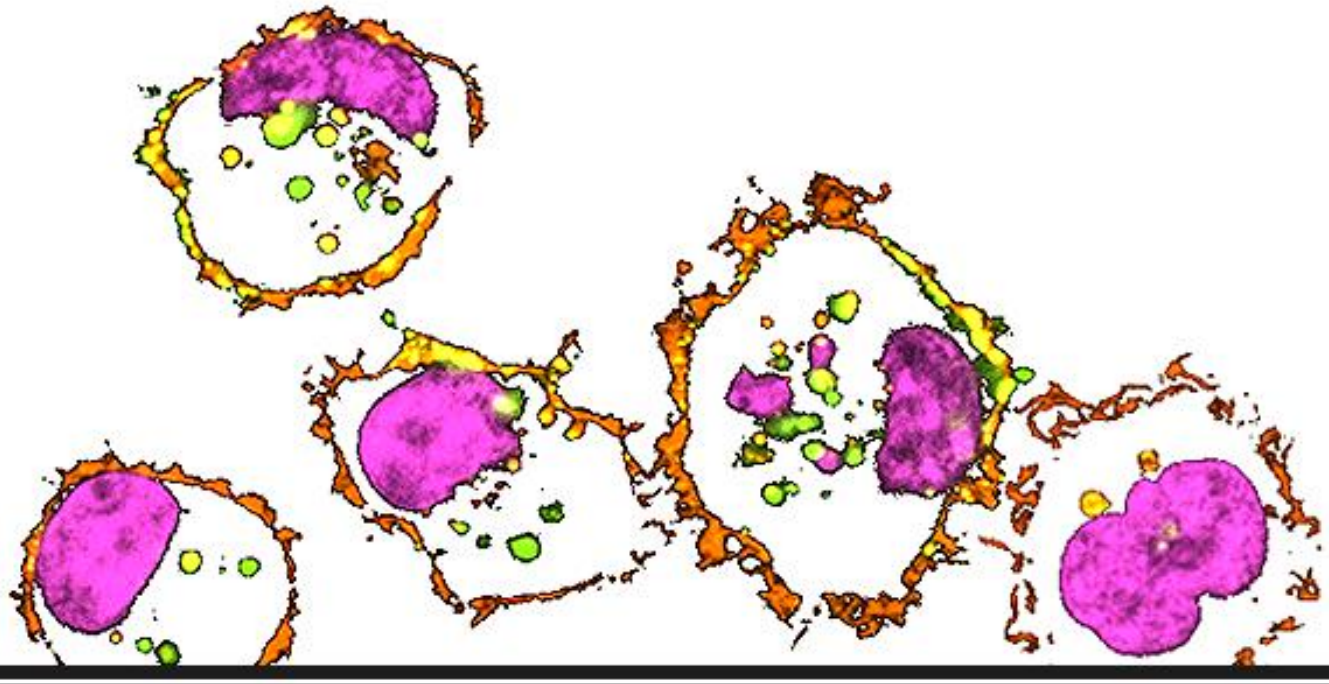
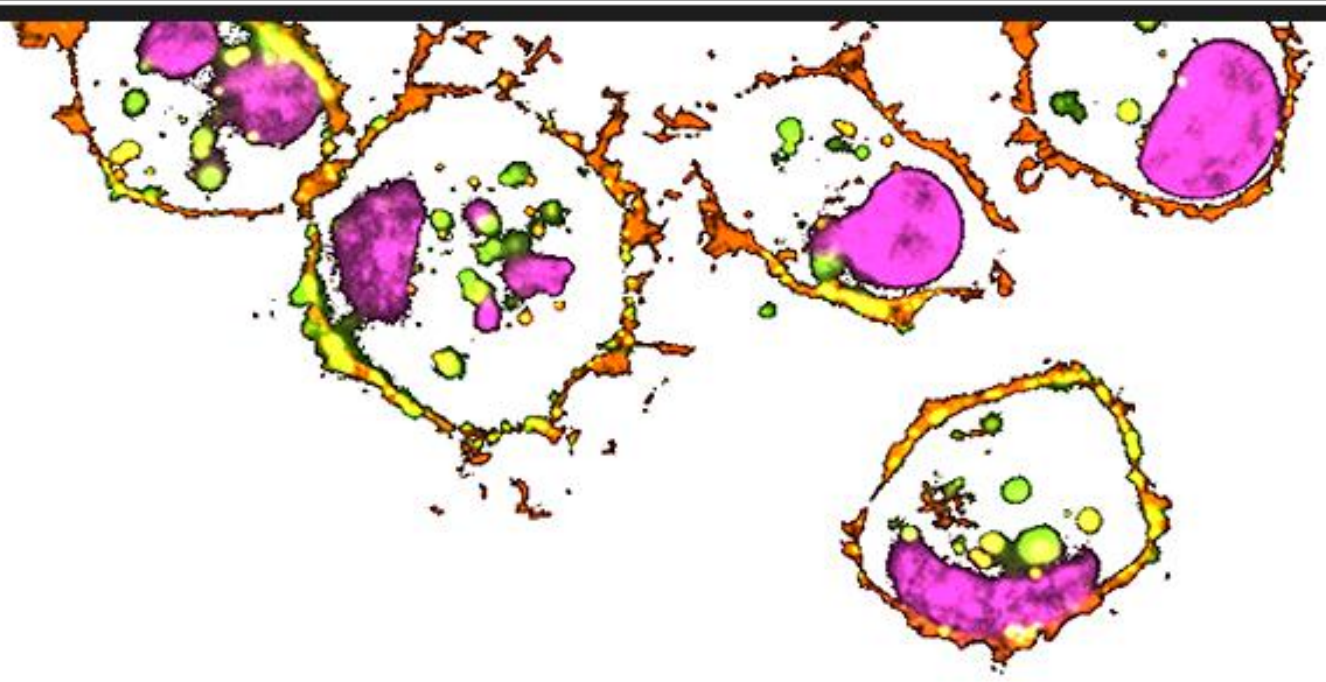
Universitat Autònoma de Barcelona

Supramolecular organisation and biological properties of tumor targeted, self-assembling protein nanoparticles

Mireia Pesarrodona Roches



SUPRAMOLECULAR ORGANISATION AND BIOLOGICAL PROPERTIES OF TUMOR TARGETED, SELF-ASSEMBLING PROTEIN NANOPARTICLES



Mireia Pesarrodona Roches

PhD. Thesis, 2017

Doctorat en Biotecnologia

Supramolecular organisation and biological properties of tumor targeted, self-assembling protein nanoparticles

Tesi doctoral – 2017

Departament de Genètica i de Microbiologia –

Facultat de Biociències

Universitat Autònoma de Barcelona

Institut de Biotecnologia i de Biomedicina

Memòria presentada per Mireia Pesarrodona Roches per optar al grau de doctor en Biotecnologia per la Universitat Autònoma de Barcelona

Mireia Pesarrodona Roches

Vist i plau dels directors de la tesis:

Antonio Villaverde Corrales Esther Vázquez Gómez Neus Ferrer-Miralles

Aquest treball ha estat realitzat principalment a l'Institut de Biotecnologia i de Biomedicina, Vicent Villar I Palasí, sota la direcció dels doctors Antonio Villaverde Corrales, Esther Vázquez Gómez i Neus Ferrer-Miralles. Una part, però, s'ha dut a terme al centre d'investigació Institut für Technische Chemie de la Leibniz University de Hannover, Alemanya.

Learning never exhausts the mind.

Leonardo da Vinci

CONTENTS

INTRODUCTION

NANOMEDICINE

IMPACT OF NANOTECHNOLOGY IN MEDICINE

Nanotechnology is the engineering of functional structures, devices or systems at the nanometer scale, ranging from 1 to 100 nm. Nanotechnologies can be designed to empower specific chemical, physical and biological properties exclusive of their nanoscale proportions and of bulk material [2]. In other words, nanotechnology allows tuning nanoparticle's properties such as size, surface charge and shape, as well as core environment, which are fundamental parameters that determine their biological performance.

Due to its versatility, nanotechnology has fast become a key instrument in many areas such as electronics, food, energy technologies, etc. When nanotechnology is applied in healthcare or medicine it is called **nanomedicine**. Therefore nanomedicine emerges from the combination of nanotechnology with pharmaceutical and biomedical sciences to develop functional nanoparticles for many medical applications. Biomaterials, imaging or *in vitro* diagnostics are some of them. However, **drug delivery** is the application dominating in nanomedicine science representing 76 % of the research activity due to the biophysical properties of nanoparticles (Figure 1). Nanoparticles are customized to adopt specific physicochemical characteristics that determine the biocompatibility and efficacy of the delivered drug. When presented in form of nanoparticles, current therapies have shown enhanced efficacy, reduced toxicity and controlled biodistribution owing to nanoparticles' ability to cross biological barriers, internalise cells and target specific tissues [1,3]. For this reason, some scientists adopt a definition of nanoparticle from the biological point of view where a nanoparticle is an object with a defined structure that possesses these unique medical properties. These functions are not limited to materials up to 100 nm but are then expanded to 1000 nm[1,3].

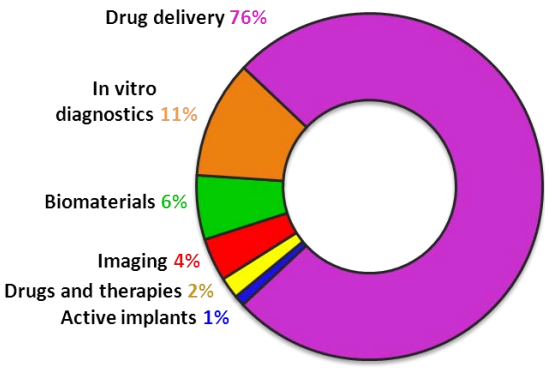


Figure 1: Sectorial dissection of publications in nanomedicine. Reprinted with permission from [1].

Research activity on nanomedicines has streamed over the last decade with a 3-fold increase on nanoparticles for medical purposes in clinical trials [4] and patent activity has exponentially grown since 2000 (Figure 2) [1]. This trend has encouraged governmental funding as well as technology and commercial perspectives worldwide [1].

Nanotechnology opens a plethora of possibilities to design and construct nanoparticles from different materials and with diverse configurations. Although the complexity and the spectrum of **materials** used for **nanoparticle** construction keeps increasingly

expanding, current nanoparticles can be classified into polymeric nanoparticles, micelles, liposomal nanoparticles, inorganic nanoparticles and protein nanoparticles.

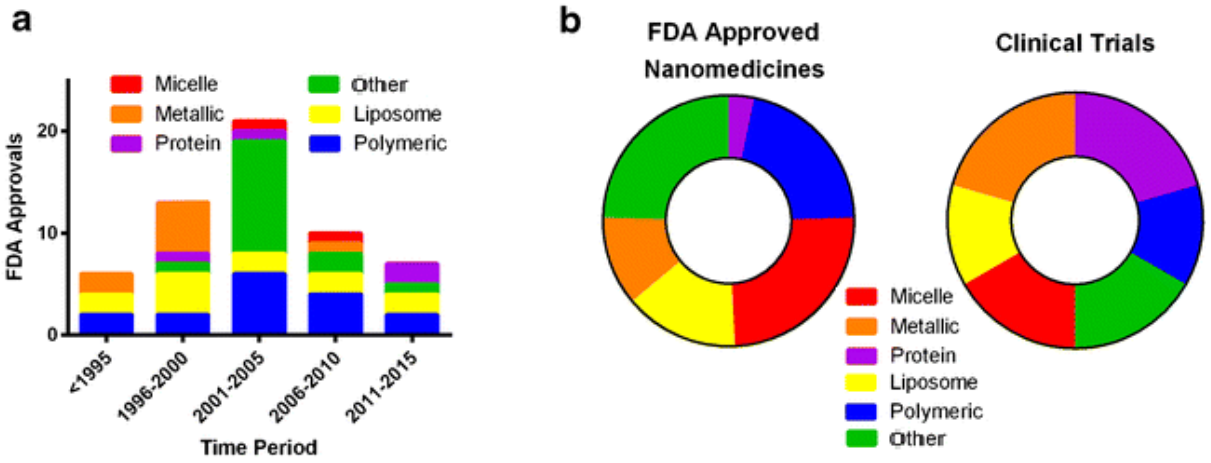


Figure 2: Trends in nanomedicine. (a) Number of FDA approvals in nanomedicines identified by material over 5-year time period. (b) Ring charts represents overall FDA-approved nanomedicines (left) and nanomedicines under clinical trials (right) stratified by material. Reprinted and modified with permission from [4].

Regarding the material used to synthesize the nanoparticles for drug delivery purposes, an important aspect to consider is **biocompatibility**. In addition to the physicochemical characteristics, the bulk material may illicit different reactions concerning immunocompatibility, toxicity and biodegradability. For example, besides showing cytotoxicity issues, carbon-based polymers or polymeric nanoparticles can cause inflammation due to over-stimulation of the immune system. Liposomes and micelles can also evoke immune response and, likewise non-PEGylated micelles [5], dendrimers display serious toxicity problems [6]. Moreover, biodegradability of these materials has been a major concern especially for metallic nanoparticles [7].

Protein-based nanoparticles, being natural biomolecules, represent an appealing alternative for its safety, biocompatibility and biodegradability[8]. Although protein-based particles were initially in the shadow of polymeric and liposomal nanomaterials, over the last 5 years they had blunt among approved nanomedicines and the number is likely to increase given the portion of protein-based nanoparticles under clinical trials [4,9].

Another advantage of protein-based nanoparticles is that they can be designed to display targeting ligands or functionalizing domains in a single chain polypeptide though protein engineering principles. Drugs can be chemically incorporated into the nanoparticles but, when using cytotoxic peptides or therapeutic proteins, these functional agents can be combined into the protein sequence constituting a therapeutic vehicle itself. However, when drugs or functional moieties are incorporated in other type of materials, they must be encapsulated into the nanoparticle or chemically conjugated on its surface being less homogenous in composition than protein-based versions [10].

NANOTECHNOLOGY TOWARDS CANCER TREATMENT

Interest and concern in cancer has gone hand in hand with its increased incidence over population reaching around 14 million new cases of cancer worldwide every year [11]. Although general traits shared among most types of cancer have been described at molecular, cellular or system level, cancer is an extremely complex disease. Among cancer therapies chemotherapy has been widely used despite its low therapeutic index and high toxicity. Its lack of cell specificity leads to the presence of strong side effects by off-target organ accumulation that limits the administrable dose. Also, many chemotherapy drugs are unable to reach metastatic sites or cross biological barriers such as blood-brain-barrier to reach the tumor.

Hence, chemotherapy efficacy would significantly improve if drugs were specifically targeted to cancer cells. This is one of the motors that propel drug delivery nanomedicines, to administrate them in functionalised nanoparticles targeted to the specific tissue thus increasing the administrable dose as well as reducing their side effects. Nanotechnology capability to design vehicles with multiple functions such as optimise biodistribution, internalise specific cells and even specific compartments within the cells turns nanomedicine to a very promising field, especially in an era of personalized cancer medicines [12].

TARGETED NANOPARTICLES

When rationally designing engineered nanoparticles, several targeting steps must be considered: first nanoparticles must be driven to the cancer site, then, internalise cancer cells and finally target the specific sub-cellular compartment [13,14].

Nanomedicines transport to the cancer site can be accomplished by passive or active targeting. So far, most of the approved nanomedicines rely on **passive targeting** via the enhanced permeability and retention (EPR) effect. This effect consists on the enriched accumulation and retention of nanoparticles in tumor tissue when crossing through the gaps between endothelial cells in tumor vessels produced by the high vascularity (angiogenesis) in solid tumors (Figure 3). First generation of nanoparticle-based therapies relies on this targeting mechanism; however, it is not effective in all tumors [15] and neither for relatively large nanoparticles [4,16]. One important additional issue regarding drug access into the cells is that passive targeting does not preclude cell internalisation once within the tumor tissue [15]. However, cell penetrating peptides (CPPs) are used to overcome this problem and facilitate non-specific cell entrance of nanoparticles.

Beyond EPR effect, next generation nanomedicines are designed to employ **active targeting** to access cell cytoplasm [3]. As it has been widely described, cancer cells upregulate certain transmembrane receptors, secreted factors or even expose protein on cell surface that are usually found in the cytoplasm. Active targeting is then based on nanoparticles presenting ligands that bind receptors overexpressed on the target cells

or specific moieties exposed on cancer cell surface (Figure 3). These targeting ligands include: endogenous proteins with known receptor specificity such as albumin or transferrin; peptide sequences from natural resources or screening libraries with described receptor binding; monoclonal antibodies and single-chain variable fragments (ScFv). In general, finding ligands that exclusively recognise a specific target is a difficult task, likewise identifying specific receptors exposed in a particular cell type as they are mostly not exclusive but overexpressed in unhealthy cells compare to normal cells. In this respect, the use of phage display technique or the development of aptamers, which are short nucleic acid sequences that bind cellular targets, ease the discovery of specific binding molecules for a particular target (Figure 6) [13,17,18]. Among the abovementioned ligands, peptides are attractive for the small size, low immunogenicity and low-cost of manufacturing [19,20].

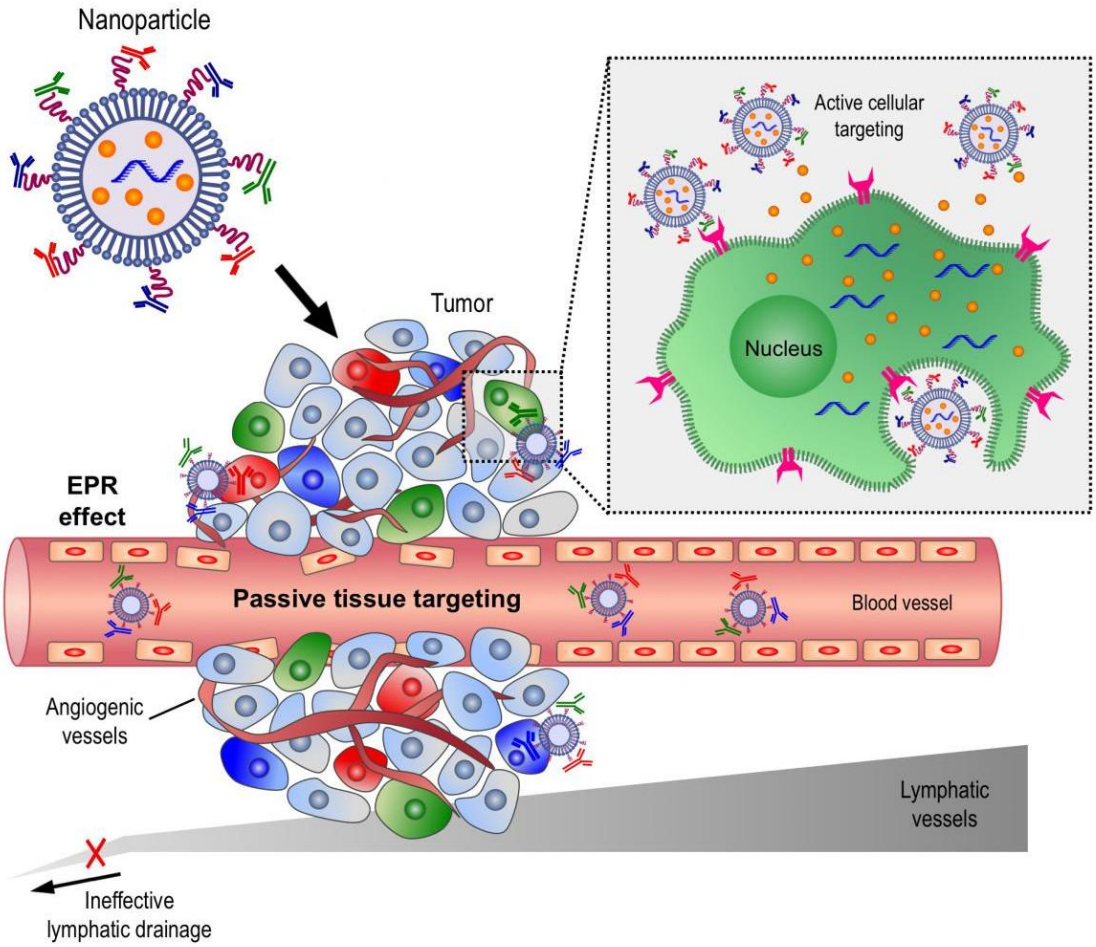


Figure 3: Representation of passive/active tissue targeting. Once nanoparticles reach the tumor tissue through passive targeting assisted by high vascularity and ineffective lymphatic drainage (EPR effect) of the tumor, active targeted nanoparticles can specifically internalise in cancer cells and release the cargo in the required subcellular compartment. Reprinted with permission from [21].

Designing nanoparticles for active targeting results in higher cellular uptake of the drug agent and therefore higher anti-tumor activity compared to non-targeted nanoparticles[3,17].

Nanoparticle's physicochemical characteristics are pivotal to ensure their recirculation and correct biodistribution upon systemic administration. However specific tissue penetrability will mainly depend on the targeting process. The efficacy of active targeting relies, on the one hand, on ligand specificity and affinity for the receptor. On the other hand, selection of the target appears to be crucial to avoid off-target accumulation of the nanoparticles. A receptor which is overexpressed or specifically presented in cancer cells compared to healthy cells is essential to obtain specific and fast nanoparticle accumulation minoring traffic to secondary organs.

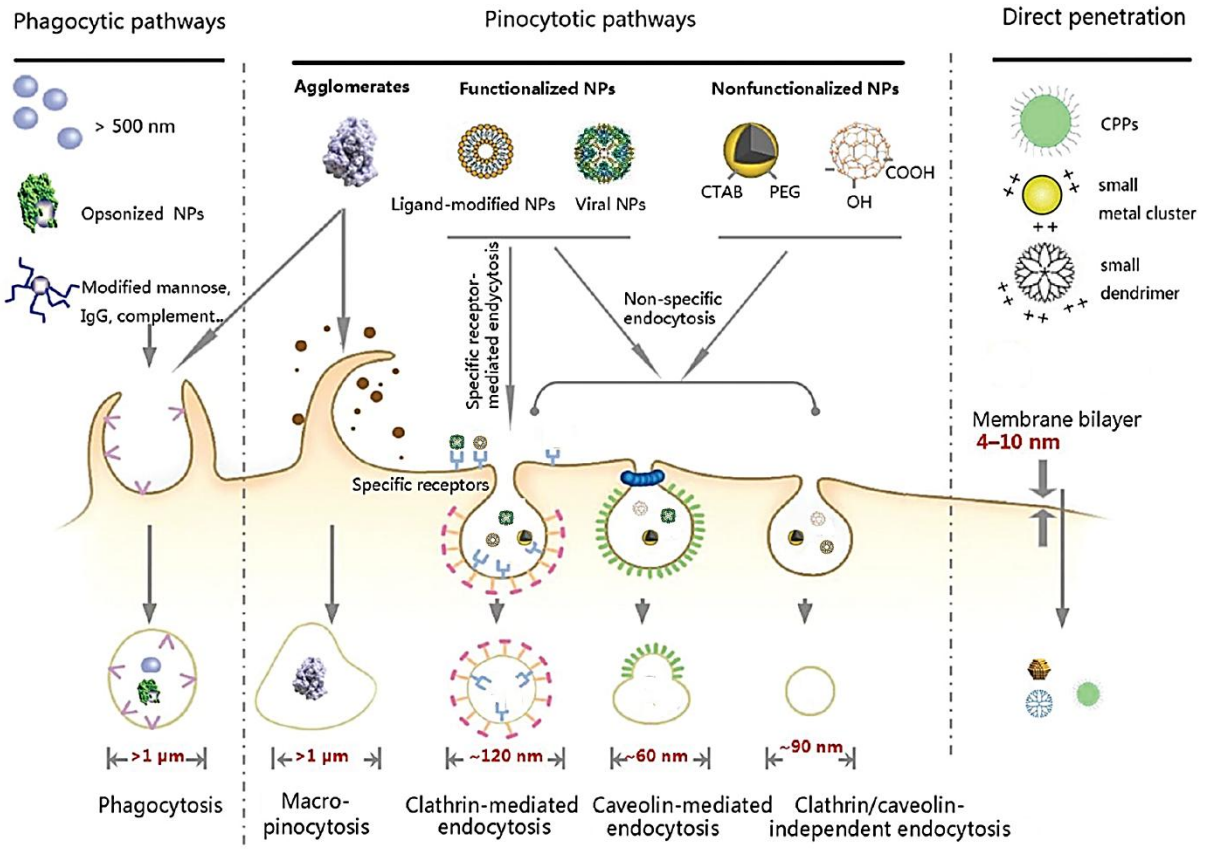


Figure 4: Nanoparticle routes of entry depending on size, material, surface charge and if they are functionalised with targeting ligands or not. Reprinted and modified with permission from [22].

In general, receptor-ligand binding allows nanoparticle cell penetration though receptor-mediated endocytosis (Figure 4). Clathrin-mediated endocytosis, contrariwise other pinocytotic pathways, results in maturation of vesicles into lysosomes leading to enriched enzyme activity and rapid acidification. Therefore, nanoparticles need to be engineered to escape lysosomal degradation. This can be accomplished through proton sponge effect by using cationic surface groups such as His-rich peptides [23] and polyethylenimine (PEI) or by adding peptides able of disrupt endosomal membrane like influenza virus hemagglutinin peptide (HA2) [24], Pseudomonas Exotoxin A (ETA) or Diphteria toxin (DT) [25,26]. Many other peptides from different origins had been described as specific sequences for sub-cellular compartment targeting. By functionalising nanoparticles with these peptides, cargo delivery to nuclei, mitochondria or endoplasmic reticulum can be directed [27,28].

A large set of peptides have been described as ligands binding specific cell receptors or as sub-cellular compartment signalling sequences that can be incorporated for nanoparticles' functionalisation. Loading drugs into active-targeted nanoparticles accounts for the release of drug in a specific location inside the body following systemic administration thus reducing side effect and increasing the administration dose without toxicity.

BREAST AND COLORECTAL CANCER TARGETING: CD44 AND CXCR4 RECEPTORS

Breast cancer is the most common cancer in women accounting for 25 % of all diagnosed cancers. Although early-stage breast cancer is highly curable through surgery removal, the survival rate for patients with metastatic breast cancer is only 20 %. It is a heterogeneous disease with multiple subtypes and its prognosis depends on the cancer subtype. Triple-negative breast cancer (TNBC) is the most aggressive one and it is characterised by the lack of expression of estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2). For this reason hormonal therapy targeted to these cell receptors is not effective, being chemotherapy the primary option for systemic treatment [29].

Current approved nanomedicines for metastatic breast cancer are based on passive tumor targeting of doxorubicine-loaded liposomes such a Doxil or Myocet (approved in 1995 and 2000 respectively) [3]. For this indication, the only active targeting nanomedicine in the market and in turn the first approved protein-based nanotechnology (2005) is Abraxane, paclitaxel-bound albumin nanoparticles. This successful approach significantly enhances paclitaxel solubility and delivery to the tumor [4]. However, albumin non-specificity is a handicap since many albumin-binding receptors have been identified in various cell lines [30].

Colorectal cancer is the third most common cancer worldwide but also the third that accounts for higher mortality rates. To date, no nanomedicines have been approved to treat this disease and standard chemotherapy is the most widely used first approach. Alternatively, when chemotherapeutic drugs are not effective, targeted therapies based on small molecule inhibitors or monoclonal antibodies are used such as Bevacizumab (Avastin) [31] or Cetuximab (Erbitux) [32] which target proteins overexpressed in cancer cells. Yet, just as chemotherapy, they present severe side effects and often fail to cure patients with advanced stages of the illness. In colorectal cancer, mortality is highly related to the appearance of metastasis; likewise TNBC, current therapies are not improving prognosis of patients due to the low penetrability into the tumor.

Metastasis appearance and primary tumor renewal have been highly related with **cancer stem cells** (CSCs), a small subset of cancer cells with self-renewal and multi-lineage differentiation capacity. Invasiveness and mobility of CSCs are relevant characteristics for driving tumour recurrence, progression and metastasis formation [33], besides their

apoptosis and drug resistance, responsible of the low prognosis of many cancers [34]. Despite CSCs molecular heterogeneity, some cell receptors have been identified to be overexpressed in these cells which are crucial for CSCs tumorigenic and metastatic potential [35,36].

In this context, nanoparticle design to specifically target those receptors is a promising tool to treat metastasis and deliver high drug-doses in CSCs to inhibit tumor progression reducing non-desired off-target toxic effects. However, the success of this innovative approach relies on the appropriate selection of those molecular targets and of the effective ligands to bind them and trigger nanoparticle internalisation.

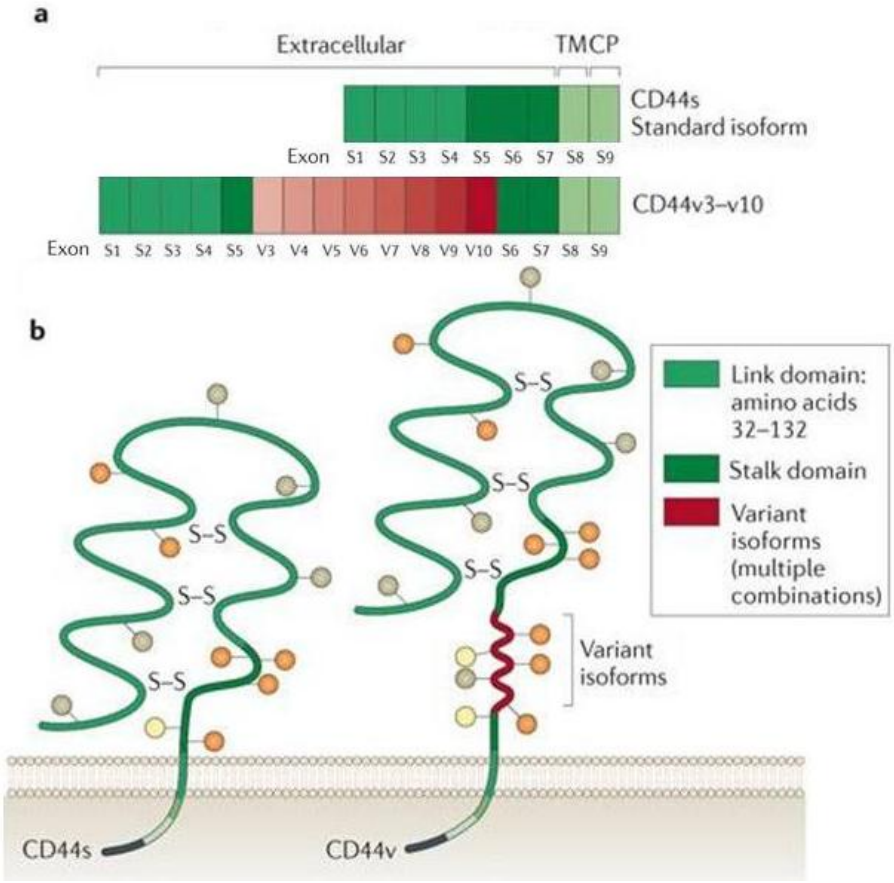


Figure 5: CD44 structure. (a) CD44 gene representation. All CD44 forms display the exons encoding for cytoplasmic tail (CP), the transmembrane region (TM), the stalk domain (dark green boxes) and the link domain (light green boxes) which contains the binding site for (HA). Alternative splicing of exons encoding for the variant regions (red boxes) results in multiple combinations representing the variant isoforms of CD44. Isoform CD44v3-v10 is depicted to compare with standard CD44 isoform. (b) Representation on standard (left) and variant isoforms (right) of CD44 receptor mapping the sites for N-glycosylation (brown circles), O-glycosylation (orange circles) and glycosaminoglycan (GAG)-binding sites (yellow circles). Reprinted with permission from [34].

Several markers shared by CSCs have been identified such as CD20, CD24, CD34, CD44, CXCR4, EpCAM or CD133. Among them, **CXCR4** is the one of choice in colorectal cancer scenario for being specifically involved in metastatic processes and being associated to bad prognosis [37]. Recently in our group, Unzueta *et al*. generated a protein-based nanoparticle empowering T22 domain (an 18-mer CXCR4 ligand derivative of

Polyphemusin II from the horseshoe crab) able to recognise primary tumor and metastatic foci upon systemic administration in an *in vivo* colorectal cancer model [38].

In breast cancer stem cells, **CD44** is the most frequent molecular marker and it is widely used as target of choice for many nanomedicines. It is a transmembrane glycoprotein which participates in cell-cell and cell-extracellular matrix interactions and plays an imperative role in promoting angiogenesis, invasion and metastasis. Although CD44 is encoded by a single gene, alternative splicing results in expression of multiple isoforms (Figure 5). Among them, variants CD44v3 and CD44v6, are detected in most metastasis while no expression is found in normal tissue suggesting their role in metastasis appearance [39].

Given the CD44 receptor expression profile in tumor, many nanoparticles have been designed to actively target CD44-positive cells. Most of them apply CD44 natural ligand, hyaluronic acid (HA), as receptor-binding moiety [40] however results so far have not been very promising. Most likely, this is because not all CD44-positive cells bind constitutively to HA, but seem to be dependent on post-translational modification patterns [36]. Other CD44-targeted nanoparticles have been chemically functionalised by adding specific antibodies recognising the receptor [41,42]. However, antibodies have high molecular weight that limit nanoparticles' multivalency and penetration; apart their manufacturing costs are expensive.

Therefore, CD44-binding proteins are put forwards as an alternative for active-targeting nanomedicines. The advantage of using protein ligands in protein-based nanoparticles is that they can serve as functionalising agents by simply fusing them through protein engineering principles. Several proteins from extracellular matrix including collagen, fibronectin or laminin have been described to bind CD44 receptor [36]. Peptide domains from these proteins have been identified to bind and, in some cases, induce internalisation through CD44 (Table 1). In addition, other non-natural CD44-ligands are discovered through phage display technology [43,44].

Table 1: Sequence, number of positively charged amino acids and specificity of CD44-ligands identified in the literature.

ORIGIN	LIGAND	SEQUENCE	+ AA	SPECIFICITY	REF
Fibronectin	FNI	KNNQKSEPLIGRKKT	5	CD44v3, CD44v6	45,46
	FNII	YEKPGSPPREVVPRPRPGV	4		
	FNV	WQPPRARI	2		
Laminin α5 chain	A5G27	RLVSYNGIIFFLK	2	CD44v3	47,48
Phage display	P7	FNLPLPSRPLLR	2	CD44, CD133	49
Thrombospondin-4	C21	NDTIPEDFQEFQTQNFDRFDN	1	CD44, ROD1, TSP-1	50,51
Fibrinogen	$(\beta 15\text{-}66)_2$	RGHRPLDKKREEAPSLRPAPPPISGGG YRARPAKAAATQKKVERKAPDAGGC	13	All variants CD44 and VE-cadherin	52

As previously mentioned, efficacy of targeted-nanoparticles relies on the appropriate selection of the functionalising ligand. Features such as ligand specificity, receptor-mediated internalisation and binding affinity are fundamental to ensure specific targeting and high drug internalisation for sub-cellular delivery.

PHYSICOCHEMICAL CHARACTERISTICS AFFECTING NANO-BIO INTERACTIONS

Biodistribution, tissue-specific delivery and internalisation efficiency constitute the key biological properties that govern cell-targeted drug delivery in nanomedicine. Apart from the targeting-ligand binding used when rationally designed nanoparticles, physicochemical properties of the whole conjugate like size, shape or surface charge directly influence these biological properties (Figure 6) [3]. Thus, full physicochemical and biological characterisation is not only recommended but it also represents a key requirement for pharmaceutical regulators' approval.

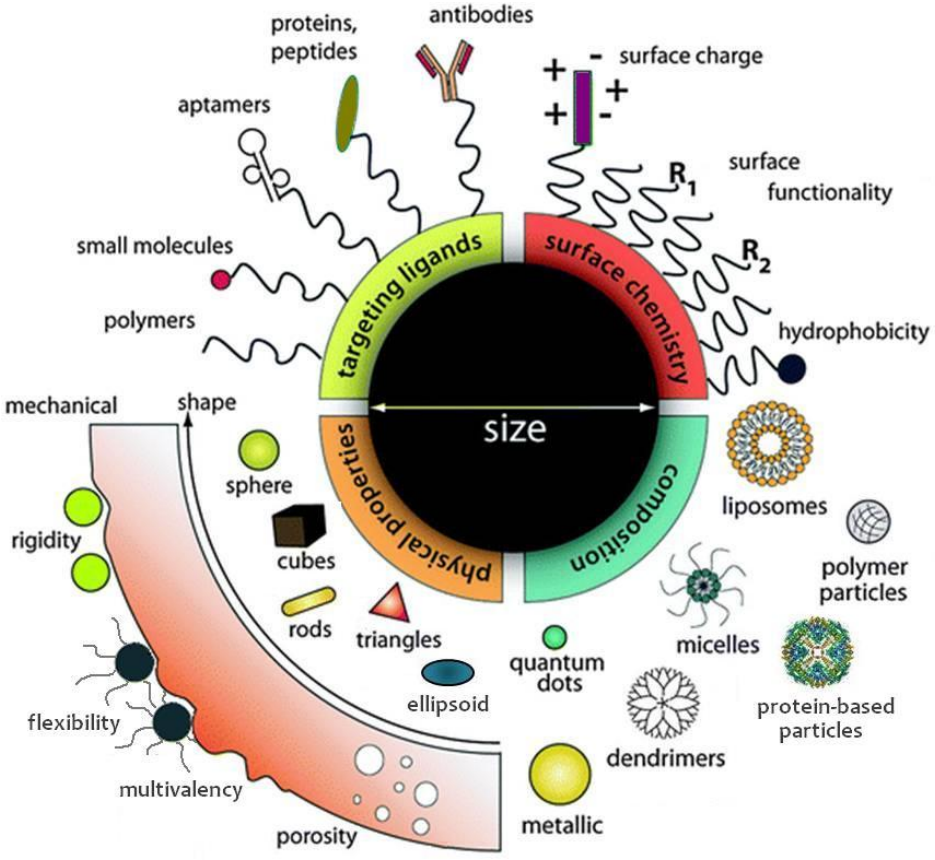


Figure 6: Schematic representation of physicochemical characteristics of nanoparticles that affect its interaction with biological systems. Reprinted with permission from [13].

SIZE is the most investigated and probably one of the most important physicochemical aspects of nanoparticles' affecting pharmacokinetics and biodistribution. Blood circulating time is decreased by non-targeted organ accumulation of nanoparticles strongly influenced by size. For example, small nanoparticles of less than 10 nm are quickly excreted through the kidneys whereas much larger nanoparticles (200 nm) tend

to accumulate in lungs, liver and spleen (where they are processed by phagocytic cells)[53] (Figure 7A) or they are unable to extravasate beyond the blood vessels. Therefore, optimal nanoparticle size to achieve longer circulating time and higher penetration into the tumor is situated between 10 and 100 nm [54,55].

At the cellular level, size can determine the cellular internalisation pathway and therefore the entrance kinetics (Figure 4)[56-58]. Some studies revealed that 15 nm nanoparticles show a higher endocytotic rate than 100 nm nanoparticles [59] although in other studies, 50 nm NPs displayed the maximum cellular uptake [60]. Large data has been gathered that offers contradictory results about size-dependent internalisation. These discrepancies support the fact that ideal internalisation size may vary between different nanoparticle material [54] and cell type[61], but also, between passive and active-targeting[62].

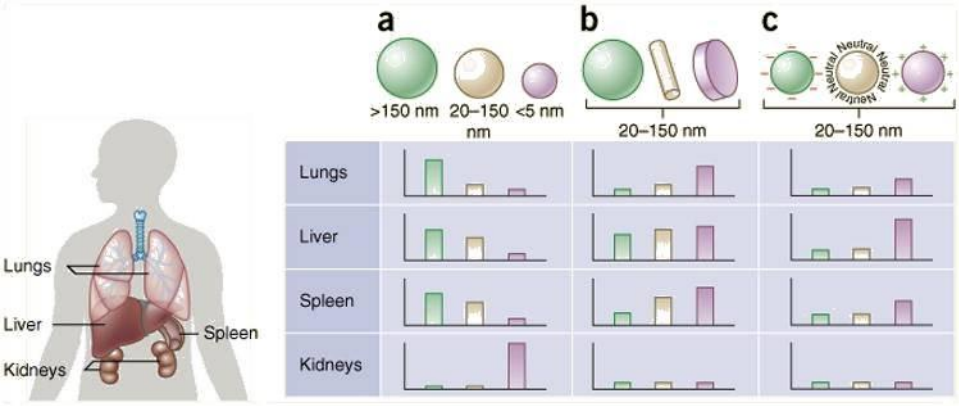


Figure 7: Non-targeted organ accumulation of nanoparticles depending on their size (a), shape (b) or surface charge (c). Reprinted with permission from [63].

One important aspect related to size is **MULTIVALENCY**, being the number of ligands exposed on nanoparticles' surface once they are physically or chemically functionalised. The possibility of multiple functionalisation is enhanced in nanoparticles characterised by a large surface/volume ratio. For protein-based nanoparticles, this aspect is termed nanoparticle's **OLIGOMERIC STATUS**. It is the number of protein building-blocks forming the nanoparticle, each presenting a ligand which has been fused on the polypeptide sequence and thus, it is homogeneously exposed along protein nanoparticles [3]. Particles with higher ligand density lead to an increased ligand-binding affinity although optimal multivalency is the one that allows stable multi-receptor binding to drive membrane bending but cedes receptor molecules for new nanoparticle binding [54,60].

Nanoparticle **SHAPE** is another characteristic that influence cell penetrability even though it is less investigated and, likewise size results are contradictory. For example, spherical nanoparticles perform higher internalisation since the membrane bending energy is minimal compare to rods, cylinders or disc-shaped nanoparticles [64,65]. However, when considering nanoparticles larger than 100 nm, rods or discoidal particles display higher surface area and thus more multivalent interactions that provide them better internalisation rates [59,66].

Regarding biodistribution, besides affecting particle blood circulation time [67], shape drives different particle accumulation in particular organs. Spherical and rod-shaped nanoparticles are rather to be accumulated in liver while disk-shaped NPs are preferably concentrated in lung, heart and spleen (Figure 7B) [68].

Another important aspect to be considered regarding biodistribution and cell internalisation is the **SURFACE CHARGE** of nanoparticles. Usually, charged-nanoparticles are adsorbed by serum proteins that covered them in a so called "protein corona". This adsorption evokes the opsonisation of this particles and later removal by macrophages in the spleen or Kupffer cells inside the liver [54], hence reducing their blood circulation time [55,65]. Liver accumulation has also been detected for hydrophobic nanoparticles (Figure 7C) [20,69]. Depending on surface charge, different levels of toxicity are observed; actually, positive nanoparticles can cause complications such as haemolysis and platelet aggregation [70].

In addition to organ accumulation, cell penetrability pathway and efficacy is altered by surface charge [71,72]. In general, positively charged particles are efficiently adsorbed as a result of cell membrane's negative surface (caused by exposed glycosaminoglycans) but instead they can trigger more non-specific internalisation compared to negatively charged or neutral nanoparticles [55,65].

Many efforts are being made to understand how the physiochemical properties of nanoparticles rule their biological performance and their mechanisms of interaction, but so far, the main conclusion obtained from all these variable results is that no single factor governs the interaction with biological systems but instead they are connected one to another. Apart from pre-clinical studies to determine toxicity and efficacy of a nanoparticle, it is of particular importance to have a comprehensive understanding of the physicochemical parameters and reproducibility of a new nanomedicine [4]. **Taking this consideration in mind and given the influence of the nanoparticle's material on biodistribution and internalisation efficiency, to date, little attention has been paid to the nano-bio interactions of protein-based nanoparticles** [73].

PROTEIN-ONLY BASED NANOPARTICLES

Proteins account for a vast array of functions in all biological systems and have unrestrictedly evolved to fine-tune the precise structure that effectively develop a unique or multiple functions [74]. Their ability to oligomerise is advantageous over monomeric forms possessing higher complexity and stability, allosteric regulation and major functional control. Viruses are an excellent example of multifunctional protein entities with the potential to deliver genetic material, specifically internalise target-cells and efficiently target a sub-cellular compartment; all in one.

The cumulative understanding of the structure-function relationship of proteins has encouraged the design of novel protein assemblies through "bottom-up" approach, in which proteins as building-blocks self-assemble to form higher order structures [75]. Thus, protein assemblies are presented as potential alternative nanoparticles for medical applications.

The defined geometries and stability of protein assemblies is based on the multiple weak interactions such as electrostatic interactions, Van der Waals forces and hydrophobic bonding between building-blocks that result on large protein-protein interactions [76,77]. Owing to their nature, proteins as building-blocks for self-assembled systems have many advantages over inorganic nanoparticles: **ordered and stable structure** at the nano-range; **less toxicity and more biodegradability**; and being synthesised by **easy and cost-effective biological production** [78,79]. Apart from some peptides which are chemically produced, proteins forming nanoparticles are manufactured in safe and environmentally friendly expression systems by recombinant production. Through genetic engineering, the cell factory can produce the recombinant protein that will be extracted and purified thereafter.

In addition, the biggest benefit of proteins is that it is a **versatile material**, allowing incorporation of multiple functionalised moieties at uniform and precise loci by simple genetic modification. These moieties, including specific ligands, DNA-binding domain or intracellular delivery domain can be combined to reach the desired functional activity for efficient therapeutic purposes. Moreover, because many therapeutic molecules present a protein nature, they can be incorporated to protein building-blocks constituting a therapeutic vehicle itself. This has been proved to be one of the most effective drug loading mechanisms [80].

TYPES OF PROTEIN-BASED NANOPARTICLES

Currently, many different protein-based nanoparticles have been applied for drug delivery spanning from natural occurring self-assembling protein nanoparticles to *de novo* design protein nanoparticles (Figure 8).

Among **natural structure-based nanoparticles**, virus-like particles (VLPs) lead the list of protein-based nanomedicines in clinical trials [79,81]. Formed solely by virus-capsid proteins and therefore being non-pathogenic, they are highly stable systems presenting virus-like capabilities such as natural cell targeting and genetic material storing [82] which can be reversibly disassembled *in vivo* [83]. Besides ranging from tens to hundreds nanometres size, VLPs combine a diverse collection of shapes from rods to spheres [82].

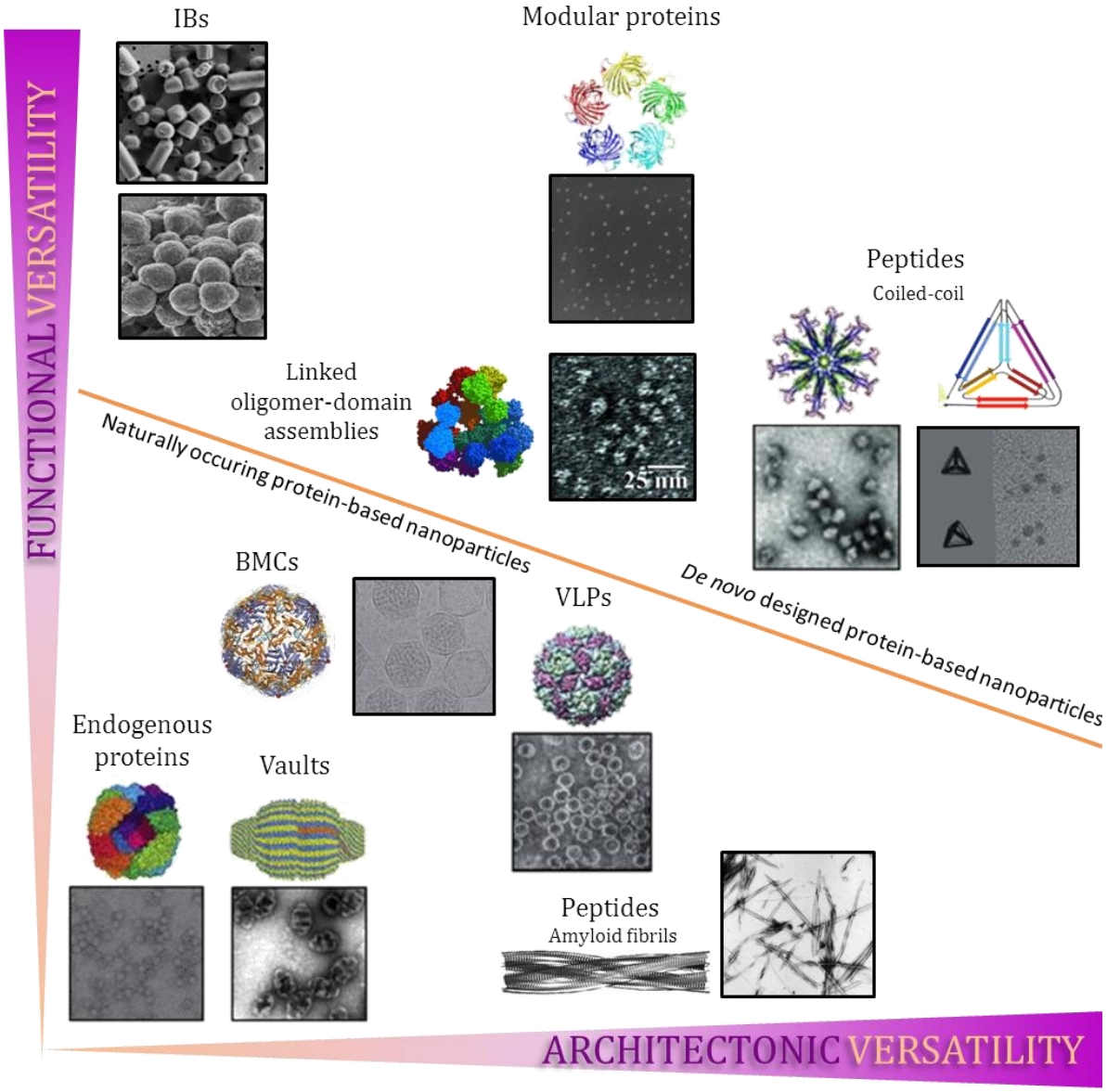


Figure 8: Matrix of most representative protein-based nanoparticles from natural occurring to *de novo* designed particles. Each family of protein-based nanoparticle is represented with an *in silico* structure and a transmission electron microscopy image and is located according to the functional (Y axis) and architectonic (X axis) versatility. All images are reprinted with permission from [38,79,84-90].

Other natural occurring assemblies from bacterial or eukaryotic origin are bacterial microcompartments (BMCs) or vaults respectively. Unlike VLPs, vaults are non-immunogenic structures but display limited architectonic versatility [79]. Endogenous self-assembling serum proteins like ferritin [91] or albumin [92] have been successfully used to enhance drug solubilisation and circulation time [4] and until now they constitute the only approved protein-based nanomedicines.

However, the narrow functional flexibility and limited controlled-geometry from the above mentioned nanoparticles have encouraged findings on *de novo* designed protein assemblies being more tuneable and functionally versatile [93].

New design protein assemblies are distinguished between peptides or full proteins. Secondary structure modules constitute peptides that spontaneously self-assemble in higher order architectures. Beta-strands can interact forming amyloid fibrils and gel-like structures [88,94-96]. In the case of α-helices, their interaction results in adaptable coiled-coil structures forming fiber-like [97], polyhedral [85,98], barrel-like [99] and ring-shape structures [100]. Unfortunately, although self-assembling peptides offer wide architectonic array, they display moderate functional versatility and so far, only few studies have investigated their role in nanomedicine.

Concerning full proteins, defined architectonic geometry and activity, although being complex, can be obtained through rational design of the engineered polypeptide. There is limited number of engineered approaches being adopted in research into nanoparticles. One approach is fusing naturally oligomeric protein domains into a single molecule to obtain cage-like particles [84,101,102]. Other methods are based on elastin-like nanoparticles [103,104] or in the *in silico* engineering *de novo* protein-protein interactions [77]. Recently, a new engineering approach has been developed in our research group which enables to obtain fully functionally versatile nanoparticles by rational design of self-assembling protein building-blocks.

MULTIFUNCTIONAL SELF-ASSEMBLY PROTEIN NANOPARTICLES

Multifunctional self-assembling protein-based systems have been considered an interesting approach as an alternative to VLPs because they allow mimicking virus-like functions required for drug delivery in just a single polypeptide chain. Towards fusion of short functional peptides or insertion into a carrier protein using genetic engineering principles, these artificial complexes can undergo multiple essential functions for many medical therapies such as cell recognition and internalisation, endosomal escape, sub-cellular compartment delivery and DNA condensation (Figure 9) [105]. Some examples are the engineered beta-galactosidase with inserted cell-binding and nuclear transport signal domains [106] or engineered polypeptides with specific cell targeting, endosomal escape, DNA binding and bacterial exotoxin fused-domains [107-109] for cancer therapy.

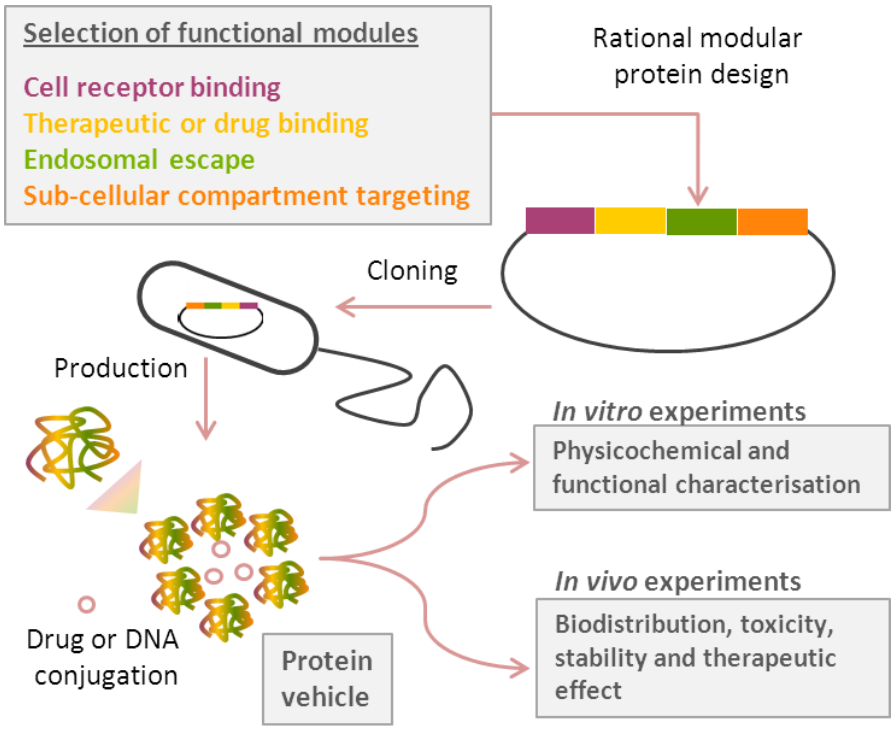


Figure 9: Schematic representation of the production of recombinant multifunctional self-assembling protein nanoparticle.

Compared with nanoparticles constructed with other materials (metals, polymers, liposomes, etc.) which must be functionalised by chemical conjugation to gain these properties, multidomain self-assembling proteins are produced in suitable expression systems being the final product homogeneous in composition [25]. However, the functional modules used in these constructs do not show architectonic potential. Therefore, final geometry and size are uncertain resulting from the tendency of the engineered protein to form supramolecular complexes [93].

Our group has recently described a new protein engineering principle resulting in **multifunctional polypeptides that self-assemble into defined architectonic nanoparticles**. This approach consists of a polypeptide combining an N-terminal cationic peptide and a C-terminal polyhistidine-tag flanking a core protein. These protein building-blocks form highly stable, regular disk-shaped nanoparticles (Figure 10) [110]. The spontaneous organisation is driven by the strong dipolar charge distribution of the polypeptide resulting in weak but multiple protein-protein interactions.

The N-terminal cationic peptide and the His-tag, a part from being enrolled in the architectonic formation, they possess functional activities. N-terminal peptide constitutes a cell-penetration peptide for non-specific (R9) [110] or specific (T22) receptor-mediated internalisation [38] whereas the polyhistidine-tag serves as an endosomal escape peptide [23]. The functional versatility was taken one step further by tethering extra functional peptides such as a DNA binding domain and/or a nuclear localisation signal [111]. These modular protein nanoparticles have shown to be able to

target specific tissue, internalise cells and migrate towards the cytoplasm [110] in absence of toxicity at cellular or systems level [38].

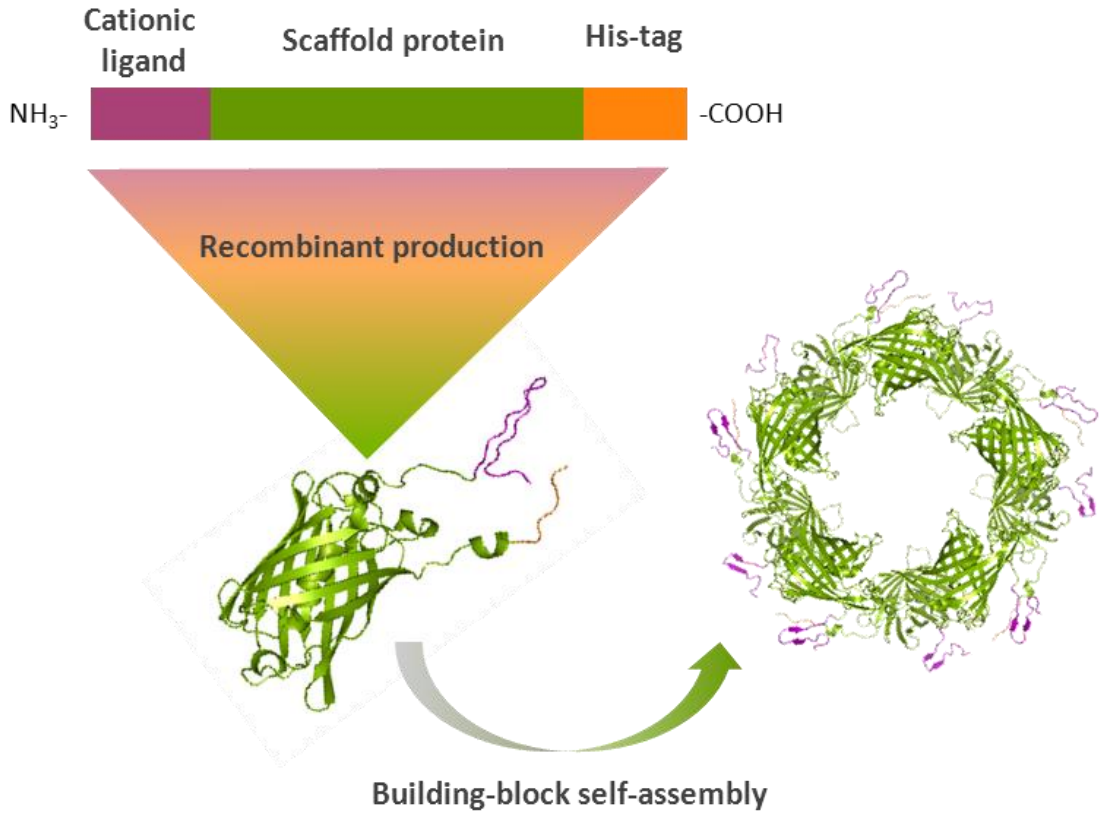


Figure 10: Representation of the developed approach in our group to engineer self-assembling protein nanoparticles. The addition of end-terminal cationic peptides (purple and orange boxes) flanking a core protein (green box) drives the assembly of protein building blocks into higher order structures. *In silico* models taken from the engineered colorectal cancer targeted nanoparticle build up through T22-GFP-H6 protein.

All these findings strength the potential of this engineering approach, based on end-terminal cationic peptides as pleiotropic tags in protein building blocks, as a virus-inspired, more versatile alternative to virus-like particles with enhanced functional adaptability for medical application.

MICROBIAL CELL FACTORIES

The development of tools and techniques in genetic engineering set a significant positive trend in biotechnology allowing genetic material manipulation for recombinant protein production. Since the first insulin produced in bacteria in 1980s, about 400 recombinant protein pharmaceuticals had reached the market and thousands of other protein drugs are under development indicating the relevance and flexibility of this technology [112].

Although many prokaryote and eukaryote protein expression systems are available, the recombinant cell factory per excellence is *Escherichia coli* (Table 2) for the production not only of many exogenous proteins [112] but also of most *de novo* self-assembling proteins [77,113,114].

Table 2: Advantages and disadvantages of *Escherichia coli* expression system. Engineered strains to overcome the limitations of this system [115].

	ADVANTAGES	DISADVANTAGES	Advanced strain
Escherichia coli	Rapid growth and expression	Unable to perform post-translational modifications	Origami B (disulphide bonds)
	Ease genetic manipulation	Presence of endotoxins (LPS)	KPM335, ClearColi™
	Wide knowledge (genetics and tools)	Soluble expression of <15% heterologous proteins	
	Cost-effective culturing medium	Improper folding of >90kDa protein	

The production yield of a heterologous protein can exceed 50 % of total cellular protein [116]. The strong induction of heterologous protein expression leads to an excess of polypeptides in the process of adapting native conformation in a crowded cytosol. To achieve efficient protein folding, the help of cellular folding modulators is required to reach the native conformation in a timely fashion. However, most of times the bacterial folding machinery is inefficient facing high and prolonged stress conditions and complexity of the recombinant protein which leads to build-up protein aggregates known as **Inclusion Bodies (IBs)** [117].

PROTEIN QUALITY CONTROL

Molecular **chaperones** are the ultimately most important folding modulators to assist in *de novo* protein folding or refolding to reach its native conformation. Chaperones, recognising unstructured hydrophobic stretches that should be embedded in the core of the native protein conformation, avoid misfolding or intermolecular binding and subsequent protein aggregation [116]. They are divided into three functional categories:

Folding chaperones (DnaK [118] and GroEL/ES [119]) consist of protein modulators that help nascent peptide to reach the native conformation in an energy-dependent manner. *Holding chaperons* bind to protein intermediates under an overwhelmed situation, while awaiting for folding chaperones to become available to refine protein folding [120].

And finally, *disaggregating chaperones* (ClpB [121]) resolubilise aggregated proteins to reactivate refolding when stress conditions are lessened (Figure 11).

Beyond these three lines of defence upon protein aggregation, a set of cytoplasmic heat shock proteases prevent unfolded polypeptide accumulation as insoluble aggregates. **Proteases** like ClpA [122], ClpP [123] or Lon [124] degrade misfolding-prone proteins, proteolytically vulnerable intermediates or prematurely terminated polypeptides [125].

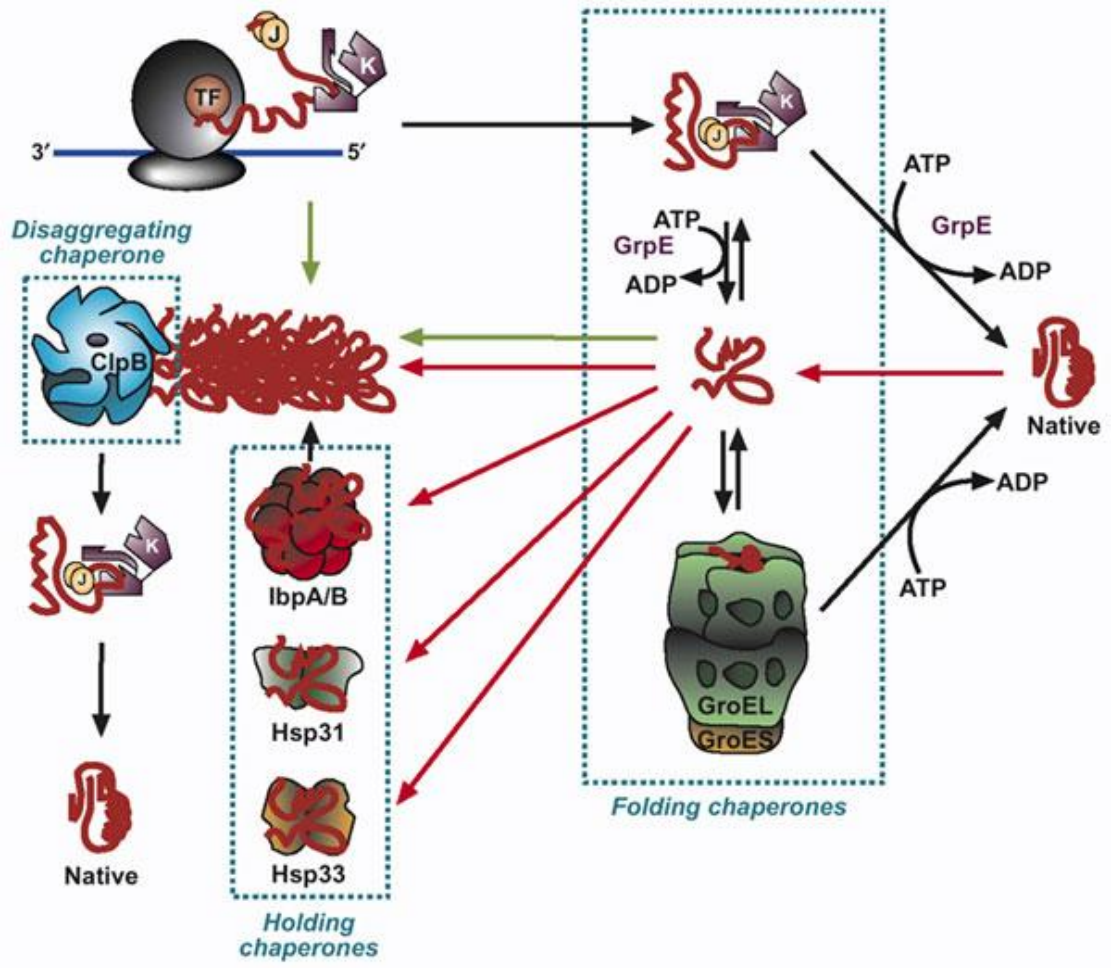


Figure 11: Flowchart representation of protein folding in *Escherichia coli* assisted by the three different functional chaperone categories. Nascent polypeptide is encountered by trigger factor (TF), DnaK (K)/DnaJ (J) chaperones and subsequently transferred to GroEL/ES folding chaperons. Red arrows indicate the unfolding and aggregation of proteins in times of stress. Green arrows indicate protein aggregates that accumulate in inclusion bodies. Reprinted with permission from [116].

Despite the above-mentioned protein folding mechanisms and enhanced expression approaches such as reducing the culture temperature, engineering the protein sequence, adding fusion partners or coproducing chaperones, the production of many heterologous proteins in *E. coli* results in IB formation [126].

IBS AS RECOMBINANT PROTEIN SOURCE

Due to the fact that insoluble aggregates formed by misfolded protein are responsible for IB formation, protein quality has largely been assumed to be intimately linked to protein solubility. IBs have been considered as purely amyloid-like aggregates of unfolded protein protected from proteases degradation. Nevertheless, some approaches have revealed that **IBs contain variable amounts of properly folded protein embedded in the aggregate** [127,128]. Eventually, IBs are composed by a wide spectrum of protein conformations and functional qualities, as well as it happens with soluble protein combining intermediate states, native-like structures or even soluble aggregates [129].

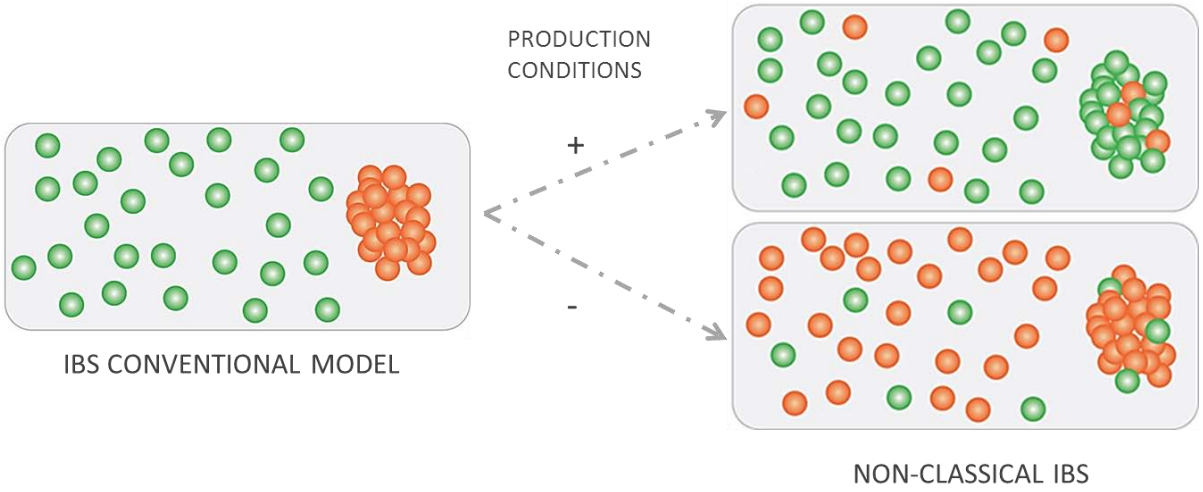


Figure 12: Old fashion structural conception of IBs (left) in which these entities were solely formed by amyloid structures of inactive unfolded protein (orange spheres) and soluble protein was exclusively composed by functional native protein (green spheres). New IBs model (right) in which aggregates are described as a build-up of unfolded protein together with fold-intermediates and native protein. Structural and functional quality is shared between soluble protein and IBs and it is modulated according to the production conditions. Reprinted with permission from [128].

Recently, Cano-Garrido and co-workers [130] described IBs as sponge-like supramolecular structures where native-like protein species are filled within an amyloid scaffold structure (Figure 13). Thus, the **protein quality** is not reflected by protein solubility but it **depends on intrinsic characteristics of the amino acid chain, cell growth temperature and genetic traits of the host** [117,129,131].

Besides IBs enclose several conformational states, they are composed of up to 95 % of the recombinant protein being the rest of components chaperones involved in IB protein disaggregation (DnaK, ClpB and IbpAB) [132], phospholipids and nucleic acids traces from the host cell [133].

Considering that IBs are composed of functionally native or native-like protein and constitute an almost pure source of recombinant protein, a new biotechnological approach adopted is to solubilise these aggregates to obtain active protein and thus reduce the need of extensive chromatographic purification steps and the associated costs [131].

Protein recovery from IBs requires IB isolation, solubilisation, in some cases refolding, one-step chromatography and quality control assessment. Many solubilisation agents have been applied for IB disaggregation. The use of high concentration of chaotropes like urea or guanidine hydrochloride (GuHCl) completely unfolds protein aggregates [134]. However, *in vitro* refolding procedure are tedious and depend on extensive trial-and-error and case-by-case strategies to find the appropriate physical and chemical conditions to obtain the native conformation and avoid protein aggregation (Figure 13) [135].

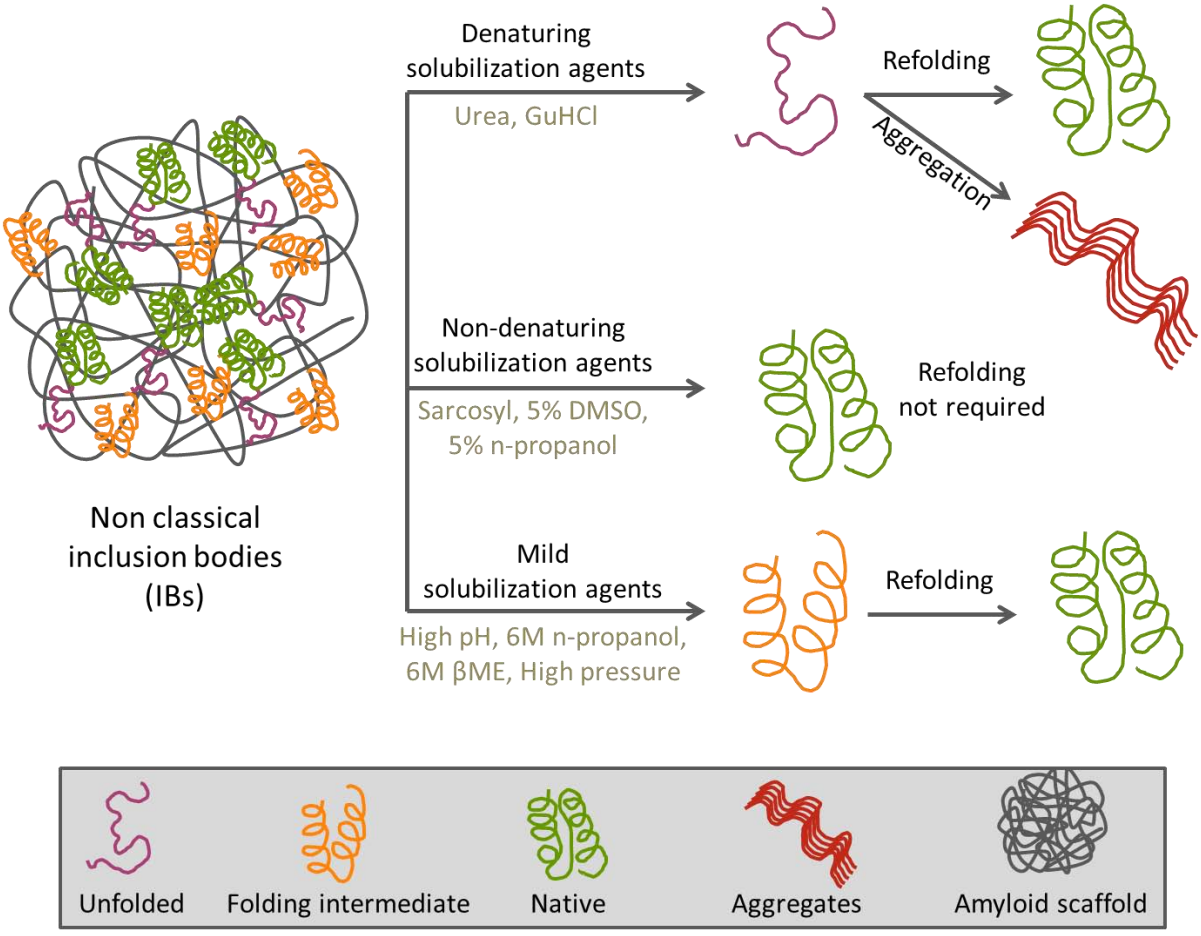


Figure 13: Schematic representation of the solubilizing methods to obtain soluble active protein from inclusion bodies. Modified with permission from [134].

Non-denaturing solubilisation agents such as Sarkosyl [86] or DMSO [136] are been successfully used to extract biologically active protein from IBs. These agents are intended to extract properly folded or native-like proteins embedded in the IBs avoiding refolding procedures (Figure 13). By using this strategy, Peternel *et al.* registered the same protein activity efficiency compare to the soluble counterpart for GFP-forming IBs[86].

Albeit many self-assembling protein nanoparticles have been purified from IB solubilisation [107,137], **how the protein material source** (soluble protein or IBs) **affects** the **biological performance of smart protein nanoparticles** remains neglected.

ENDOTOXIN FREE SYSTEMS

All Gram-negative bacteria, including *E. coli*, present LPS (lipopolysaccharide) on the outer membrane which is a highly immunogenic compound in *in vivo* tests [138]. Inevitably, recombinant proteins produced in gram-negative cell factories will contain LPS contamination. This is an issue, on the one hand, because **LPS can lead to toxicity** problems for protein pharmaceuticals and, on the other hand, because LPS-driven biological activities have been wrongly attributed to the produced recombinant protein[139,140].

Since LPS removal options are not universally applicable or completely efficient [141], the approach of using **LPS-free *E. coli* mutant strains** represents an appealing alternative [138]. The potential of these mutants as cell factories for the production of soluble proteins without endotoxin activity has been recently demonstrated[138,142]. Thus, they are an urgently needed option to be explored for the production of recombinant protein nanoparticles for biomedical applications.

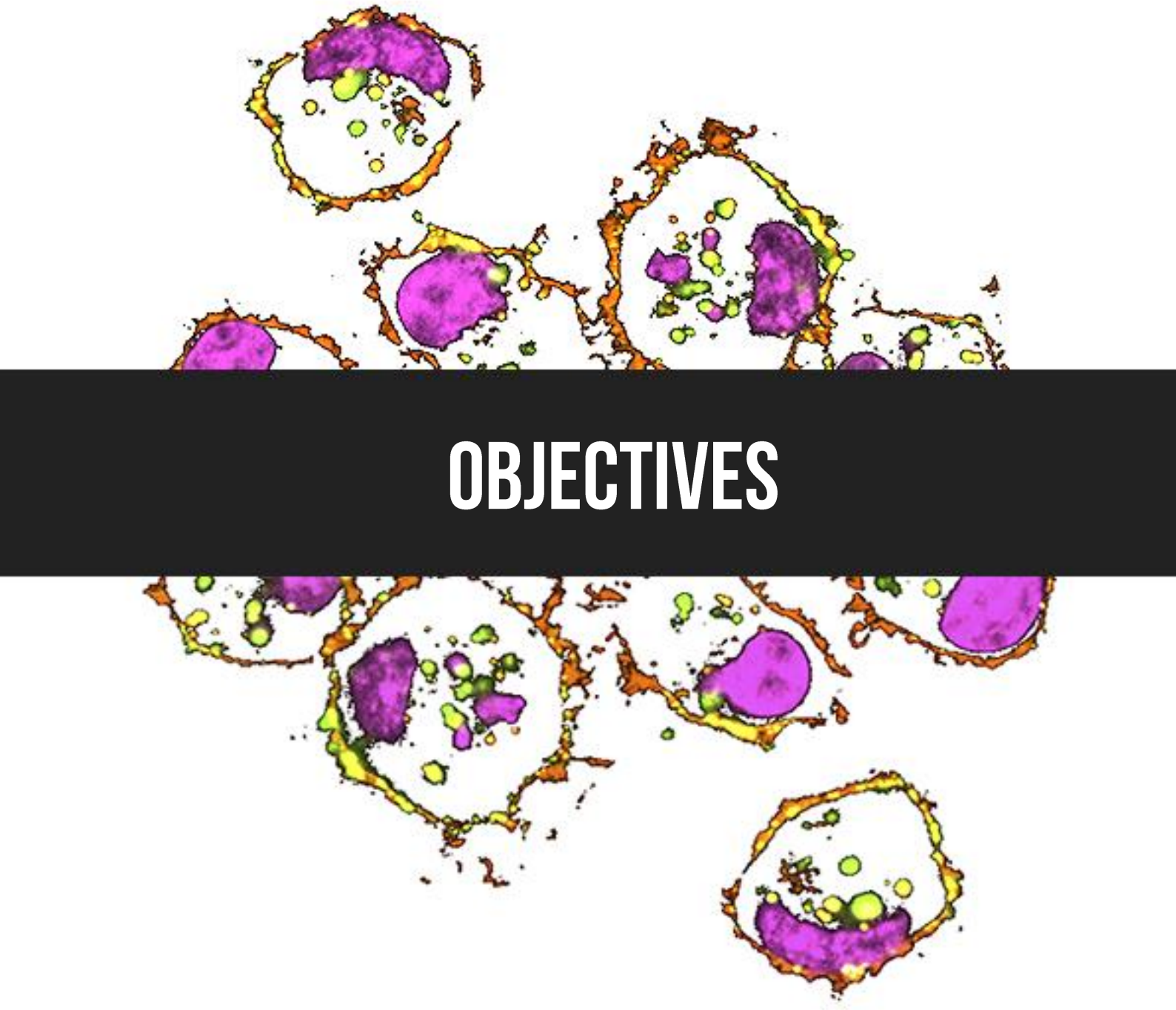
OVERVIEW

Recombinant multidomain protein building-blocks have demonstrated to be an appealing biotechnological approach for the development of cell-targeted, nanoscale drug carriers on account of protein functional and architectonic versatility. The colorectal tumor and metastases targeting efficacy as well as non-toxicity of T22-empowered nanoparticles validates the potential of the *de novo* engineering approach developed in our research group. An approach based on end-terminal cationic peptides as pleiotropic tags in self-assembling protein building blocks [38].

In an attempt to explore the flexibility of this approach, the first section of this thesis' research has been focused on the design, production and characterisation of self-assembling protein nanoparticles targeted to triple negative breast cancer though CD44 receptor. These nanoparticles might be used in a future as nanoscale tumor-targeted drug carriers in the treatment of breast cancer. Given the limited performance of CD44 targeting through HA-presenting nanoparticles [36], the use of a peptide ligand to bind CD44 receptor is appealing and can be adapted in the building-block as a single polypeptide.

Referring to microbial cell factories, many of the recombinant proteins are likely accumulated in IBs upon induction stress; however, considering non-classical structure description of these protein build-ups they represent an abundant and pure protein source [131]. Although active protein has been obtained from IBs through mild solubilisation procedures [86], it is still unknown how the recovery of active protein from IBs affects the self-assembling process into protein nanoparticles. Thus, the second part of this work analyses the impact on nanoparticle structure and functionality considering the protein material source of the developed CD44-targeted protein vehicles.

Albeit it has been recently known that the bacterial host directly influence the architecture and the performance of the produced protein-nanoparticles [69], how the genetic background of *E. coli* strains deficient of chaperons or LPS-free affect these protein assemblies has not been addressed. Thus, the last section of this thesis consists on a fine structural understanding of CXCR4 and CD44 targeted nanoparticles produced in different *E. coli* strains and how their physicochemical characteristics affect the nanoparticles' biointeractions.



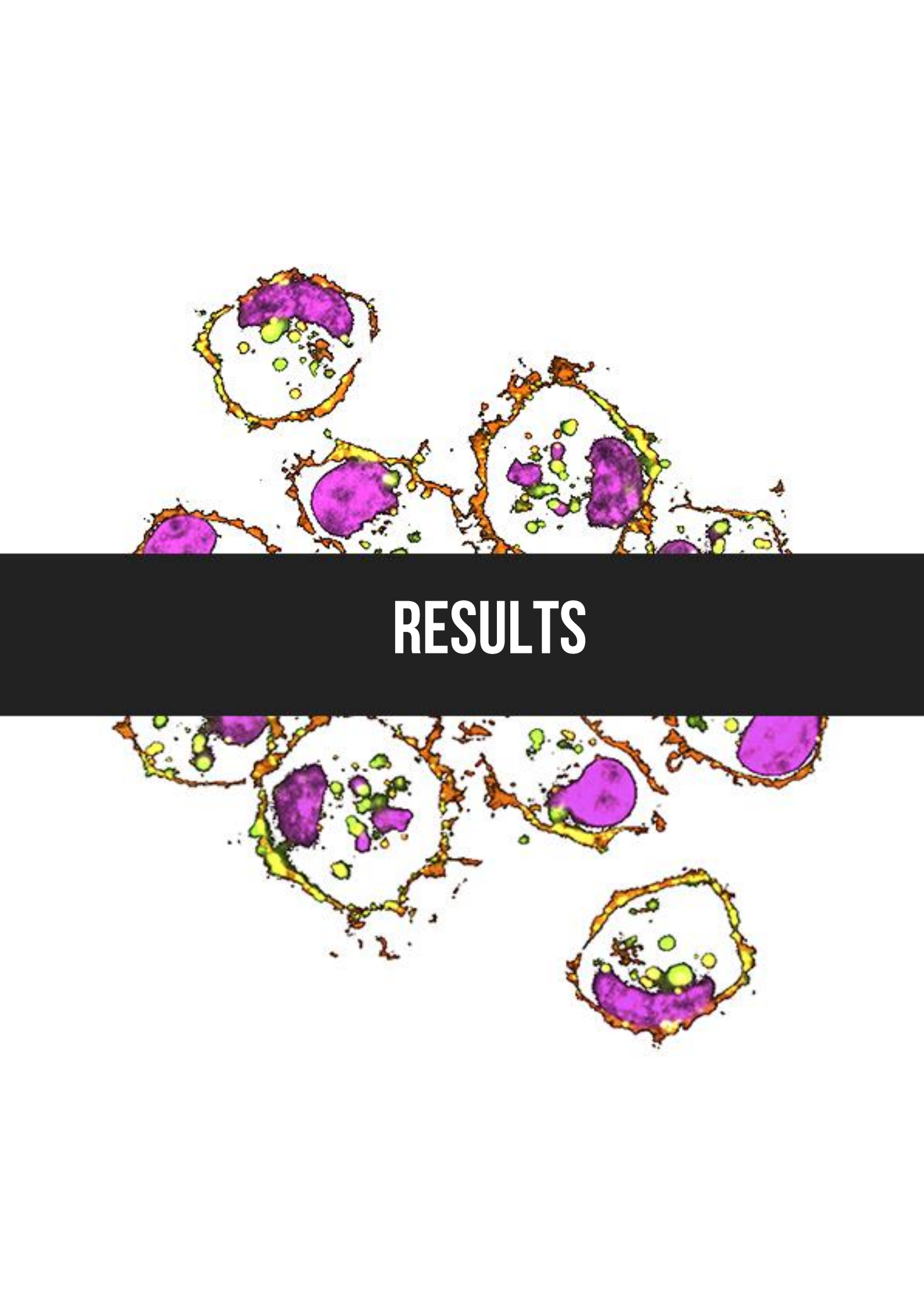
OBJECTIVES

The approach developed in our group based on the application of architectonic peptide tags to induce the self-assembly of multifunctional protein building blocks into nanoparticles has yielded outstanding *in vitro* and *in vivo* results in a colorectal cancer model. The aim of this thesis is to explore the potential of this approach in a breast cancer system and characterise the morphometric and biological properties upon nanoparticle's bio-production. For this intention, we planned the following objectives:

1) To engineer self-assembling protein nanoparticles targeted to breast cancer cells by incorporating cationic peptide ligands binding CD44 receptor. To analyse their assembly capacity and to evaluate their biological performance in terms of specific receptor-mediated cell internalisation.

2) To comparatively analyse the structural and functional performance of the developed CD44-targeted protein vehicles when obtained from soluble cell fraction or from bacterial IBs through non-denaturing solubilisation procedures.

3) To characterise the supramolecular organisation and biological properties of CXCR4 and CD44 targeted protein nanoparticles when produced in several *E. coli* strains with different genetic background.



RESULTS

ARTICLE 1

Intracellular targeting of CD44+ cells with self-assembling, protein only nanoparticles

Pesarrodona M, Ferrer-Miralles N, Unzueta U, Gener P, Tatkiewicz W, Abasolo I, Ratera I, Veciana J, Schwartz S Jr, Villaverde A, Vazquez E

International Journal of Pharmaceutics. 2014 Oct 1; 473(1-2):286-95

CD44 is a multifunctional cell surface receptor that has a role in cell proliferation and differentiation, thus CD44 expression has been connected to the tumorigenic and metastatic potential of CSCs. In a drug delivery context, CD44 has represented an appealing marker for nanoparticles targeting in cancer therapies. Most of the formulated nanoparticles have been designed by the conjugation of the natural ligand of CD44 (hyaluronic acid, HA) however it has been described that not all $CD44^+$ cells constitutively bind HA. Along these lines, proteins binding CD44 represent a potential alternative to HA allowing, through conventional protein engineering, functionalized and adaptable protein-only nanocarriers.

Taking these alternative nanoparticles into consideration and applying the nano-architectonic principle based on the combined use of end terminal cationic peptides and polyhistidine tag, we have explored five cationic CD44-specific protein ligands as components of multifunctional chimerical proteins composed of CD44-ligand and a His tag flanking the GFP protein with potential for self-oligomerisation.

Among tested ligands, two peptides from laminin and fibronectin, respectively, drive the formation of self-assembling, fully biocompatible protein-only nanoparticles of 14nm in form of stable ring-shaped entities. Those protein nanoparticles efficiently bind and internalise target cells upon exposure with demonstrated receptor-driven specificity. Such particulate protein organisation confers added value properties to the constructs favouring cellular penetrability, what opens a plethora of possibilities for the rational design of protein-based, fully biocompatible nanomedicines.

IJP-D-14-00737 REVISED VERSION

Intracellular targeting of CD44$^+$ cells with self-assembling, protein only nanoparticles

Mireia Pesarrodona [1,2,3], Neus Ferrer-Miralles [1,2,3], Ugutz Unzueta [1,2,3], Petra Gener [3, 4], Witold Tatkiewicz [3, 5], Ibane Abasolo [3, 5], Imma Ratera [3, 4], Jaume Veciana [3, 4], Simó Schwartz Jr [3, 5], Antonio Villaverde [1,2,3 *], Esther Vazquez [1,2,3]

[1]Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain

[2]Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain

[3]CIBER en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Zaragoza, Spain

[4] CIBBIM-Nanomedicine, Hospital Universitari Vall d'Hebron and Vall d'Hebron Institut de Recerca, Universitat Autònoma de Barcelona, 08035 Barcelona, Spain.

[5] Department of Molecular Nanoscience and Organic Materials, Institut de Ciència de Materials de Barcelona (CSIC), Bellaterra, 08193 Barcelona, Spain

*** Corresponding author:** A. Villaverde. Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain. Phone: (+34) 935813086; Fax (+34) 935812011; Email antoni.villaverde@uab.cat

Abstract

CD44 is a multifunctional cell surface protein involved in proliferation and differentiation, angiogenesis and signaling. The expression of CD44 is up-regulated in several types of human tumors and particularly in cancer stem cells, representing an appealing target for drug delivery in the treatment of cancer. We have explored here several protein ligands of CD44 for the construction of self-assembling modular proteins designed to bind and internalize target cells. Among five tested ligands, two of them (A5G27 and FNI/II/V) drive the formation of protein-only, ring-shaped nanoparticles of about 14 nm that efficiently bind and penetrate $CD44^+$ cells by an endosomal route. The potential of these newly designed nanoparticles is evaluated regarding the need of biocompatible nanostructured materials for drug delivery in CD44-linked conditions.

Keywords: CD44; Multifunctional protein; Nanoparticle; Self-assembling; Biomaterials; Drug delivery

1. Introduction

CD44 is a transmembrane adhesion glycoprotein which participates in cell-cell and cell-extracellular matrix interactions, being hyaluronic acid (HA), fibronectin and laminin its natural ligands. CD44 is expressed in many cell lines including leukocytes and fibroblasts and it participates in a wide range of physiological processes such as cell migration, lymphocyte homing, cell activation and hematopoiesis (Gee et al. 2004). Among several surface receptors overexpressed in cancer stem cells (CSCs), including CD133, CD44, CD49 and ITGA6, CD44 is the most frequent molecular marker, being present in a large variety of tumor types. It is prevalent in highly recurrent colon, liver, prostate or breast cancers (Zoller, 2011).The tumorigenic and metastatic potential of CSCs have been associated to CD44 expression. Many evidences strongly support that an alteration of CD44 expression levels promotes tumor cell survival and aggressiveness and it also induces tumorogenesis and metastasis. In this regard, cell lines which highly express CD44 are capable of forming more aggressive tumors in the invading tissue (Goodison, Urquidi et al., 1999). In breast cancer, CD44 is not only a useful stem cell marker but also a promising therapeutic target (Marangoni, Lecomte et al., 2009; Sauter, Kloft et al., 2007; Tijink, Buter et al., 2006), and targeting to CD44 reduces tumor growth and prevents post-chemotherapy relapse of human breast cancer xenografts (Marangoni, Lecomte et al., 2009). In humans, anti-CD44 antibodies used as targeting agents for either radiolabels or anticancer chemotherapeutics have shown promise in clinical trials, and disease stabilization was observed in patients with breast or head and neck tumors treated with anti-CD44-based conjugates (Sauter, Kloft et al., 2007; Tijink, Buter et al., 2006).

In the drug delivery scenario, a diversity of materials including natural polymers, carbon nanotubes and lipid-based and inorganic nanoparticles have been proposed for the specific targeting CD44-expressing cells, upon convenient functionalization. Most of them have been formulated by the conjugation of a given nano-vehicle with HA acid

(Peer, Karp et al., 2007), but in general, preliminary results have been not dramatically promising. Apart from toxicity issues (Goodison, Urquidi et al., 1999), not all CD44 + cells constitutively bind HA, as the binding capacity appears to be influenced by structural variation and/or CD44 glycosilation patterns (Gee, Kryworuchko et al., 2004; Zoller, 2011). In this regard, CD44-binding proteins are promising alternatives to HA as targeting agents, and the use of these proteins as functionalizing agents would offer the flexibility of protein engineering in nanoparticle design. However, while CD44 has been described to undergo receptor-mediated endocytosis when bound to HA and collagen (Koo, Huh et al., 2011; Rezler, Khan et al., 2007), evidences supporting internalisation of laminin and fibronectin are still missing.

In the present study and by applying a nano-architectonic principle based on the combined use of end terminal cationic peptides and polyhistidines (Ferrer-Miralles, Corchero et al., 2011; Unzueta, Cespedes et al., 2012; Unzueta, Ferrer-Miralles et al., 2012), we have explored five CD44-specific protein ligands as components of multifunctional chimerical proteins, with potential for oligomerization. Among them, two peptides from laminin and fibronectin, respectively, efficiently target CD44+ cells and in addition, they promote the self-assembling of the carrier protein as functional nanoparticles of 14 nm. Such particulate organization confers added value properties to protein constructs, favouring cellular penetrability and opening a plethora of possibilities for the rational design of protein-based, fully biocompatible nano-medicines, that in form of viral biomimetics, target CD44-overexpressing cells.

2. Materials and methods

2.1. Cell lines and media. MDA-MB-231 cell line was maintained in RPMI 1640 supplemented with 10 % foetal calf serum (FBS) and 6 mM GlutaMAX (Invitrogen), and MCF-7 in Dulbecco's Modified Eagle Medium (DMEM) F12 supplemented with 10 % foetal bovine serum (FBS) and 2 mM GlutaMAX (Invitrogen). HepG2 (ATCC HB-8065) cell line was maintained in MEM-α (Invitrogen) supplemented with 10 % FBS and 2 mMGlutaMAX. All cell types were incubated at 37°C and 5 % CO_2 except HEK-293-T, which was maintained at 10% CO_2 in DMEM supplemented with 10% FBS and 2 mMGlutaMAX.

2.2. Protein design, production, purification, and characterization. Five chimeric genes were designed in-house and provided by GenScript (Hong Kong, China). The resulting fusion proteins (Figure 1 A) were named according to their modular organization as N_{term}-Peptide-GFP- H6-$C_{term,}$ being all N terminal peptides known CD44 ligands (Table 1). Using *Nde*I/*Hind*III restriction sites, segments of these genes were inserted into pET22b expression plasmid (Novagen 69744-3). Fibroblast growth factor 2 (FGF2-H6), as CD44 overexpression activator, was produced as described for the proteins above. All the encoded proteins were produced in *Escherichia coli* BL21 (F^- o*mpThsdS*$_B$(r_B^- m_B^-) *gal dcm*DE3, Novagen) overnight at 16°C for A5G27-GFP-H6,FNI/II/V-GFP-H6 and FGF2-H6, overnight at 20ºC for FNI-GFP-H6 and P7-GFP-H6 and during 3 h at 37ºC for FNV-GFP-H6. Gene expression was induced upon the addition of 1 mM IPTG. Bacterial cells were then centrifuged for 45 min (5000 g at 4°C) and resuspended in Tris buffer (20 mM Tris, pH 8.0, 500 mM NaCl, 10 mM imidazole) in the presence of protease inhibitor (Complete EDTA-Free, Roche, Basel, Switzerland). The cells were disrupted at 1200 psi in a French press (Thermo Scioentific). Lysates were centrifuged at 15,000g for 45 min and soluble fractions were filtrated before His-tag affinity chromatography using HiTrap Chelating HP 1 mL

columns (GE Healthcare, Piscataway, NJ) with an AKTA purifier FPLC (GE Healthcare). Elution was achieved by a linear gradient of 20 mM Tris, pH 8.0, 500 mM NaCl, and 500 mM imidazole, and the eluted material was analyzed by Western Blotting with an anti-His monoclonal antibody (Santa Cruz Biotechnology Inc., Heidelberg, Germany) to observe the presence of the protein of interest. An anion-exchange chromatography was additionally performed for A5G27-GFP-H6 and FNI/II/V-GFP-H6 proteins purification using DEAE HP and QFF HP 1 ml columns (GE Healthcare) respectively and a linear gradient of 10 mM Tris-HCl pH 8.0, 1 M NaCl for A5G27-GFP-H6 and 166 mM$NaHCO_3$, 1 M NaCl for FNI/II/V-GFP-H6. Proteins were finally dialyzed overnight at 4ºC against sodium bicarbonate buffer (166 mM $NaHCO_3$ at pH 7.4) except for FGF2-H6 that was dialysed with phosphate buffer (0.1 M $NaPO_3$, pH 6.3). Protein integrity and purity were checked by Coomassie Brilliant Blue staining, by mass spectrometry (MALDI-TOF) and N-terminal sequencing using the Edman degradation method. Protein concentration was determined by Bradford assay as described elsewhere (Bradford, 1976). A5G27-GFP-H6, FNV-GFP-H6 and FNI/II/V-GFP-H6 had been preliminarily screened for self-assembling in a previous study (Unzueta, Ferrer-Miralles et al., 2012). Protein production was supported by Protein Production Platform (CIBER-BBN – UAB, http://www.ciber-bbn.es/en/programas/89-plataforma-de-produccion-de-proteinas-ppp).

2.3. Atomic force microscopy, fluorescence determination, and dynamic light scattering. The fluorescence of chimerical proteins was determined in a Cary Eclipse fluorescence spectrophotometer (Varian Inc, Palo Alto, CA) at 510 nm using an excitation wavelength of 450 nm. Volume size distribution of nanoparticles was determined by dynamic light scattering at 633 nm (Zetasizer Nano ZS, Malvern Instruments Limited, Malvern, UK). Atomic force microscopy (AFM) analyses were performed with a commercial atomic force microscope (PicoSPM 5100 from Molecular Imaging Agilent Technologies, Inc., Santa Clara, CA, USA) operating in acoustic mode.

142 Proteins suspended in 166 mM $NaCO_3H$ at pH 7.4 were dropped onto a freshly cleaved
143 mica surface. The substrate was rinsed gently with miliQ water to eliminate the salts of
144 the buffer solution and let to air dry before imaged. For the acoustic mode
145 measurements, a monolithic supersharp silicon SSS-NCH-50 (Nanosensors, Inc.) tip,
146 with a radius of 2 nm, a nominal spring constant of 10–130 N/m and a resonance
147 frequency of 204–497 kHz were used.

148 **2.4. Flow cytometry.** MDA-MB-231, MCF-7, HEK-293-T and HepG2 cells were
149 cultured on a 24-well plate at $8 \cdot 10^4$, $1 \cdot 10^5$, $8 \cdot 10^4$ and $1.2 \cdot 10^5$ cells/well respectively with
150 their correspondent medium for 24 h until reaching 70 % confluence. Medium was
151 removed and cells were washed twice with PBS (Sigma-Aldrich Chemie GmbH,
152 Steinheim, Germany), and then OptiPro medium supplemented with L-glutamine and
153 peptide at the desired concentration was added and incubated for 24 h at 37ºC and 5%
154 CO_2 in a humidified atmosphere. Cells were detached using 1 mg/mL trypsin for 15 min
155 followed by the addition of complete medium and centrifuged at 1200 rpm for 15 min.
156 After supernatant was removed, the cell pellet was resuspended in 300 µl DPBS Ca^{2+},
157 Mg^{2+}-free (Invitrogen). Protein internalization was analyzed using a FACS-Canto
158 system (Becton Dickinson, Franklin Lakes, NJ) using a 15 mW air-cooled argon ion
159 laser at 488 nm excitation. Experiments were performed in duplicate.

160 **2.5. Protein internalisation monitored by confocal laser scanning microscopy.**
161 MDA-MB-231 cells were plated on a MatTek culture dish (MatTek Corporation,
162 Ashland, MA) at 200.000 cells/plate for 24 h. Medium was removed and cells were
163 washed with DPBS, OptiPro medium supplemented with L-glutamine and peptide at
164 1.5 µM was added and incubated for 24 h at 37ºC and 5 % CO_2 in a humidified
165 atmosphere before confocal analysis. Plasma membranes were labelled with 2.5
166 µg/mL CellMask™ Deep Red (Molecular Probes, Eugene, OR) and cell nuclei with 0.2
167 µg/mL Hoechst 33342 (Molecular Probes) for 10 min. in the dark before confocal
168 analysis. Cells were washed in PBS and complete medium was added. Analysis was

169 performed using a TCS-SP5 confocal laser scanning microscope (Leica Microsystems,
170 Heidelberg, Germany) using a Plan Apo 63 ×/1.4 (oil HC × PL APO lambda blue)
171 objective as described elsewhere (Vazquez, Roldan et al., 2010). Images were
172 processed using Imaris version 6.1.0 software (Bitplane, Zürich, Switzerland).

173 **2.6. Analysis of protein stability and cytoxicity.** Stability of A5G27-GFP-H6 and
174 FNI/II/V-GFP-H6 was analysed in duplicate in human serum (S2257-5ML, Sigma, St
175 Louis, MO) at 37°C, with agitation and at a final concentration of 0.115 µg/µL and 0.055
176 µg/µL respectively. Fluorescence was determined as previously described. Cell viability
177 was determined by 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT)
178 assay using VICTOR3V 1420 (Waltham, USA).

179 **2.7. Determination of CD44 expression.** $1.5 \cdot 10^6$ of each MDA-MB-231, MCF-7, HEK
180 293 T and HepG2 cells were tripsinised and centrifuged (1,200 rpm, 5 min, 4ºC) and
181 fixed in 100 µl of 2 % formaldehyde for 10 min at room temperature. The pellet was
182 washed with PBS-BSA. Cells were then resuspended in 75 µl of blocking solution
183 (Human Ig 200 µg/mL PBS) for 15 min at room temperature, and aliquoted into 3x25 µl
184 fractions. 5 µl of APC Mouse IgG2b κ Isotype control (BD Pharmingen, 555745) and 5
185 µl of APC Mouse Anti-Human CD44 (BD Pharmingen, 559942) were added to two of
186 these samples respectively. 45 min after incubation at 4ºC out from light, samples were
187 centrifuged (8,000 rpm, 15 sec, 4ºC) and washed with 1 ml PBS-BSA (0.5 %). Pellet
188 was resuspended in 500 µl PBS. Samples were analysed in duplicate with a
189 FacsCalibur cytometer (Becton Dickinson, Franklin Lakes, NJ).

190 **2.9. Peptide internalisation under CD44 deregulation and competition assays.**
191 MDA-MB-231 cells were cultured on a 24 well-plate at $8 \cdot 10^4$ c/w for 24 h until 70 %
192 confluence. Medium was removed and cells washed twice with PBS. 250 µl of OptiPro
193 medium supplemented with L-glutamine with the CD44 expression regulator/competitor
194 was added. Competitor anti-CD44 and downregulator anti-IL-10 polyclonal antibodies

195 (HCAM sc-7946 and IL-10 sc-7888 respectively, Santa Cruz Biotechnology Inc.,
196 Heidelberg, Germanyand) were used at 0.3 µM whereas positive regulatorFGF2 was
197 added at different ratios (1:1, 1:10, 1:30). After 1 h incubation of regulator/competitor at
198 37°C, proteins were added at 0.3 µM. Cells were detached and prepared for cytometry
199 analysis as described above.

200 **2.10. Statistical analysis.** Mean data and other statistics were calculated with
201 Sigmaplot 10.0.

202 **Table 1. CD44 peptidic ligands used to tag GFP-H6 monomers.**

Source protein and reference	CD44 ligand protein segment	Length (in amino acids)	Number of positively charged residues (arg + lys)	Amino acid sequence	Particle formation and size of monomer or oligomer (nm/PDI[a])	Construct name. The N-terminal module corresponds to the peptide ligand.
Fibronectin, HBFN-fragment I, (Jalkanen and Jalkanen, 1992; Yasuda, Poole et al., 2003)	1977-1991	15	5	KNNQKSEPLIGRKKT	No, 6.3/0.7	FNI-GFP-H6
Fibronectin, HBFN-fragment V, (Jalkanen and Jalkanen, 1992; Yasuda, Poole et al., 2003)	1923-1930	8	2	WQPPRARI	No, 5.5/0.8	FNV-GFP-H6
Fibronectin, HBFN-I/II/V containing fragment, (Jalkanen and Jalkanen, 1992; Yasuda, Poole et al., 2003)	1923-1991	69	12	WQPPRARITGYIIKY EKPGSPPREVVPRP RPGVTEATITGLEPG TEYTIYVIALKNNQKS EPLIGRKKT	Yes, 13.7/0.4	FNI/II/V-GFP-H6
Laminin α5 chain, peptide A5G27, (Hibino, Shibuya et al., 2004; Hibino, Shibuya et al., 2005)	2975-2987	13	2	RLVSYNGIIFFLK	Yes, 13.8/0.4 [b]	A5G27-GFP-H6

P7, phage display derived peptide, (Park, Lee et al., 2012)	-	12	2	FNLPLPSRPLLR	No, 6.5/0.4	P7-GFP-H6

[a] The indicated size refers to the peak determined by DLS. PDI is the polydispersity index in the DLS measurements.

[b] Slight discrepancies between the size of A5G27-GFP-H6 nanoparticles determined here and in a previous report (Unzueta, Cespedes et al., 2012) are due to the different composition of the buffers used in these studies.

206 **3. Results**

207 Five known peptidic ligands of CD44 (FNI, FNV, FNI/II/V, A5G27 and P7, Table 1)
208 were fused to C-terminal H6-tagged GFP following a previously described approach
209 (Figure 1 A, (Unzueta, Ferrer-Miralles et al., 2012)), to construct multidomain GFP
210 protein versions with affinity for CD44+ cells. All these constructs were efficiently
211 produced in bacteria, resulting in full length polypeptides with expected mass and
212 predicted N-terminal amino acid sequence (see Supplementary Figure 1). Before
213 testing for biological properties, the potential self-assembling of protein monomers into
214 higher order entities was explored. According to a previously proposed model
215 (Unzueta, Ferrer-Miralles et al., 2012), the highly cationic peptide FNI/II/V was
216 expected to promote the formation of ordered oligomers within the nanoscale. In this
217 regard, FNI/II/V-GFP-H6 in solution peaked at around 14 nm by DLS (Figure 1 B, Table
218 1), indicative of nanoparticle formation. The unassembled GFP-H6 protein showed a
219 size of 6 nm (Vazquez, Roldan et al., 2010), and FNI, FNV and P7-empowered GFP
220 proteins peaked at the same size, indicating that they remained unassembled like the
221 parental GFP-H6 (Figure 1 B, Table 1). Unexpectedly, A5G27, even being poorly
222 cationic (Table 1), also promoted the formation of supramolecular entities of 14 nm
223 (Figure 1 B, Table 1).

224 When the nanoparticulate architecture of FNI/II/V-GFP-H6 and A5G27-GFP-H6 was
225 assessed by AFM, these proteins clearly organized as regular nanoparticles with a size
226 fully compatible with DLS determinations (Figure 1C-G). Some A5G27-GFP-H6
227 particles appeared as rings or pseudorings with a centred cavity (Figure 1E), and such
228 a ring-based architecture was generic in FNI/II/V-GFP-H6 samples, in which some of
229 the nanoparticles showed a pentameric organization (Figure 1F). In this regard,
230 molecular modelling of R9-GFP-H6 nanoparticles indicated a pentameric organization
231 of the constructs (Unzueta, Ferrer-Miralles et al., 2012; Vazquez, Cubarsi et al., 2010;
232 Vazquez, Roldan et al., 2010) and size-exclusion chromatography of several GFP and

233 iRFP-based nanoparticles also showed pentamers as a basic module resulting from
234 self-assembling in protein particles empowered by cationic stretches plus H6
235 (Cespedes, Unzueta et al., 2014).

236 To evaluate the biological properties of the fusion proteins, namely cell penetration and
237 eventual intracellular trafficking, we first checked the specific fluorescence of all fusion
238 proteins, as fluorescence emission is a convenient reporter to monitor cell binding and
239 internalization. As observed (Figure 2 A), fluorescence emission of the enhanced GFP
240 was not straightforward affected by protein fusion but it was reduced up to around 30-
241 40 % of the parental protein in A5G27-GFP-H6 and FNI/II/V-GFP-H6. The coincidence
242 between fluorescence dropping and nanoparticle formation might be indicative of a
243 slight quenching effect associated to oligomer formation. Irrespectively of the precise
244 cause, fluorescence levels were well acceptable and sufficient for further analyses. An
245 important fraction of $CD44^+$ MDA-MB-231 cells were fluorescent when exposed to
246 A5G27-GFP-H6 and FNI/II/V-GFP-H6 (around 80 % *vs* 40 % or less in the case of
247 other fusions, Figure 2B), and a higher global fluorescence emission was observed in
248 cell cultures exposed to these proteins when compared with alternative GFP versions
249 (between 1000 and 2000 *vs* up to 600, Figure 2C). In both experimental approaches
250 and as it was expected, GFP-H6 only rendered background values. In agreement with
251 quantitative data, confocal analyses of protein-exposed MDA-MB-231 cell cultures
252 confirmed the high penetrability of A5G27-GFP-H6 and FNI/II/V-GFP-H6 *vs* alternative
253 GFP versions, mostly showing a background uptake (Figure 2D). Note that as indicated
254 above, the lower specific fluorescence of these two constructs resulted in an
255 underestimation of the internalized material when compared to alternative ligands.
256 Considering the values presented in the Figure 2B, A5G27-GFP-H6 and FNI/II/V-GFP-
257 H6 internalized 20 times more efficiently than the alternative fusion proteins. Again, the
258 parental, H6-tagged GFP was seen as totally excluded from cultured cells. The

incorporation of the internalized recombinant proteins into membranous vesicles was clearly observed and it was indicative of endosomal uptake (Figure 2D).

Internalization of A5G27-GFP-H6 and FNI/II/V-GFP-H6 occurred very fast, and the fraction of target cells and the amount of intracellular protein reached a plateau at about 10 h (Figure 3A). Interestingly, the penetration of these nanoparticles did take place without compromising cell viability (Figure 3B), again in agreement with the occurrence of an endosomal route and supporting the full biocompatibility nature of protein-only nanoparticles. In this context, both A5G27-GFP-H6 and FNI/II/V-GFP-H6 were fully stable when incubated in human serum (Figure 3C), a fact that confirmed the structural robustness of the particles and prompted us to envisage a potential for proper biodistribution of these materials in targeted drug delivery or diagnostic applications.

To assess the receptor-driven specificity of the cell penetration, the uptake of protein nanoparticles was explored in several cell lines, expressing and not expressing CD44, namely MDA-MB-231, MCF-7, Hep G2 and HEK-293T. A differential cell penetrability of A5G27-GFP-H6 and FNI/II/V-GFP-H6 was observed (Figure 4A), coincident with the amount of cellular CD44 in the target cells (Figure 4B). This fact strongly supported the CD44-dependence of cell binding. In this context, a commercial polyclonal anti-CD44 serum inhibited the entrance of both type of nanoparticles (Figure 5). Finally, we wanted to determine if the externally mediated up- or down-regulation of CD44 expression could have enhancing or inhibiting effects on the penetration of nanoparticles, as expected. The fibroblast growth factor 2 (FGF-2), a positive regulator of CD44 (Grimme, Termeer et al., 1999; Jones, Tussey et al., 2000), stimulated the cell penetrability of A5G27-GFP-H6 (Figure 6A), while the blocking of interleukin-10 (IL-10; also a positive regulator of CD44, (Gee, Kryworuchko et al., 2004)) by a specific antibody reduced the uptake of the nanoparticle (Figure 6B). By the combination of

these data, the CD44-targeting of the protein constructs developed here was fully demonstrated.

4. Discussion

Developing tools for targeted drug delivery is a priority in targeted medicines of cancer and other prevalent diseases (Ruoslahti, Bhatia et al., 2010; Vicent and Duncan, 2006). While of the number of cell surface proteins identified as valuable markers are rapidly expanding (Gonzalez-Angulo, Hennessy et al., 2010; Klonisch, Wiechec et al., 2008; Mocellin, Lise et al., 2005; Nguyen and Massague, 2007; Ruoslahti, Bhatia et al., 2010; Sawyers, 2008; Tjalsma, 2010), efficient and secure vehicles are still missing. While consensus exists in that nanoscale containers are ideal for competent systemic transport, diffusion in the tissue and cell penetrability (Mastrobattista, van der Aa et al., 2006), a fully biocompatible material remains to be developed. In this regard, the number of nanomedicines so far approved by the medicament agencies is still limited (Duncan and Gaspar, 2011).

A promising route to the generation of efficient vehicles for drug delivery is conventional genetic engineering, since proteins are fully biocompatible and functional macromolecules. The ability to recruit diverse peptides from distinct origins and with different complementing activities in a single chain protein offers promise to generate constructs showing the biological properties exhibited by viruses during infection. These include stable systemic circulation, receptor targeting, internalization, intracellular trafficking and accumulation into the appropriate compartment (Aris and Villaverde, 2004; Ferrer-Miralles, Vazquez et al., 2008; Vazquez, Ferrer-Miralles et al., 2009; Vazquez, Ferrer-Miralles et al., 2008). Despite this potential, the exploration of how protein-protein contacts could be engineered to construct protein-only nanoparticles has been in general neglected. The adaptation of virus-like particles (VLP) (Ma, Nolte et al., 2012), bacterial micro compartments (BMC) (Corchero and Cedano, 2011), eukaryotic vaults (Rome and Kickhoefer, 2012) or other natural protein constructs (Rodriguez-Carmona and Villaverde, 2010) poses severe limitations, as conferring new tropisms might in general alters the stability of the particle. However,

recent lessons about how protein-protein interactions can be engineered for self-assembling in fully *de novo* designed protein constructs {Neus Ferrer-Miralles, 2013 1106 /id;Unzueta, 2014 1308 /id;Villaverde, 2012 7494 /id} should permit to approach an 'artificial virus' strategy for the design of novel nanomedicines (Mastrobattista, van der Aa et al., 2006). Importantly, the recent advances in systems and synthetic biotechnology and industrial microbiology (Lee, Mattanovich et al., 2012) allow the large scale biosynthesis of natural proteins and protein constructs, being biological biofabrication highly versatile (Vazquez and Villaverde, 2013) and progressively more competitive with regard to conventional chemical synthesis (Chen, 2012).

We have here constructed a set of five modular polypeptides intended to target CD44+ cells (Table 1). Among them, two constructs (empowered by peptides A5G27 and FNI/II/V respectively), self-organize as stable nanoparticles (Figure 1 and 3C) that efficiently bind and internalize CD44+ target cells (Figure 4), accumulating in the perinuclear and nuclear regions (Figure 2). Since the five peptides used here are very well known ligands of CD44 (Table 1), the coincidence between nanoparticle formation and internalization strongly suggests that the presentation of a given protein in form of nanoparticle (versus the plain monomeric form) stimulates cell penetration. This is probably because of the multiple ligand display and multivalent cross-linking at the cell surface, favoring membrane wrapping (Jiang, Kim et al., 2008), and in the line of nanoparticle size being a main determinant of interactions with cells (Jiang, Kim et al., 2008). Also, this is in agreement with the high penetrability found in natural oligomers when displaying cell-binding peptides (Aris and Villaverde, 2003; Villaverde, Feliu et al., 1998).

While other engineering principles have been recently proposed to control protein self-assembling (Bai, Luo et al., 2013) (King, Sheffler et al., 2012), the approach based on the end terminal fusion of a cationic peptide and a polyhistidine is not restricted to a unique core protein. This versatility would be convenient to avoid immunogenicity of the

342 constructs by selecting homologous proteins in next generation-constructions. By using
343 the end-terminal peptide-pair strategy, the formation of nanoparticles with pentameric,
344 toroid-like organization had been previously predicted (Unzueta, Ferrer-Miralles et al.,
345 2012; Vazquez, Roldan et al., 2010) and demonstrated by FESEM for a certain
346 category of proteins (those empowered by the cationic peptide T22) (Cespedes,
347 Unzueta et al., 2014). The highly resolutive AFM has confirmed this particular
348 architecture also for A5G27-GFP-H6 and FNI/II/V- GFP-H6 (Figure 1E,F), indicative
349 that the ring shaped distribution of the building blocks is not restricted to an unique type
350 of end-terminal tags. Being highly cationic (Table 1), the oligomerization of FNI/II/V-
351 GFP-H6 was fully anticipated at the upstream stage, while the formation of A5G27-
352 GFP-H6-based nanoparticles (being A5G27 poorly cationic) was unexpected. Acting
353 the amino terminal stretch both as architectonic tag and cell ligand, the promotion of
354 protein self-assembling by a non-cationic peptide expands, in any case, the spectrum
355 of potential ligands usable for the formation of cell-targeted nanoparticles, so far
356 restricted to cationic protein regions.

357 Laminin and fibronectin ligands bind CD44 through chondroitin and heparin-like GAG
358 side chains, especially by the heparin-sulphate found in CD44v3 and CD44v6 isoforms.
359 On the other hand, the alternative CD44 ligand HA binds to a binding site termed the
360 “link module” which is a domain expressed in all CD44 isoforms and it is located on
361 CD44 most exposed region (far from V3 and v6 variant regions) (Peach, Hollenbaugh
362 et al., 1993). Therefore, although HA could not be used as a competitor to probe the
363 CD44 targeting of the protein constructs developed here, the specificity in binding and
364 internalization was successfully demonstrated by the coincidence between CD44 levels
365 and penetrability (Figure 4), by inhibition mediated by a polyclonal anti-CD44 serum
366 (Figure 5), and though the external alternate regulation of CD44 levels and the
367 consequent variation in the efficiency of nanoparticle uptake (Figure 6).

In summary, we have developed smart and stable protein-only nanoparticles (A5G27-GFP-H6 and FNI/II/V-GFP-H6) as plastic agents that bind specifically CD44+ cells and that are efficiently internalized by receptor-mediated endocytosis in absence of cell toxicity. Interestingly, the vehicle itself is composed by fully functional (fluorescent, in our model system) proteins, what opens a plethora of possibilities regarding the targeted delivery of therapeutic polypeptides in form of nanoparticles. The nanostructure gained by these proteins would desirably allow them escaping from renal clearance, as the final size of the constructs is largely over the 6-7 nm cut-off. In addition, protein nanoparticles could be loaded with chemically coupled conventional drugs, as the principle of protein-drug coupling has been largely proved in already licensed drugs (Elzoghby, Samy et al., 2012). Altogether, and according to recent data proving the high stability *in vivo* (Cespedes, Unzueta et al., 2014) and efficient biodistribution of similar protein-only modular constructs (Unzueta, Cespedes et al., 2012), the tools generated here are promising platforms as vehicles for drugs and imaging agents, in the context of emerging nanomedines for breast cancer and other metastatic CD44-linked tumours based on biocompatible and versatile protein materials.

Acknowledgments

We appreciate the technical support of Fran Cortés from the Cell Culture Unit of Servei de Cultius Cel.lulars Producció d'Anticossos i Citometria (SCAC, UAB), of the Servei de Microscòpia and of Amable Bernabé from Soft Materials Service (ICMAB-CSIC/CIBER-BBN) and from Proteomics facility from UAB (a member of ProteoRed-ISCIII network) . We are also indebted to the Protein Production Platform (CIBER-BBN - UAB) for helpful technical assistance in protein production and purification (http://www.ciber-bbn.es/en/programas/89-plataforma-de-produccion-de-proteinas-

394 ppp). The authors also acknowledge the financial support granted to E.V. from FIS
395 (PI12/00327), to S. S. from FIS (PI11/01079), to E.V. and S.S. from The Marató de TV3
396 (TV32013-133930), to J.V from DGI (Grant POMAs CTQ2010-19501) and to A.V. from
397 MINECO (BIO2013-41019-P), from Generalitat de Catalunya (2014SGR-132) and from
398 the Centro de Investigación Biomédica en Red (CIBER) de Bioingeniería,
399 Biomateriales y Nanomedicina (NANOPROTHER and PENTI projects), financed by the
400 Instituto de Salud Carlos III with assistance from the European Regional Development
401 Fund. M.P. and U.U. received PhD fellowships from ISCIII and from UAB respectively.
402 W.T.is grateful to the Consejo Superior de Investigaciones Científicas (CSIC) for a
403 "JAE-pre"fellowship. A.V. has been distinguished with an ICREA ACADEMIA Award.

404

405

406 Reference List

407

408 Allen,M.J., Hud,N.V., Balooch,M., Tench,R.J., Siekhaus,W.J., Balhorn,R., 1992. Tip-radius-
409 induced artifacts in AFM images of protamine-complexed DNA fibers. Ultramicroscopy, 42-44 (
410 Pt B), 1095-1100.

411 Aris,A., Villaverde,A., 2003. Engineering nuclear localization signals in modular protein vehicles
412 for gene therapy. Biochem. Biophys. Res. Commun., 304, 625-631.

413 Aris,A., Villaverde,A., 2004. Modular protein engineering for non-viral gene therapy. Trends
414 Biotechnol., 22, 371-377.

415 Bai,Y., Luo,Q., Zhang,W., Miao,L., Xu,J., Li,H., Liu,J., 2013. Highly ordered protein nanorings
416 designed by accurate control of glutathione s-transferase self-assembly. J. Am. Chem Soc., 135,
417 10966-10969.

418 Bradford,M.M., 1976. A rapid and sensitive method for the quantitation of microgram
419 quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem., 72, 248-254.

420 Cespedes,M.V., Unzueta,U., Tatkiewicz,W., Sanchez-Chardi,A., Conchillo-Sole,O., Alamo,P.,
421 Xu,Z., Casanova,I., Corchero,J.L., Pesarrodona,M., Cedano,J., Daura,X., Ratera,I., Veciana,J.,
422 Ferrer-Miralles,N., Vazquez,E., Villaverde,A., Mangues,R., 2014. In Vivo Architectonic Stability
423 of Fully de Novo Designed Protein-Only Nanoparticles. ACS Nano., 8, 4166-4176.

424 Chen,G.Q., 2012. New challenges and opportunities for industrial biotechnology. Microb. Cell
425 Fact., 11, 111.

426 Corchero,J.L., Cedano,J., 2011. Self-assembling, protein-based intracellular bacterial
427 organelles: emerging vehicles for encapsulating, targeting and delivering therapeutical
428 cargoes. Microb Cell Fact., 10, 92.

429 Duncan,R., Gaspar,R., 2011. Nanomedicine(s) under the microscope. Mol. Pharm., 8, 2101-
430 2141.

431 Elzoghby,A.O., Samy,W.M., Elgindy,N.A., 2012. Albumin-based nanoparticles as potential
432 controlled release drug delivery systems. J. Control Release, 157, 168-182.

433 Ferrer-Miralles,N., Corchero,J.L., Kumar,P., Cedano,J.A., Gupta,K.C., Villaverde,A., Vazquez,E.,
434 2011. Biological activities of histidine-rich peptides; merging Biotechnology and Nanomedicine.
435 Microb Cell Fact., 10, 101.

436 Ferrer-Miralles,N., Vazquez,E., Villaverde,A., 2008. Membrane-active peptides for non-viral
437 gene therapy: making the safest easier. Trends Biotechnol., 26, 267-275.

438 Gee,K., Kryworuchko,M., Kumar,A., 2004. Recent advances in the regulation of CD44
439 expression and its role in inflammation and autoimmune diseases. Arch. Immunol. Ther. Exp.
440 (Warsz.), 52, 13-26.

441 Gonzalez-Angulo,A.M., Hennessy,B.T., Mills,G.B., 2010. Future of personalized medicine in
442 oncology: a systems biology approach. J. Clin. Oncol., 28, 2777-2783.

443 Goodison,S., Urquidi,V., Tarin,D., 1999. CD44 cell adhesion molecules. Mol. Pathol., 52, 189-
444 196.

445 Grimme,H.U., Termeer,C.C., Bennett,K.L., Weiss,J.M., Schopf,E., Aruffo,A., Simon,J.C., 1999.
446 Colocalization of basic fibroblast growth factor and CD44 isoforms containing the variably
447 spliced exon v3 (CD44v3) in normal skin and in epidermal skin cancers. Br. J. Dermatol., 141,
448 824-832.

449 Hibino,S., Shibuya,M., Engbring,J.A., Mochizuki,M., Nomizu,M., Kleinman,H.K., 2004.
450 Identification of an active site on the laminin alpha 5 chain globular domain that binds to CD44
451 and inhibits malignancy. Cancer Res, 64, 4810-4816.

452 Hibino,S., Shibuya,M., Hoffman,M.P., Engbring,J.A., Hossain,R., Mochizuki,M., Kudoh,S.,
453 Nomizu,M., Kleinman,H.K., 2005. Laminin alpha5 chain metastasis- and angiogenesis-inhibiting
454 peptide blocks fibroblast growth factor 2 activity by binding to the heparan sulfate chains of
455 CD44. Cancer Res., 65, 10494-10501.

456 Jalkanen,S., Jalkanen,M., 1992. Lymphocyte CD44 binds the COOH-terminal heparin-binding
457 domain of fibronectin. J. Cell Biol., 116, 817-825.

458 Jiang,W., Kim,B.Y., Rutka,J.T., Chan,W.C., 2008. Nanoparticle-mediated cellular response is
459 size-dependent. Nat. Nanotechnol., 3, 145-150.

460 Jones,M., Tussey,L., Athanasou,N., Jackson,D.G., 2000. Heparan sulfate proteoglycan isoforms
461 of the CD44 hyaluronan receptor induced in human inflammatory macrophages can function
462 as paracrine regulators of fibroblast growth factor action. J. Biol. Chem, 275, 7964-7974.

463 King,N.P., Sheffler,W., Sawaya,M.R., Vollmar,B.S., Sumida,J.P., Andre,I., Gonen,T., Yeates,T.O.,
464 Baker,D., 2012. Computational design of self-assembling protein nanomaterials with atomic
465 level accuracy. Science, 336, 1171-1174.

466 Klonisch,T., Wiechec,E., Hombach-Klonisch,S., Ande,S.R., Wesselborg,S., Schulze-Osthoff,K.,
467 Los,M., 2008. Cancer stem cell markers in common cancers - therapeutic implications. Trends
468 Mol. Med., 14, 450-460.

469 Koo,H., Huh,M.S., Sun,I.C., Yuk,S.H., Choi,K., Kim,K., Kwon,I.C., 2011. In vivo targeted delivery
470 of nanoparticles for theranosis. Acc. Chem. Res., 44, 1018-1028.

471 Lee,S.Y., Mattanovich,D., Villaverde,A., 2012. Systems metabolic engineering, industrial
472 biotechnology and microbial cell factories. Microb. Cell Fact., 11, 156.

473 Ma,Y., Nolte,R.J., Cornelissen,J.J., 2012. Virus-based nanocarriers for drug delivery. Adv. Drug
474 Deliv. Rev., 64, 811-825.

475 Marangoni,E., Lecomte,N., Durand,L., de,P.G., Decaudin,D., Chomienne,C., Smadja-Joffe,F.,
476 Poupon,M.F., 2009. CD44 targeting reduces tumour growth and prevents post-chemotherapy
477 relapse of human breast cancers xenografts. Br. J. Cancer, 100, 918-922.

478 Mastrobattista,E., van der Aa,M.A., Hennink,W.E., Crommelin,D.J., 2006. Artificial viruses: a
479 nanotechnological approach to gene delivery. Nat. Rev. Drug Discov., 5, 115-121.

480 Mocellin,S., Lise,M., Nitti,D., 2005. Targeted therapy for colorectal cancer: mapping the way.
481 Trends Mol. Med., 11, 327-335.

482 Neus Ferrer-Miralles, Escarlata Rodriguez-Carmona, Jose Luis Corchero, Elena Garcia-Fruitos,
483 Esther Vazquez, Antonio Villaverde, 2013. Engineering protein self-assembling in protein-based
484 nanomedicines for drug delivery and gene therapy. Crit Rev. Biotechnol, in press.

485 Nguyen,D.X., Massague,J., 2007. Genetic determinants of cancer metastasis. Nat. Rev. Genet.,
486 8, 341-352.

487 Park,H.Y., Lee,K.J., Lee,S.J., Yoon,M.Y., 2012. Screening of peptides bound to breast cancer
488 stem cell specific surface marker CD44 by phage display. Mol. Biotechnol, 51, 212-220.

489 Peach,R.J., Hollenbaugh,D., Stamenkovic,I., Aruffo,A., 1993. Identification of hyaluronic acid
490 binding sites in the extracellular domain of CD44. J. Cell Biol., 122, 257-264.

491 Peer,D., Karp,J.M., Hong,S., Farokhzad,O.C., Margalit,R., Langer,R., 2007. Nanocarriers as an
492 emerging platform for cancer therapy. Nat. Nanotechnol., 2, 751-760.

493 Rezler,E.M., Khan,D.R., Lauer-Fields,J., Cudic,M., Baronas-Lowell,D., Fields,G.B., 2007. Targeted
494 drug delivery utilizing protein-like molecular architecture. J. Am. Chem. Soc., 129, 4961-4972.

495 Rodriguez-Carmona,E., Villaverde,A., 2010. Nanostructured bacterial materials for innovative
496 medicines. Trends Microbiol., 18, 423-430.

497 Rome,L.H., Kickhoefer,V.A., 2012. Development of the Vault Particle as a Platform Technology.
498 ACS Nano..

499 Ruoslahti,E., Bhatia,S.N., Sailor,M.J., 2010. Targeting of drugs and nanoparticles to tumors. J.
500 Cell Biol., 188, 759-768.

501 Sauter,A., Kloft,C., Gronau,S., Bogeschdorfer,F., Erhardt,T., Golze,W., Schroen,C., Staab,A.,
502 Riechelmann,H., Hoermann,K., 2007. Pharmacokinetics, immunogenicity and safety of

503 bivatuzumab mertansine, a novel CD44v6-targeting immunoconjugate, in patients with
504 squamous cell carcinoma of the head and neck. Int. J. Oncol., 30, 927-935.

505 Sawyers,C.L., 2008. The cancer biomarker problem. Nature, 452, 548-552.

506 Tijink,B.M., Buter,J., de,B.R., Giaccone,G., Lang,M.S., Staab,A., Leemans,C.R., van Dongen,G.A.,
507 2006. A phase I dose escalation study with anti-CD44v6 bivatuzumab mertansine in patients
508 with incurable squamous cell carcinoma of the head and neck or esophagus. Clin. Cancer Res.,
509 12, 6064-6072.

510 Tjalsma,H., 2010. Identification of biomarkers for colorectal cancer through proteomics-based
511 approaches. Expert. Rev. Proteomics., 7, 879-895.

512 Unzueta,U., Cespedes,M.V., Ferrer-Miralles,N., Casanova,I., Cedano JA, Corchero JL, Domingo-
513 Espin,J., Villaverde A, Mangues,R., Vazquez E, 2012. Intracellular CXCR4^{+} cell targeting with
514 T22-empowered protein-only nanoparticles. Int. J. Nanomedicine, 7, 4533-4544.

515 Unzueta,U., Ferrer-Miralles,N., Cedano,J., Zikung,X., Pesarrodona,M., Saccardo,P., Garcia-
516 Fruitos,E., Domingo-Espin,J., Kumar,P., Gupta,K.C., Mangues,R., Villaverde,A., Vazquez,E.,
517 2012. Non-amyloidogenic peptide tags for the regulatable self-assembling of protein-only
518 nanoparticles. Biomaterials, 33, 8714-8722.

519 Unzueta,U., Saccardo,P., Domingo-Espin,J., Cedano,J., Conchillo-Sole,O., Garcia-Fruitos,E.,
520 Cespedes,M.V., Corchero,J.L., Daura,X., Mangues,R., Ferrer-Miralles,N., Villaverde,A.,
521 Vazquez,E., 2014. Sheltering DNA in self-organizing, protein-only nano-shells as artificial
522 viruses for gene delivery. Nanomedicine.10:535-541.

523 Vazquez,E., Cubarsi,R., Unzueta,U., Roldan,M., Domingo-Espin,J., Ferrer-Miralles,N.,
524 Villaverde,A., 2010. Internalization and kinetics of nuclear migration of protein-only, arginine-
525 rich nanoparticles. Biomaterials, **31**, 9333-9339.

526 Vazquez,E., Ferrer-Miralles,N., Mangues,R., Corchero,J.L., Schwartz S Jr, Villaverde,A., 2009.
527 Modular protein engineering in emerging cancer therapies. Curr. Pharm. Des, 15, 893-916.

528 Vazquez,E., Ferrer-Miralles,N., Villaverde,A., 2008. Peptide-assisted traffic engineering for
529 nonviral gene therapy. Drug Discov. Today, 13, 1067-1074.

530 Vazquez,E., Roldan,M., Diez-Gil,C., Unzueta,U., Domingo-Espin,J., Cedano,J., Conchillo,O.,
531 Ratera,I., Veciana,J., Daura,X., Ferrer-Miralles,N., Villaverde,A., 2010. Protein nanodisk
532 assembling and intracellular trafficking powered by an arginine-rich (R9) peptide.
533 Nanomedicine. (Lond), 5, 259-268.

534 Vazquez,E., Villaverde,A., 2013. Microbial biofabrication for nanomedicine: biomaterials,
535 nanoparticles and beyond. Nanomedicine (Lond), 8, 1895-1898.

536 Vicent,M.J., Duncan,R., 2006. Polymer conjugates: nanosized medicines for treating cancer.
537 Trends Biotechnol., 24, 39-47.

538 Villaverde,A., Feliu,J.X., Aris,A., Harbottle,R.P., Benito,A., Coutelle,C., 1998. A cell adhesion
539 peptide from foot-and-mouth disease virus can direct cell targeted delivery of a functional
540 enzyme. Biotechnol Bioeng, 59, 294-301.

541 Villaverde,A., Garcia-Fruitos,E., Rinas,U., Seras-Franzoso,J., Kosoy,A., Corchero,J.L., Vazquez,E.,
542 2012. Packaging protein drugs as bacterial inclusion bodies for therapeutic applications.
543 Microb Cell Fact., 11, 76.

544 Yasuda,T., Poole,A.R., Shimizu,M., Nakagawa,T., Julovi,S.M., Tamamura,H., Fujii,N.,
545 Nakamura,T., 2003. Involvement of CD44 in induction of matrix metalloproteinases by a
546 COOH-terminal heparin-binding fragment of fibronectin in human articular cartilage, in
547 culture. Arthritis and Rheumatism, 48, 1271-1280.

548 Zoller,M., 2011. CD44: can a cancer-initiating cell profit from an abundantly expressed
549 molecule? Nat. Rev. Cancer, 11, 254-267.
550
551
552

553 Figure legends:

554

555 **Figure 1.** Construction and nanoscale characterization of CD44-targeted protein
556 nanoparticles. A) Schematic representation of the gene fusion scheme used in this
557 study. L represents a CD44 ligand that in some cases also has an architectonic role. B)
558 Size distribution of the protein constructs determined by DLS. Numerical values are
559 given in Table 1. C) AFM images of randomly selected A5G27-GFP-H6 nanoparticles.
560 D) Topography cross- section of one randomly selected isolated A5G27-GFP-H6
561 nanoparticle. E) Topography cross- section of two ring shaped A5G27-GFP-H6 nano
562 particles. F) AFM images of randomly selected FNI/II/V-GFP-H6 nanoparticles,
563 showing the topography cross- section of one isolated particle. G) Images of a
564 pentameric particle are shown. Measurements have been done with a tip radius of 2
565 nm and thus the width (but not the high) of the particles is inherently overestimated. An
566 AFM image is a convolution of the imaging tip shape/size with the actual shape of the
567 imaged object (Allen, Hud et al., 1992). Thus, one will observe broadening of the
568 sample features.

569

570 **Figure 2.** Internalization of CD44-targeted protein nanoparticles. A) Specific
571 fluorescence of the different protein constructs in comparison to that of the parental
572 GFP-H6 (in green). The specific green fluorescence of the parental protein is 1,021
573 fluorescence units (FU)/ug. B) Percentage of MDA-MB-231 cells that are fluorescent
574 after 24 h exposure to the multidomain proteins. C) Green fluorescence emitted by
575 MDA-MB-231 cells after 24 h exposure to multidomain proteins. D) Confocal sections
576 or projections of MDA-MB-231 cells upon 24 h of exposure to multidomain proteins.
577 Bars indicate 20 µm. E) Details of target cells during the uptake of fusion proteins,
578 indicating the exogenous material (in green) included in endosomes (red signal).

579 Merging into yellow is evident in some cases. A 3D projection is included in the case of
580 FNI/II/V-GFP-H6.

581

582 **Figure 3.** Kinetics of cellular internalization of CD44-targeted protein nanoparticles,
583 cytotoxicity and protein serum stability. A) Time course cell penetration of protein-only
584 nanoparticles at 1.5 μM. Percentatge of protein-internalised cells (left) and cell EGFP
585 fluorescence intensity (right). B) MTT viability analysis of MDA-MB-231cells upon
586 exposure to different doses of protein nanoparticles for 24,48 and 72h.. C) Stability of
587 protein nanoparticles in human. Fluorescence emission of serum samples with
588 nanoparticle incubation at different time point up to 24 hours. Soluble 24h corresponds
589 to fluorescence of soluble fraction from 24 hours sample after centrifugation to discard
590 nanoparticle aggregation/precipitation

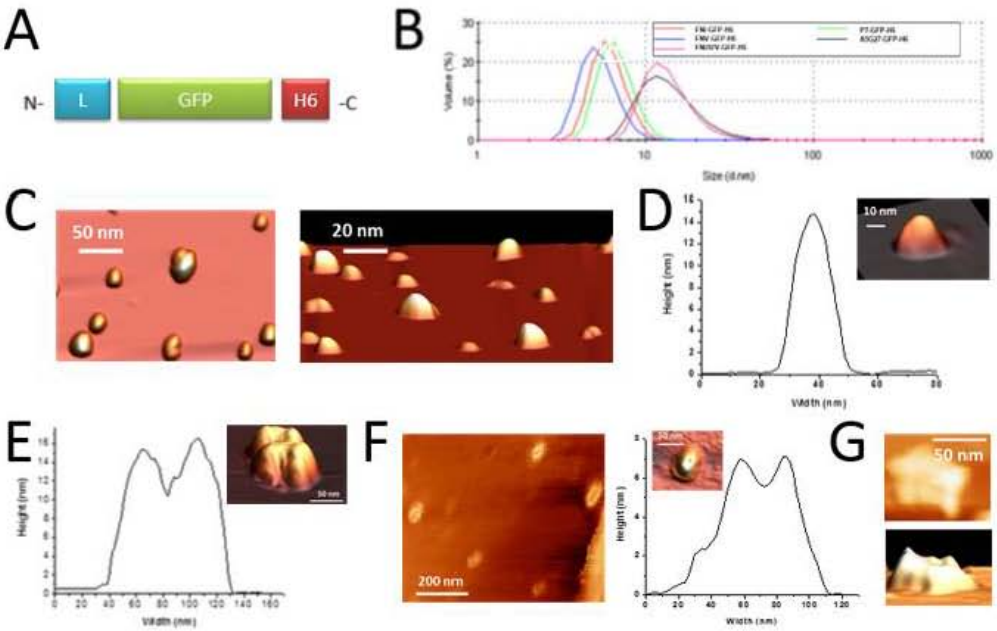
591 **Figure 4:** Specific internalization of CD44-targeted protein nanoparticles in CD44-
592 expressing and not expressing cells. A) Percentage of cells uptaking A5-GFP-H6 (left)
593 and FNI/II/V-GFP-H6 (right) in CD44-overexpressing MDA-MB-231 and MCF-7 lines
594 and in Hep G2 and HEK 293 T lines. B) Histograms of CD44-expressing cell population
595 from $CD44^+$ cell lines MDA-MB-231 and MCF-7 and $CD44^-$ cell lines HepG2 and HEK-
596 293-T. APC-anti-CD44 marked cells (red) are compared with isotopic control cells
597 (black). FL4-H axis corresponds to APC intensity. The percentage of $CD44^+$ cells is
598 indicated.

599 **Figure 5:** Anti-CD44 mediated inhibition of nanoparticle internalization. Decrease on
600 the intensity of intracellular fluorescence mediated by 0.3 μM of either A5G27-GFP-H6
601 or FNI/II/V-GFP-H6 (control: back bar) when adding 0.3 μM polyclonal anti-CD44 (ratio
602 1:1). A control of GFP-H6 incubation with anti-CD44 is presented.

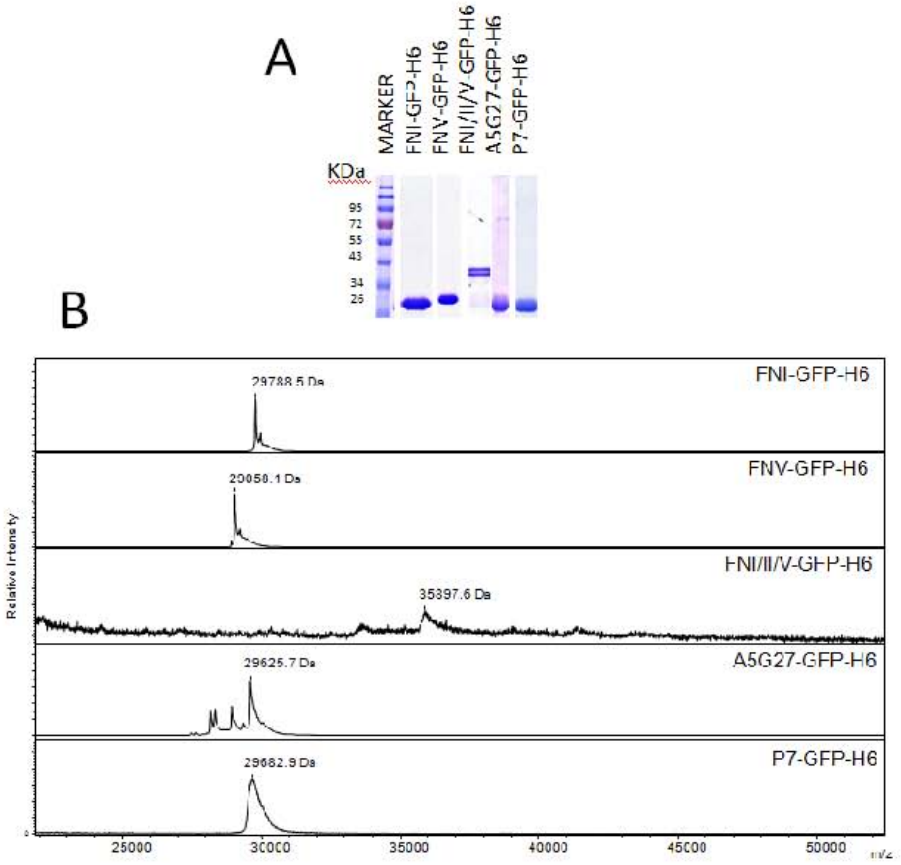
603

604 **Figure 6:** Modulation of nanoparticle internalization through CD44 regulation. A)
605 Enhanced cell entry (left) and fluorescence intensity (right) of A5G27-GFP-H6
606 incubation at 0.3 µM after 1 h of cell exposure to increasing amounts of FGF2 (ratios
607 1:1, 1:10 and 1:30) due to CD44 receptor overexpression mediated by FGF2-H6. B)
608 Reduced cell entry (left) and fluorescence intensity (right) of A5G27-GFP-H6 incubation
609 at 0.3 µM after 1 h of cell exposure to anti-IL10 antibody due to an inhibition of IL-10-
610 induced CD44 expression mediated by polyclonal anti-IL10.

FIGURE 1



COMPLEMENTARY FIGURE 1



Supplementary Figure 1: Characterization of recombinant protein constructs. A) Purified protein constructs visualized by Coomassie Brilliant Blue staining upon electrophoresis on SDS- polyacrylamide gel. B) MALDI-TOF MS analyses of the five constructs confirms protein integrity. A sample of 1 µl 1:1 (sample: DHAP matrix) mixture was applied on a ground steel plate. Samples were analysed using a lineal operation mode and 25kv acceleration voltage.

FIGURE 2

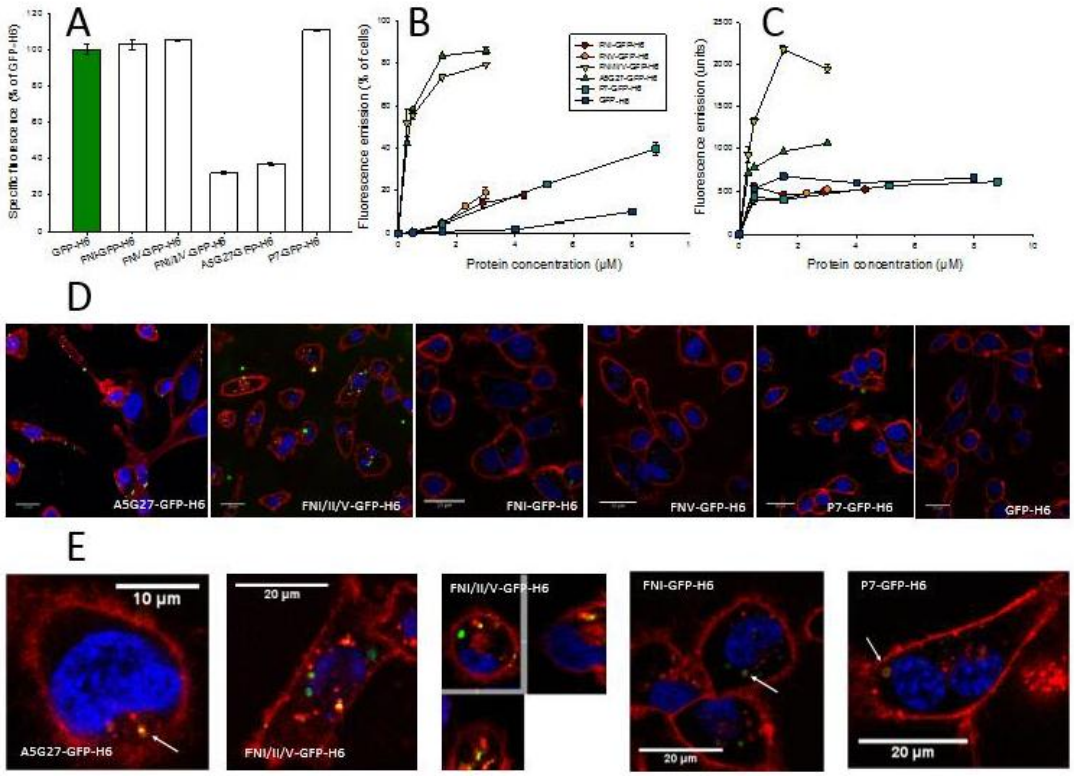


FIGURE 3

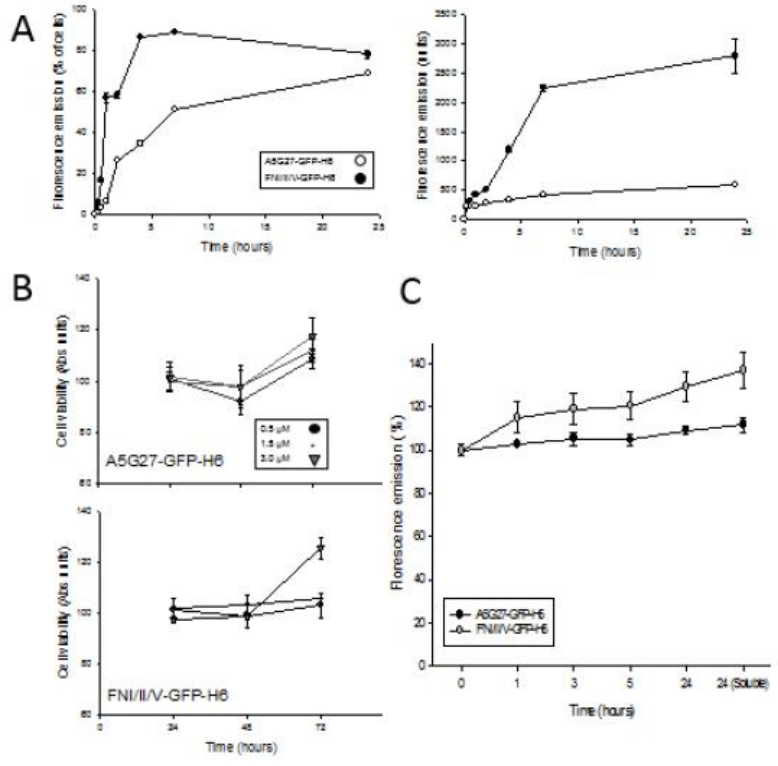


FIGURE 4

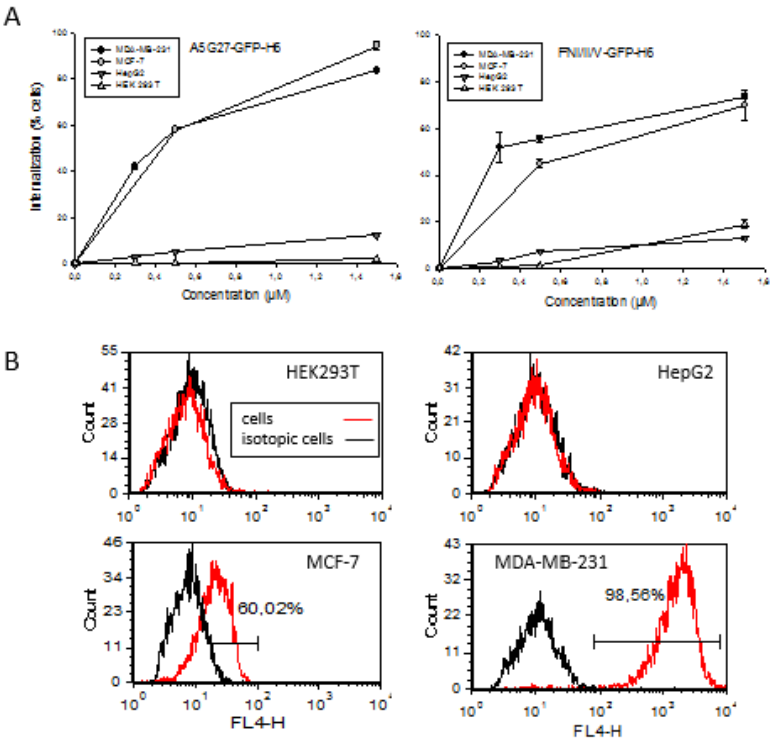


FIGURE 5

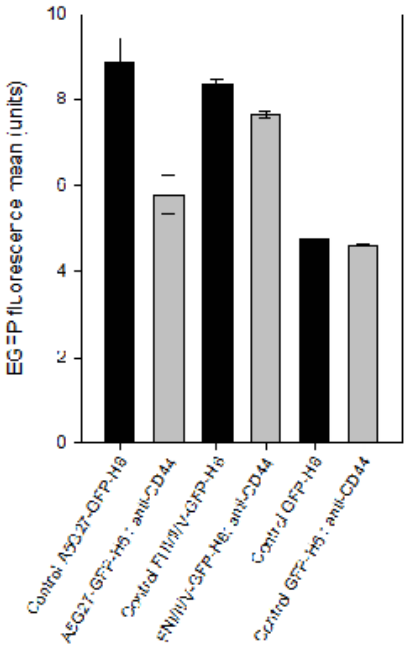
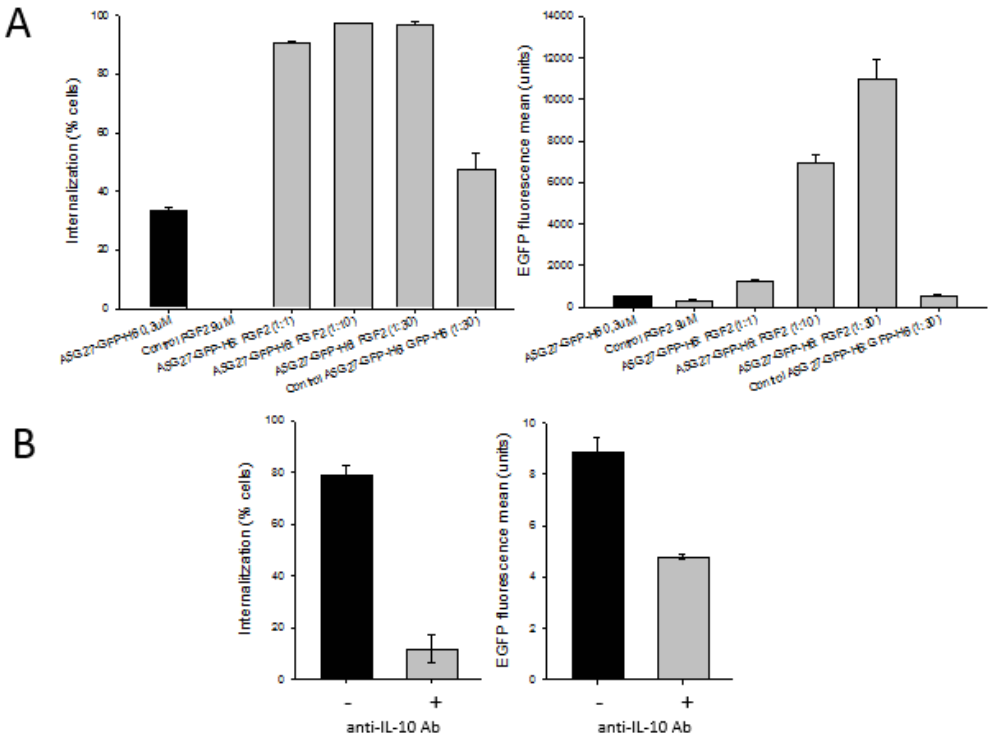


FIGURE 6


A
Internalization (% cells)
EGFP fluorescence mean (units)
ASG27-GFP-H6 0,3uM
Control RGF2 9uM
ASG27-GFP-H6 RGF2 (1:1)
ASG27-GFP-H6 RGF2 (1:10)
ASG27-GFP-H6 RGF2 (1:30)
Control ASG27-GFP-H6 GFP-H6 (1:30)
B
Internalization (% cells)
EGFP fluorescence mean (units)
-
+
anti-IL-10 Ab

ARTICLE 2

Conformational and functional variants of CD44-targeted protein nanoparticles bio-produced in bacteria

Pesarrodona M, Fernández Y, Foradada L, Sánchez-Chardi A, Conchillo-Solé O, Unzueta U, Xu Z, Roldán M, Villegas S, Ferrer-Miralles N, Schwartz S Jr, Rinas U, Daura X, Abasolo I, Vázquez E, Villaverde A.

Biofabrication. 2016 Apr 14; 8(2):025001

Protein-based nanoparticles are especially appealing in nanomedicine because their structure and functionality is adaptable via genetic engineering and their biological fabrication is scalable in recombinant cells. While some recombinant polypeptides are straightforward to obtain in solution, others, in particular those of non-bacterial origin, form insoluble aggregates called inclusion bodies. Another challenge is protein stability, and therefore activity, can be sensitive to production conditions and the purification strategies applied.

We have explored to what extent downstream purification strategy influences the structure and biological performance, *in vitro* and *in vivo*, of the two engineered CD44-targeted protein-only nanoparticles produced in *Escherichia coli*. As these proteins are produced as soluble species and also partially as inclusion bodies, this study compares: particles built by soluble protein species purified from the soluble cell fraction and protein versions obtained by *in vitro* extraction from inclusion bodies, applying non-denaturant solubilizing chaotropic agents. Comparative analysis includes nanoparticle architecture, structure conformation, cell internalisation and *in vivo* biodistribution.

Our data demonstrate that the choice of downstream procedure influences biological performance as well as the physicochemical properties of the material. Extracted protein from IBs, besides assembling into larger architectonically stable nanoparticles after purification, shows enhanced *in vitro* protein cell internalisation, compared to its soluble counterpart. Moreover it displays a different protein structure and *in vivo* distribution profile.

Since the following study continuing this work has not been still accepted for publication, it will be presented in the Annex I of this PhD thesis.

MANUSCRIPT 1

Intrinsic functional and morphometric variability of tumor-targeted GFP nanoparticles produced in bacteria

Hereafter a short abstract of this manuscript is presented, still I kindly request to look up the corresponding annex to find the whole text.

Other works connected with this thesis but that are not part of the results are presented in Annex II to VI.

MANUSCRIPT 1

Intrinsic functional and morphometric variability of tumor-targeted GFP nanoparticles produced in bacteria

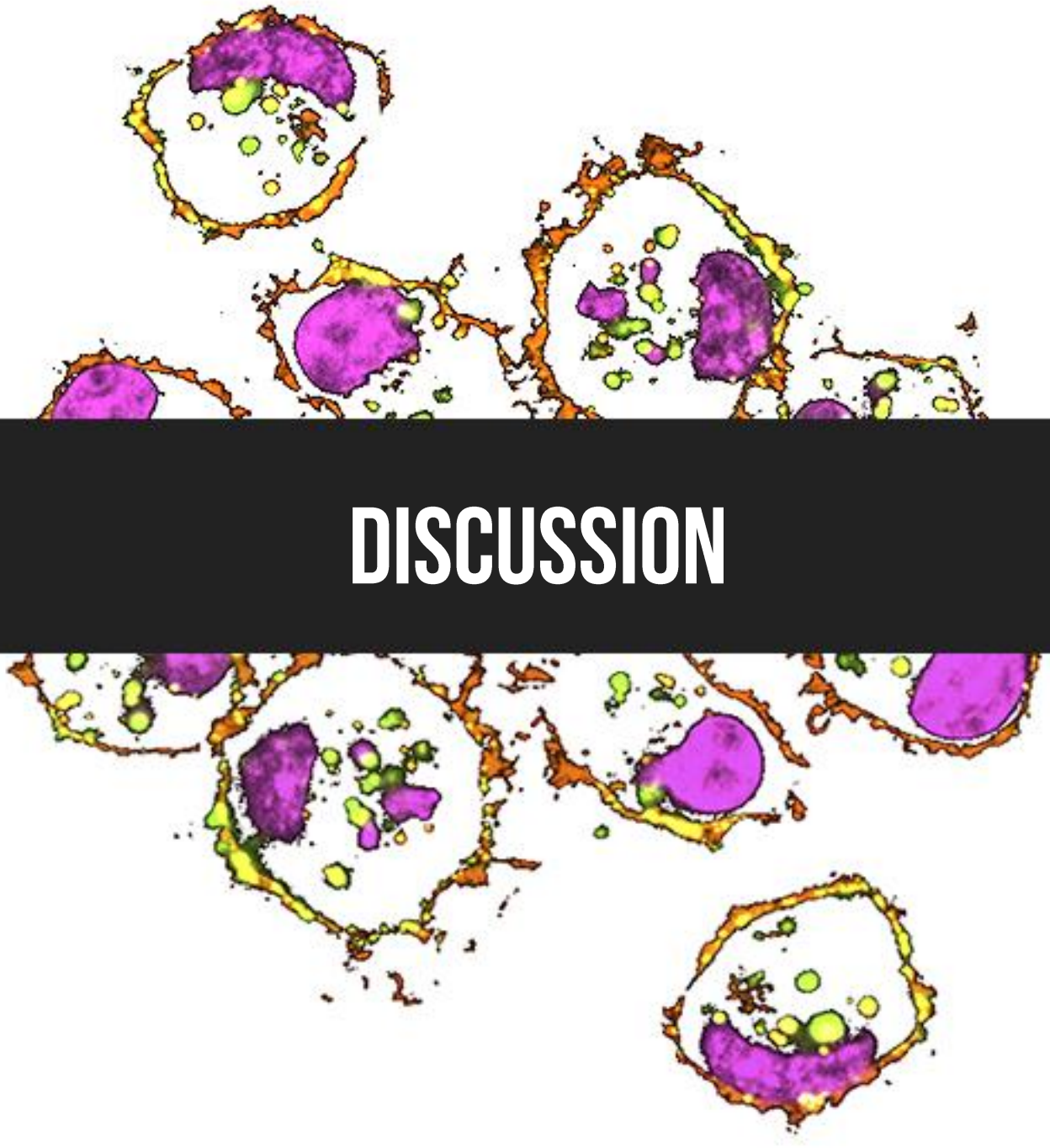
Pesarrodona M, Crosas E, Cubarsí R, Sánchez-Chardi A, Saccardo P, Unzueta U, Rueda F, Sanchez-García L, Serna N, Mangues R, Ferrer-Miralles N, Vázquez E, Villaverde A

Submitted to ACS Nano

Protein nanoparticles are attractive for their structural and functional versatility along with the cost-effective, rapid and bio-friendly biological production in recombinant living cells. In a previous study, presented on Annex IV of this thesis, we demonstrated that *de novo* designed self-assembling protein building-blocks acquire diverse supramolecular organisations depending on the originating cell source. Likewise, their structural and physicochemical variability lead to notable functional diversity. However, it is still poorly unknown how a cell factory with altered protein folding machinery impact on the ultimate configuration and function of smart protein nanoparticles.

This study has been focused on tumor-targeted protein nanoparticles fine separation and characterisation when produced in alternative *E. coli* strains with variant protein folding networks. Multiple analytical approaches including dynamic light scattering, small-angle X-ray scattering and field emission scanning electron microscopy have been applied to determine their architecture and geometry as well as evaluate their nano-bio interactions.

Results show that production yields discrete but regular populations of morphometric profiles. Oligomeric entities range from 2.4 to 28 nm and from spherical to rod-shaped geometries. The relative proportions of these architectonic variants where determined by features of the producing strain. In turn, contrasting physiochemical properties of these species were intimately influencing their performance with regard to fluorescence emission, cell penetrability and receptor specificity. Therefore, bacterial protein quality control system defines the ultimate geometry of designed GFP constructs and, by consequence, the functional output. This brings a benefit since nanoparticles with optimal properties could be analytically identified and isolated from producing cells for their ultimate goal.



DISCUSSION

Nano-scaled structures are ideal vehicles in drug delivery scenario due to the physical and biological properties exclusive to their size and bulk material [2]. Nanoparticles, extending above 7 nm, avoid renal filtration thus incrementing circulation time of the carried drug [143]. Furthermore, properties such as diffusion in the tissue and cell penetrability convert nanoparticles into a potent tool for competent systemic transport [144]. Next generation nanoparticles are designed to promote drug's active targeting to specific tissue in order to reduce side-effect by off-target organ accumulation [3]. In fact, the development of targeted nanoparticles is expanding with the increasing number of cell surface proteins identified as valuable markers [145]. However, achieving a full biocompatible material for the synthesis of nanoparticles remains a major challenge [146]. In this regard, protein-based nanoparticles are presented as a promising approach that benefit from non-toxicity, high biodegradability and functional versatility of proteins.

VLPs are an excellent example of the functional and structural tunability and stability of protein supramolecular structures. The biological properties that exhibit viruses during infection, such as stable systemic circulation, receptor targeting, cell internalisation, intracellular trafficking and accumulation into the appropriate compartment, are attempted to be reflected on designed nanoparticles [105]. In this regard, multifunctional proteins, in comparison to other protein-based constructs like BMCs or VLPs that have limited functional flexibility, can combine multiple domains from different origins conferring multiple functions into a single polypeptide that can be safely and cost-effectively produced in cell factories as recombinant proteins. Despite this potential, how to control protein-protein interactions between building-blocks to build protein nanoparticles with defined architectonic geometry has remained neglected [93,147].

Therefore, our group has lately proposed a button-up approach to obtain multifunctional self-assembling protein nanoparticles which consist on the addition of end-terminal cationic tags flanking a core protein. The building-blocks spontaneously organise into highly stable and regular nanoparticles driven by the electrostatic field of the polypeptide resulting in weak but multiple protein-protein interactions [38]. The potential of this *de novo* engineering approach was demonstrated with the tumor-targeting efficacy, *in vivo* stability and biocompatibility of the rationally designed T22-GFP-H6 protein building block. This self-assembling protein, empowering T22 as peptide ligand of CXCR4 marker, display excellent *in vitro* and *in vivo* targeted accumulating in primary tumor and metastasis on a colorectal cancer model [38].

In breast cancer, CD44 is widely used as biomarker for targeted nanoparticles because of its overexpression in CSC and its role in promoting angiogenesis, invasion and metastasis. Although some peptides are described as CD44 binding ligands, the natural ligand, HA, and anti-CD44 antibodies are the preferable choice when decorating carbon nanotubes, polymers, dendrimers or inorganic nanoparticles for breast cancer targeting [148,149]. Despite that HA show cell type binding dependence and antibodies have high molecular weight hindering nanoparticle's multivalency and are expensive to produce,

limited studies have investigated the targeting efficacy of CD44 peptide ligands displayed in protein-based nanoparticles.

In an attempt to exploit the abovementioned **protein nanoparticle formation approach**, the FIRST OBJECTIVE of this work was to engineer self-assembling protein nanoparticles **targeted to breast cancer** cells by incorporating peptide ligands binding CD44 receptor.

The results of this investigation show that, although the selected targeting peptides are well known as CD44 ligands, only the laminin, A5G27, and the fibronectin, FNI/II/V, peptides display cell internalisation which, at the same time, take part in the two polypeptides with the ability to self-assemble into regular nanoparticles. The other tested ligands remain unassembled likewise GFP-H6 protein control and display limited cell internalisation. This supports the fact that nanoscale structure formation presenting large multivalency is required over the monomeric polypeptide for cell internalisation. This is probably due to the presentation of multiple ligands on nanoparticulated material that allows multiple cross-linking at the cell surface and thus, favouring membrane wrapping. This could be also due to nanoparticle's size, which is described as critical property that influences cell internalisation [60]. This also applied to natural oligomers when displaying cell-binding peptides [106,150].

Regarding size and structure of the developed protein nanoparticles, A5G27 and FNI/II/V peptides trigger building-block oligomerisation into particles of around 14 nm of diameter (Article 1_Fig. 1B), a size that would enable them to escape from renal filtration. Toroid-like architecture, previously observed for nanoparticles presenting cationic peptide T22 (Annex III_Fig. 6 C) and R9 [110], has been observed for the CD44 ligand –empowered building blocks by AFM (Article 1_Fig. 1 C-G), indicating that the ring-shaped distribution is not limited to a unique type of end-terminal tags. Taking into consideration the previously reported relationship between the nanoparticle size and the number of cationic amino acids in the N-terminal tag sequence (Annex II_Fig. 3), nanoparticle formation could be anticipated for FNI/II/V peptide ligand with a total of 12 cationic amino acids. Indeed, the cationic nature of the CD44-binding peptides that remain unassembled is far below the number of cationic amino acids required (Article 1_Table 1). However, A5G27 peptide represents a particular case. Being poorly cationic and considering the regression analysis we previously performed (Annex II_Fig. 3), A5G27-empowered protein should be excluded from nanoparticle formation. Yet, regular ring-shaped nanoparticles are observed. Noticeably, we noted that the distribution (a part form the number) of cationic amino acids within the sequence is also affecting the protein-protein interactions needed for self-assembling into nanoparticles [151]. The particular case of A5G27 peptide opens the spectrum of possible self-assembly inducing peptides. Further research is required to establish which protein-protein interactions are driving the oligomerisation in this instance.

In vitro experiments with A5G27-GFP-H6 and FNI/II/V-GFP-H6 proteins confirm the stability and biocompatibility of protein nanoparticles in line with the previous results of T22-GFP-H6 nanoparticles [38]. Besides lacking cell toxicity at concentration above the optimal for internalisation (Article 1_Fig. 3B), the functional and structural stability in plasma and serum respectively (Article 1_ Fig. 3C and Article 2_ Fig. 8B-C) demonstrate the structural robustness of the resulting nano-scaled material.

Surprisingly, green fluorescence activity of the two developed protein-only nanoparticles is lessened around 30-40% compared to the unassembled proteins (Article 1_Fig. 2A). Although the possible interference of quenching effect cannot be ruled out, the impaired fluorescence activity is likely to be arisen from the compactness and conformational variation forced by self-assembling arrangement. This hypothesis has been supported in the third study of this thesis (Annex I_Fig. 2B and 5A), where a correlation between specific fluorescence of nanoparticles and the protein flexibility is detected. In a previous study, we have demonstrated that the nanoparticle formation strategy based on end-terminal cationic peptides as pleiotropic tags, so far worked with all tested core proteins being GFP, p53 (Annex II) and IRFP (Annex III). As observed in this presented study, the functionality in the final vehicle of the core protein (in this case, GFP), multiplies the number of possible multifunctional polypeptide designs and encourage their displacement for therapeutic peptides that would results in a therapeutic vehicle itself.

Regarding biological properties, A5G27-GFP-H6 and FNI/II/V-GFP-H6 nanoparticles show a 20-fold increase on cell internalisation compared to the other CD44-ligand GFP constructs that remain unassembled. Internalisation occurs in a dose-dependent manner reaching a plateau at 1,5 µM in 8 hours. Indeed, this results demonstrate the, until now unexplored, internalisation capacity of laminin and fibronectin peptides. Both nanoparticles' cellular entry occur through receptor mediated endocytosis by CD44 receptor and accumulate in the perinuclear and nuclear regions (Article 1_Fig. 2 D-E). CD44 specific internalisation was demonstrated by the coincidence between CD44 levels and penetrability in different cell lines (Article 1_Fig. 4). Furthermore cell internalisation is enhanced or diminished by the alternate regulation of CD44 using a positive regulator, FGF2, and a negative regulator, anti-IL10, respectively (Article 1_ Fig. 6).

Summarizing this first study, we have developed two protein-only nanoparticles that specifically bind CD44-receptor, namely A5G27-GFP-H6 and FNI/II/V-GFP-H6. These designed building-blocks promote the formation of ring-shaped structures which are approximately 14 nm in size, stable in plasma that can internalise cells through endocytosis in absence of cell toxicity. These targeted protein nanoparticles can therefore be chemically coupled to a conventional drug as has been already accepted in approved nanomedicines [152]. Alternatively, regarding the activity of the core protein (in these case GFP) in the resulting nanoparticles, therapeutic peptides can be incorporated or exchanged constituting a protein-only therapeutic vehicle for many medical purposes. Together with the stability and biodistribution of other tested multifunctional self-

assembling protein nanoparticles (Annex III), the CD44-targeted nanoparticles developed here are promising platforms for the transport of drugs or imaging agents as nanomedicines for breast cancer or other CD44-linked afflictions.

In the preceding lines we have discussed the biological performance *in vitro* of the generated CD44-targeted nanoparticles in a drug delivery scenario. However, protein nanoparticles are produced in living cells, biological systems that are more complex than chemical synthesis applied for the production of other material-based nanoparticles. In this regard, the other two studies that embody this thesis aim to understand how the production system influences nanoparticles' configuration and the derivative biological performance.

Besides protein-based nanoparticles, polymers [153] or metallic nanoparticles [154] for biomedical application are synthetized in diverse microbial cell factories. In fact, the production of nano-scaled materials through microbial synthesis is increasingly in demand over conventional chemical procedure owing to the adaptability, scalability, cost-effectiveness and eco-friendly nature of biological production [155,156]. Despite the broad availability of recombinant systems, the wide knowledge acquired and characteristics such as simplicity and rapid growth of *Escherichia coli* convert it on the cell factory per excellence [112], especially since the development of LPS-free strains.

Most of the recombinant proteins, notably for chimeric proteins, are accumulated in IBs after inefficient protein folding. Being an almost pure source of recombinant protein, IBs became an appealing protein source and many pharmaceutical protein products are obtained from them. Therefore, protein solubilizing methods, alternative to the tedious refolding procedure, have been developed to extract active protein from these aggregates. In this regard, the AIM OF THE SECOND STUDY was to investigate how the **protein origin** and purification method **influence** the **organisation and bio-performance** of smart and structurally complex **protein materials**. Previously developed CD44-targeted nanoparticles are excellent samples for this objective since these proteins are produced as soluble versions but also accumulate in IBs.

Concerning product degradation during production, A5G27-GFP-H6 and FNI/II/V-GFP-H6 proteins (Article 2_Fig. 1B and 2B-C) supported the observation that protein embedded within the IBs is protected from bacterial proteases compared to the soluble counterpart [157]. This would initially represent an advantage for IBs as protein material source of choice.

Interestingly, structural and functional differences are detected when evaluating the nanoparticles successfully purified from the soluble fraction or resolubilised from IB. Larger nanoparticles (30 nm) with higher oligomeric status are observed when extracted from the insoluble fraction (Article 2_Fig. A-C) and fluorescence activity yields half the intensity of the plain soluble species (Article 2_Fig. 3E). Surprisingly, difference on cell penetrability of the nanoparticles from both origins is enhanced or diminished depending on the model protein (Article 2_Fig. 4 and 5). Accordingly, we selected A5G27-

GFP-H6 protein, which exhibit better performance to proceed for further investigation in order to comprehend those events.

The similar values regarding lipids and carbohydrates content (Article 2_Fig. 6A) or nanoparticles' net surface charge between proteins obtained from the soluble fraction (A5G27-GFP-H6 sol) and the ones resolubilised from IBs (A5G27-GFP-H6 rIB) (Article 2_Fig. 3D), let us to discard a putative role of contaminants or the detergent used for extraction that could be interfering on the differential behavior of the material (Article 2_Fig. 6B). Therefore, considering the differential size and specific fluorescence, the fact that A5G27-GFP-H6 sol and A5G27-GFP-H6 rIB adopt different structural conformations seems plausible. Through in silico analysis, residues that are involved in the modulation of GFP fluorescence emission intensity were detected in a predicted protein-protein interaction patch (Article 2_Fig. 9B). Here, fluorescence is lessened for the nanoparticulated material derived from IBs; however, fluorescence of monomeric GFP protein resolubilised from IB in previous studies was unchanged [158]. Thus, given the abovementioned data, we suggest that the functional difference derives from an altered assembly pattern of the building blocks rather than a fine conformational change upon polypeptide folding. This idea is further supported by the contrasting behaviour regarding intrinsic fluorescence intensity observed upon thermal denaturation (Article 2_Fig. 7B-C). Conventional A5G27-GFP-H6 nanoparticles tend to disassemble or unfold at increasing temperatures whereas nanoparticles extracted from IBs display a compactness or aggregation tendency. Such behaviour on supramolecular protein complexes upon IBs extraction have been later observed for a homotetrameric enzyme, beta-galactosidase (Dr. Julieta Sánchez personal communication). One possible explanation for this might be that fine conformational changes on monomers (irrelevant for green fluorescent activity) and the different molecular environment in the IB structure favour alternative arrangements of the building-blocks (Article 2_Fig. 9A). These nanoparticles would be assembled into the higher porous and hydrated IBs [159]. Alternative conformations would affect the spatial accessibility of the targeting peptide by limiting or enhancing ligand exposure as occur on FNI/II/V-GFP-H6 and A5G27-GFP-H6 proteins respectively and therefore influence the nanoparticle's performance (Article 2_Fig. 4). As both protein-nanoparticles resolubilised from IB show a size of around 30 nm, the 6-fold increase in cell penetrability for A5G27-GFP-H6 extracted from IBs compared to its soluble counterpart (Article 2_Fig. 4A) would be derived from an increase on nanoparticle multivalency rather than from size issue considering the inhibited performance of FNI/II/V-GFP-H6 rIB material.

Both materials show excellent stability in working buffer; however aggregation tendency is detected for nanoparticles obtained from IB when incubated in physiologic medium, note the peak shift at larger sizes by DLS (Article 2_Fig. 8B). After centrifugation (Article 2_Fig. 8C), fluorescence is unchanged meaning that the resulting aggregates constitute soluble clusters.

Interestingly, differences observed between nanoparticles obtained from cellular soluble fraction or extracted from IBs regarding structure and cellular internalisation, are extended at systems level. Biodistribution of A5G27-GFP-H6 rIB and A5G27-GFP-H6 sol nanoparticles display different side accumulation in non-target organs (Article 2_ Fig. 8A). Nanoparticles obtained from IBs tend to accumulate in organs of the reticuloendothelial system, such as liver and lung. Protein clustering in physiological medium, probably favoured by physicochemical properties or by an increase on aggregation-prone regions, provokes enlargement of particles to hundreds nanometers (Article 2_ Fig. 8B) which easily accumulate in these organs [53]. A possible interference of the aggregation-prone material with protein corona cannot be ruled out. The prevalence in liver might be enlarged by the increase on hydrophobicity of IB's extracted nanoparticles [20,69]. These results seem to be consistent with the limited accumulation in kidney compared to the soluble counterpart, since the enlarged particles are preserved from renal filtration.

The great cell internalisation in target cells observed *in vitro* for the developed CD44-targeted nanoparticles and the increase on cell penetrability in presence of FGF2 (a growth factor overexpressed in tumor tissue [160]) (Article 1_Fig. 6A) suggested that these nanoparticles would show an excellent penetration within the tumor. Noteworthy, however, biodistribution in a breast cancer model (Article 2_Fig. 8A) was insufficiently targeted to the cancer site. Two facts could explain that: in one hand, due to physicochemical characteristics that lead to rapid accumulation in other organs or; on the other hand, could be due to an expression of CD44 receptor in other tissues that hampers the accessibility of these nanoparticles to the tumor. The former wouldn't challenge the targeting capacity of the developed nanoparticles but the selected biomarker instead. Nevertheless, during biodistribution, the administration on any of the tested nanoparticles didn't cause to any undesirable side effect, providing additional evidence with respect to the biocompatibility of this engineered materials.

To sum-up the second study of this thesis, we have identified altered supramolecular organisation and consequently distinct *in vitro* and *in vivo* performance of cell-targeted protein nanoparticles. These differences have been identified to be intrinsically related to the material's origin and not to the separation method used. Thus, two important points that are of major concern are gathered in this study. First, protein architecture and functionality are highly influenced depending on whether the materials derive from the *E. coli* soluble fraction or extracted from IBs. It is a particularly important aspect regarding recombinant production of smart and complex protein structures that, until now, has not been addressed. Secondly, the differed building-block organisation into nanoparticles showed a huge impact on the whole biological performance, and this is why its characterisation is crucial when applying nanoparticles for biomedical applications.

Protein nanoparticles settled in soluble or insoluble cell fraction from a particular *E. coli* strain differ from the physical and biological point of view. As protein quality depends

on the genetic traits of the host in terms of folding modulators, in a recent study we characterised tumor-homing protein nanoparticles produced in a set of strains with different genetic background. Results indicate that nanoparticles produced in *E. coli* K-12 strains with knockout of folding modulators show modified properties not only at molecular level but also at their macroscopic performance (Annex IV). In line with this study, production of this smart nanomaterial in endotoxin-free *E. coli* strain (KPM335), with altered genetic background, was assessed. Resulting nanoparticles did not compromise the biodistribution profile but presented cell penetrability and structural changes (Annex V). Surprisingly, conformational variants were not only observed among explored bacterial strains (Annex V_ Fig. 8A) but also within a particular strain. Upon protein affinity purification, two separate fractions (IMAC fraction 1 and 2) were eluted that correspond to poorly or efficiently functioning particles respectively. This combination of findings indicates that bacterial cells can fabricate GFP-based nanoparticles of biomedical interest in alternative conformations.

Therefore the THIRD STUDY of this thesis was set out with the AIM of understanding the **supramolecular conformation** of tumor-targeted nanoparticles produced in **bacterial cell factories with different genetic background** and determining the **influence** of structure on their **biological function**.

First, upon fine separation of protein material resulting from bioproduction, we observed multiple peaks corresponding to different oligomerisation states from P1 to P5 (depicted in colors in Annex I_Fig. 1) that, until now, remained unexplored. Eluted protein ranged from 5 nm (unassembled protein, P5) to 25 nm (nanoparticles, P1) (Annex I_Table 1). Observed variants were eluted in both IMAC fractions, however, while small oligomers were mainly present in IMAC fraction 1, larger structures were accumulated in IMAC fraction 2 (Annex I_Fig. 1). The most striking observation to emerge from the size-exclusion chromatography is that dimensions of each oligomer population are highly consistent comparing bacterial strains and tumor-targeted protein nanoparticles, indicative of the regularity in building-block self-assembling. The main difference remains on the relative proportions of the protein populations comparing bacterial strains. Besides different sizes that would correspond to a spectrum of oligomeric assemblies, these populations display distinguishable structural properties such as shape and protein flexibility as well as altered functionality regarding fluorescence and cell penetrability. These findings suggest that the different performance of the pooled material examined in previous studies (Annex IV and V) would result from the inter-strain irregular proportions of structural and functionally diverse oligomeric variants.

Concerning morphometry, GFP oligomers reveal a broad range of shapes (Annex I_Fig. 2A). Results are corroborated in parallel by electron microscopy (Annex I_ Fig 3) and SAXS data (Supplementary_Fig. 3). Larger nanoparticles range from oblate and spherical shapes while unassembled versions display rod form. These smaller and more elongated protein particles adopt a less compact conformation that confers an extended degree of protein flexibility (Annex I_Fig. 2B). Moreover they exhibit high fluorescence values. In fact, a

negative correlation between fluorescence and size of the nanoparticles is drawn (Annex I_Fig. 5A). This data supports alternative conformations and protein-protein interactions with GFP fluorescence modulation patches that might limit or favour different oligomeric organisations.

Regarding internalisation, a positive correlation between nanoparticle size and cell entry is observed. Large nanoparticles with oblate/sphere shape efficiently internalise target cells compared to poorly penetrating more elongated, unassembled version (Annex I_ Fig. 4). This data supports that size is the principal physicochemical characteristic, together with multivalency, affecting nanoparticles performance with respect to receptor-mediated internalisation. Therefore, the mere presence of a cell-ligand is not sufficient to assist internalisation. This is in agreement with previous data comparing *in vitro* and *in vivo* performance of assembled and unassembled versions of IRFP-based protein nanoparticles [161].

Noteworthy, protein populations from IMAC fraction 1, although displaying size/penetrability correlation, internalisation levels are significantly weakened compared to its close related assemblies from fraction 2. This observation, in line with previous results using the material pool from Rueda *et al* (Annex IV_ Fig. 2D), has been also repeated for FNI/II/V-GFP-H6 nanoparticles extracted from IB in the second study of this thesis which, despite displaying larger size, internalisation is not enhanced compared to A5G27-GFP-H6 nanoparticles (Article 2_Fig. 4A). This combination of findings provides some support for the conceptual premise that inefficient internalisation would be strongly related to an alternative assembling conformation that limits ligand exposure for active targeting purposes. A lessened exposure of His-tag and cationic-ligand that result on loose receptor interaction was previously suggested based on the poor penetrability and the weak nickel affinity of abovementioned IMAC fraction (Annex IV_Fig. 2D).

Although internalisation of T22-GFP-H6 oligomeric variants was highly specific for CXCR4 receptor, specificity appeared to be influenced by size and protein compactness (Annex I_ Fig. 2B and 4A). Overall, specificity of flexible assemblies corresponding to P4 and P5 plain structures is increased compared to highly compact, larger nanoparticles. Contrary, endosomal escape capacity is enhanced with oligomers complexity probably because of a proton cumulative effect with the number of His-tag enrolled. In the case of P1, the non-significant influence of chloroquine could also suggest an alternative internalisation route which doesn't lead to late lysosome degradation (Annex I_Fig. 4B). Given the robustness of the GFP assembling pattern and the high yield of production in the bacterial strains, it is possible to separate and select the population that meet the optimal biological properties. In this particular case, in which we aim to develop targeted nanoparticles as future drug delivery agents, spherical oligomers from population P2 show the highest internalisation with the best receptor specificity probably by encountering optimal size and ligand multivalency (Annex I_Fig. 5B).

Elucidating the parameters of nanoparticles that individually or in combination enable their correct cell targeting, biodistribution and biocompatibility is a pivotal issue [59]. The impact of nanoparticles' morphometry on their exhibition in biological systems regarding organ specificity [162-164], biodistribution [165,166], toxicity [167] and cellular uptake [54,168,169] depends on the bulk material. So far, it has been addressed in several types of nanostructures but protein-based nanoparticles remained neglected. Here, the impact of nanoparticles' geometry on their functionality is described for a sort of tumor-targeted protein nanoparticles.

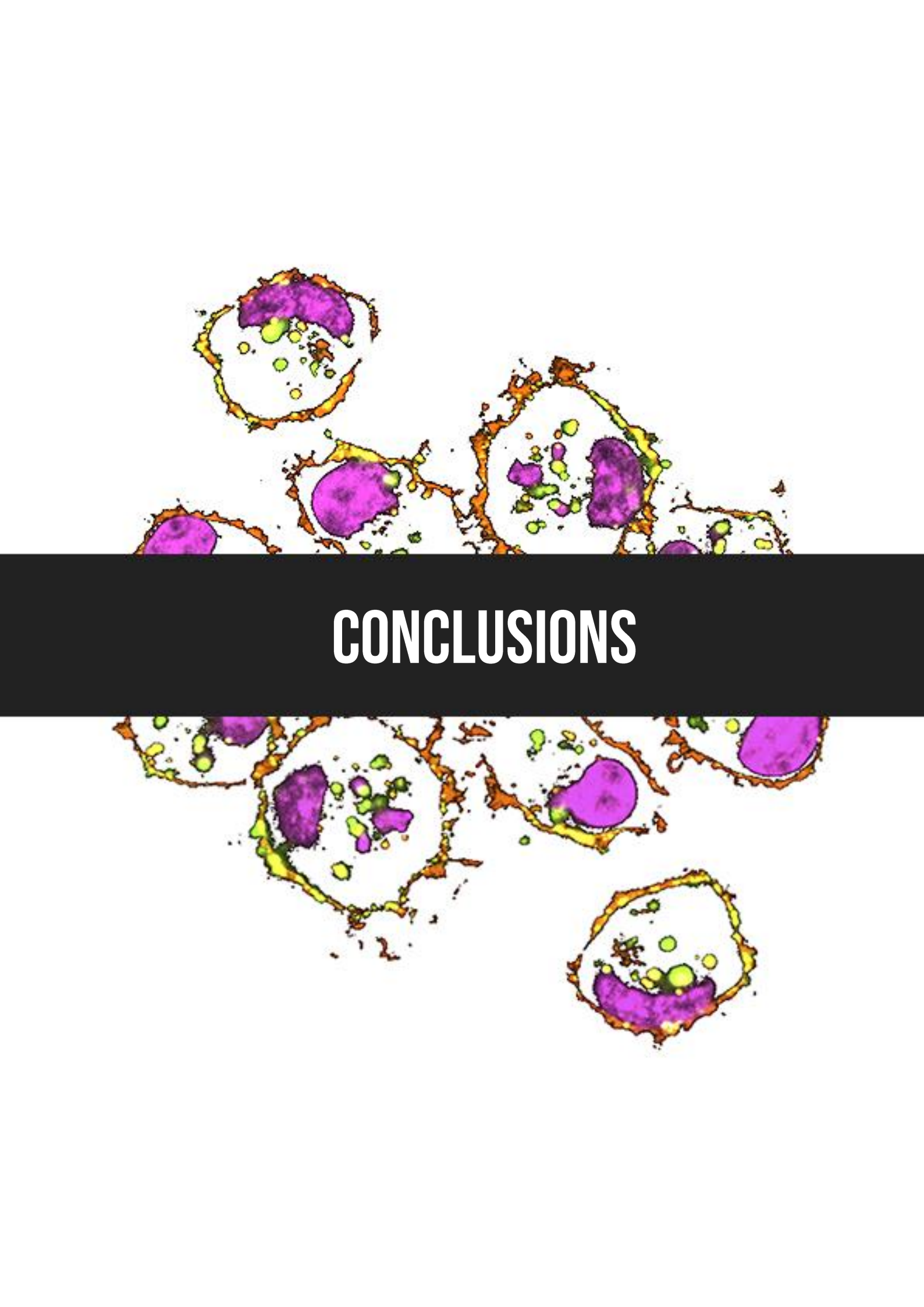
This study illustrates for the first time the robustness of self-assembling patterns among tumor-targeted GFP nanoparticles and also the influence of genetic background of the producing cells on shifting the distribution of oligomeric population in the pooled material. So far, the poor understanding on how quality control modulators impact on the fine protein conformation hampers to control or predict the assembling pattern and functional profile of a protein nanoparticle produced in a specific bacterial strain. However, upon production and fine separation of oligomeric populations is possible to characterise them and downstream select the most suitable one for a given application.

As a global summary, the presented studies had been focused on the application and characterisation of self-assembling protein nanoparticles developed through building block interaction by means of polyhistidines and cationic peptide tags. The fact that self-assembly is not limited to a unique core protein or specific peptide stretch converts it into a highly versatile approach. Here, this approach has been successfully extended for the development of CD44-targeted protein nanoparticles. The absence of toxicity and stability of the engineered vehicles strengthen their biocompatibility at cellular and systemic level. Compared to the efficient biodistribution of T22-GFP-H6 nanoparticles, the developed A5G27-GFP-H6 nanoparticles are rather accumulated in off-target organs probably derived from an unspecific systemic CD44 expression. This enforces the importance of target selection when designing active-targeted nanoparticles. Far from being effective as potential drug delivery nanomedicines for cancer, these targeted, self-assembling protein nanoparticles can be applied for the treatment of other illnesses or other applications such as imaging.

Besides microbial cell rapid and reproducible biological production of protein-based nanoparticles, here we have investigated the flexibility of these bio-factories. Alternative supramolecular organisations and therefore divergent interactions with biological systems have been described depending on the bacterial material source. Although protein from soluble cell fraction or extracted from bacterial IBs may, in general, serve indistinctly, fine conformational changes can have a huge impact on complex peptide systems such as smart protein-nanoparticles whose ultimate function depends on building block assembly.

Furthermore, the conformational and functional variability has been detected within a particular producing bacterial strain. A spectrum of morphometric variants of tumor-

targeted GFP nanoparticles is observed and identified as robust oligomeric populations. Each variant percentage within the crude material is shifted between bacterial strains with different genetic background indicating that the protein quality control somehow influences on the assembling pattern. These morphometric variants with contrasting physical properties such as size, shape and compactness exhibit distinguishable biological properties including fluorescence intensity, cell internalisation and receptor specificity. The intrinsic variability derived from the bioproduction, although being unpredictable or uncontrollable, offers a scope of nanoparticulated material from where optimal population can be selected considering their ultimate application.



CONCLUSIONS

1. Two CD44 ligands, namely A5G27 and FNI/II/V, when coupled to GFP-H6 polypeptide, drive the formation of regular nanoparticles.

2. Considering the cationic nature of the tested ligands and their ability to induce building-block self-assembly, all CD44-ligands fit on the previous mathematical prediction that relates amino terminal cationic charge and particle size, except for A5G27.

3. The cell internalisation capacity of A5G27 and FNI/II/V peptides ligands was demonstrated. A5G27-GFP-H6 and FNI/II/V-GFP-H6 nanoparticles internalise specifically through CD44 receptor via endocytosis.

4. The mere presence of a cell-ligand might not be sufficient to assist internalisation but multivalency and protein assembly in nanoparticles may be required.

5. All studied targeted self-assembling protein nanoparticles along this thesis have been proven not to be toxic *in vitro* or *in vivo* upon systemic administration in mice.

6. Efficiency of A5G27 ligand targeting CD44 *in vivo* is limited owing to non-targeted organ accumulation lead by nanoparticle physicochemical features or due to ubiquitous expression of the receptor

7. Non-denaturing solubilisation agents can be successfully applied for the extraction of smart complex protein nanoparticles from IBs.

8. Nanoparticles obtained from IBs compare with nanoparticles from soluble cell fraction show different supramolecular organisation and biological performance.

9. *In vitro* internalisation of nanoparticle conformational variants varies depending on the building block probably as a consequence of ligand exposure modification.

10. Protein nanoparticles extracted from IB display altered physicochemical characteristics compared to the soluble counterpart that:

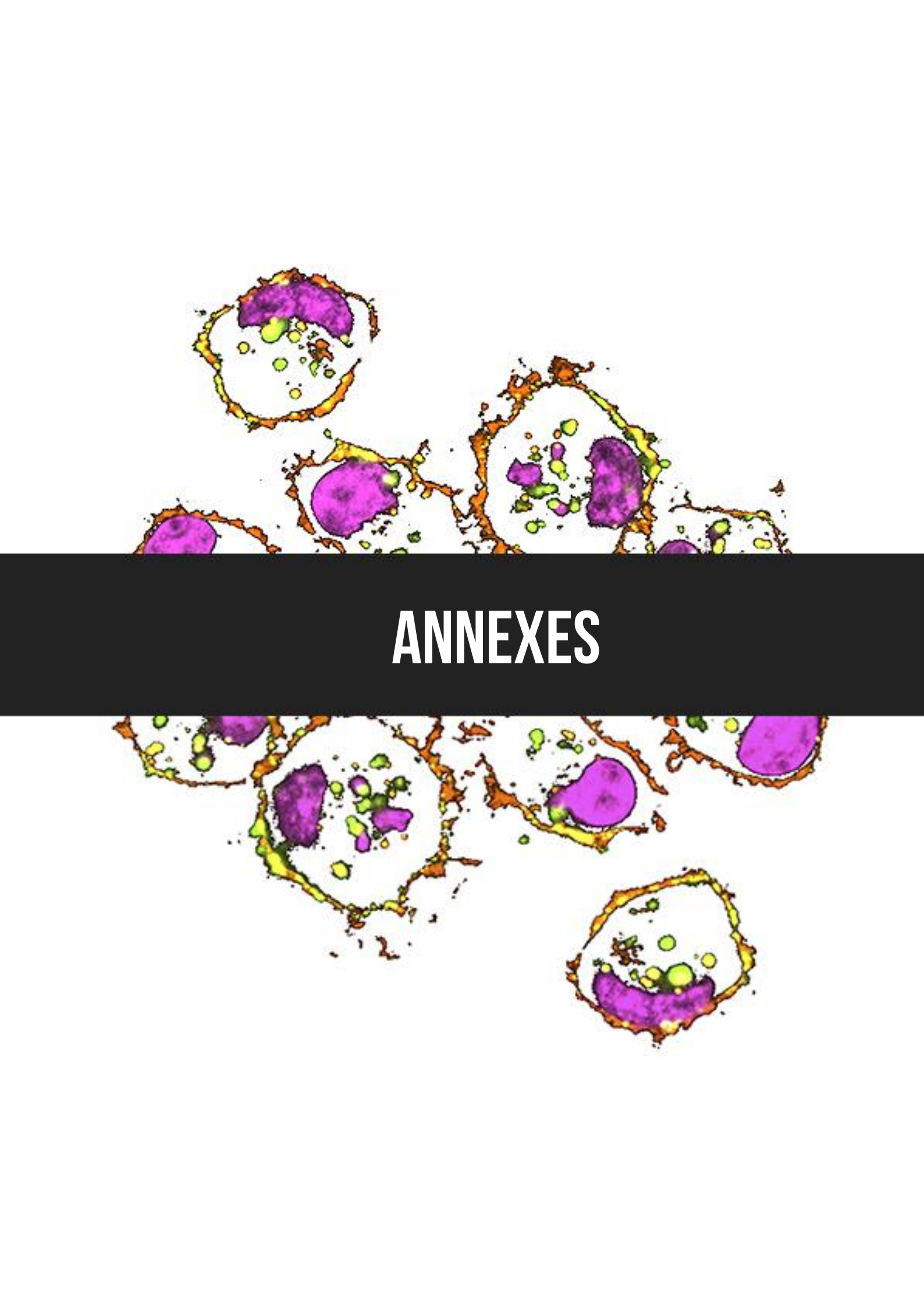
    a. Increase the aggregation tendency in physiological media and plasma in form of soluble aggregates
    b. Lead to accumulation in reticuloendothelial system (liver and lung) upon systemic administration

11. A spectrum of discrete morphometric and functional variants of tumor-targeted GFP nanoparticles is produced in microbial cell factories. Physicochemical and functional features of oligomeric populations are robust regardless the cationic ligand sequence or the producing bacterial strain.

12. The genetic background regarding protein folding machinery favours or lessens the proportion of each population. This explains the diverse performance of pool material produced in different *E.coli* strains.

13. Internalisation of these variants is proportional to oligomers size and multivalency. Assembly compactness presents a negative correlation with GFP fluorescence intensity and receptor specificity.

14. Production of GFP constructs with architectonic peptide tags in *E. coli* strains results in a wide range of conformational and functional species that cannot be controlled but allows downstream selection of optimal functional variants. In this instance of tumor-targeted protein nanoparticles, spherical oligomers of around 13 nm display optimal internalisation and receptor specificity.



ANNEXES

ANNEX I

MANUSCRIPT 1

Intrinsic functional and morphometric variability of tumor-targeted GFP nanoparticles produced in bacteria

Pesarrodona M, Crosas E, Cubarsí R, Sánchez-Chardi A, Saccardo P, Unzueta U, Rueda F, Sanchez-García L, Serna N, Mangues R, Ferrer-Miralles N, Vázquez E, Villaverde A

Submitted to ACS Nano

[1] Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, 08193 Cerdanyola del Vallès, Spain.
[2] Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Bellaterra, 08193 Cerdanyola del Vallès, Spain.
[3] CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, 08193 Cerdanyola del Vallès, Spain.
[4] ALBA Synchrotron, Carrer de la llum, 2-26, 08290 Cerdanyola del Vallès, Barcelona, Spain
[5] Department of Mathematics, Campus Diagonal Sud, Edifici U, Universitat Politècnica de Catalunya, Carrer de Pau Gargallo, 5, 08028 Barcelona, Spain
[6]Servei de Microscòpia, Universitat Autònoma de Barcelona, Bellaterra 08193 Cerdanyola del Vallès, Barcelona, Spain.
[7] Biomedical Research Institute Sant Pau (IIB-Sant Pau) and Josep Carreras Leukemia Research Institute, Hospital de la Santa Creu I Sant Pau, 08025 Barcelona, Spain.

ABSTRACT

Self-assembling proteins are gaining interest as building blocks of application-tailored nanoscale materials. This is mostly due to biocompatibility, biodegradability, and functional and structural versatility of peptide chains, regulatable through their amino acid sequence. Such potential for adaptability is particularly high in the case of recombinant proteins, which produced in living cells are suited for genetic engineering. However, how the cell factory itself and the particular protein folding machinery influence architecture and function of the final material is still poorly explored. In this study we have used diverse analytic approaches, including small-angle X-ray scattering (SAXS) and field emission scanning electron microscopy (FESEM) to determine the fine architecture and geometry of recombinant, tumor-targeted protein nanoparticles of interest as drug carriers, constructed on a GFP-based modular scheme. A set of related oligomers were produced in alternative Escherichia coli strains with variant protein folding networks. This resulted into highly regular populations of morphometric types, ranging from 2.4 to 28 nm and from spherical to rod-shaped materials. These differential geometric species, whose relative proportions were determined by features of the producing strain, were found associated to particular fluorescence emission, cell penetrability and receptor specificity profiles. Then, nanoparticles with optimal properties could be analytically identified and further isolated from producing cells for use. The cell's protein folding machinery greatly modulates the final geometry reached by the constructs, which in turn defines key parameters and biological performance of the material.

INTRODUCTION

The controlled self-assembling of synthetic peptides and recombinant proteins is a powerful tool in the generation of functional, micro and nanostructured materials. Recombinant proteins benefit, over synthetic peptides, from the versatility of biological fabrication. Based on a generic set of genetic engineering procedures, recombinant protein production has largely impacted on biotechnological and biopharmaceutical industries, with more than 400 protein drugs approved for human use [1]. The identification [2-3] and exploitation [4-5] of oligomerisation domains, the tailored fibrillation of amyloidal protein forms [6] and the de novo design of protein-protein interacting patches [7-8] offer a wide spectrum of possibilities regarding the generation of supramolecular materials to be used in biological interfaces [9-11]. Being functional but also biocompatible and biodegradable, protein materials show a still unexplored biomedical potential in both regenerative medicine and conventional or cell-targeted drug delivery [12-13]. The natural tendency of GFP to oligomerise [14] and the more recent manipulation of GFP assembling [15-16] have attracted interest as this beta-sheet rich protein represents a compact, structurally stable building block for the assay of controlled oligomer formation and material characterisation.

In previous studies, we have developed a protein engineering platform to promote the self-assembly of modular GFP constructs, based on the combination of end-terminal cationic stretches and polyhistidines [17-18]. Driven by electrostatic interactions and with a strong involvement of the histidine-rich tail, these peptides promote the formation of stable oligomers of defined average size in the nanoscale irrespective of the amino acid sequence and origin of the core protein placed in between. The resulting nanoparticles, with a toroid-like shape and usually ranging between 12 and 40 nm, are full stable *in vivo* [19] and are able to escape from renal clearance. When displaying appropriate peptide ligands of cell surface cancer markers CXCR4 or CD44 (T22 and A5G27 respectively) they specifically accumulate in primary tumor and metastasis in colorectal and mammary cancer models respectively [20-21], being suited as antitumoral drug carriers. The same platform has been used to construct fluorescent nanoparticles that cross the blood-brain barrier and target the brain [18].

Recently [22], we have determined that self-assembled T22-GFP-H6 oligomers elute from Immobilized Metal Affinity Chromatography (IMAC) in two separate fractions with slightly different sizes and CXCR4+ cell penetrability, suggesting alternative solvent exposure of both the His tag and the T22 tumor homing peptide. Also, the bacterial species and strain used for production influence the size and biodistribution of the material upon systemic administration in animal models of CXCR4+ colorectal cancer [23-24].

Altogether, these data indicate that bacterial cells can fabricate GFP-based nanoparticles of biomedical interest in alternative conformations. Such a possibility could be relevant to the *in vivo* use of these materials in a therapeutic context but in general, to the

production of self-assembling protein materials with specialized functions. As this suspected architectonic deviation has been so far elusive, we have examined here the functional traits as well as the subunit organisation of closely related, GFP-based nanoparticles produced in alternative Escherichia coli strains, specially focusing to intrinsic functional and morphometric variability resulting from the biofabrication process. We demonstrate, for the first time, subtle size- and shape-dependent heterogeneity of protein nanoparticles linked to their functional properties, which determines the performance of the materials as intracellular, cell-targeted vehicles. The identified segregation allows the selection, by their geometry, of specific oligomer populations in which receptor-specificity and cancer cell uptake are dramatically improved.

RESULTS AND DISCUSSION

The de novo designed A5G27-GFP-H6 and T22-GFP-H6 nanoparticles show, by dynamic light scattering (DLS), average size peaks of 14 and 12 nm respectively when produced in the conventional *E. coli* strains BL21 (DE3) and Origami B respectively [19-20] (Supplementary Figure 1). Any intrinsic morphometric heterogeneity, if existing, has been so far unobserved and eclipsed in the analysis of the raw material. However, when T22-GFP-H6 is purified from bacterial cell extracts by IMAC, it is eluted in two separated fractions [22], a fact that indicates alternative solvent exposures of the histidine-rich terminal tail. In this context, and to examine in detail the potential intrinsic heterogeneity in the architecture of these materials, we have screened by SEC the intra-strain size spectrum of T22-GFP-H6 nanoparticles occurring in these two separate IMAC fractions, upon production in different *E. coli* strains.

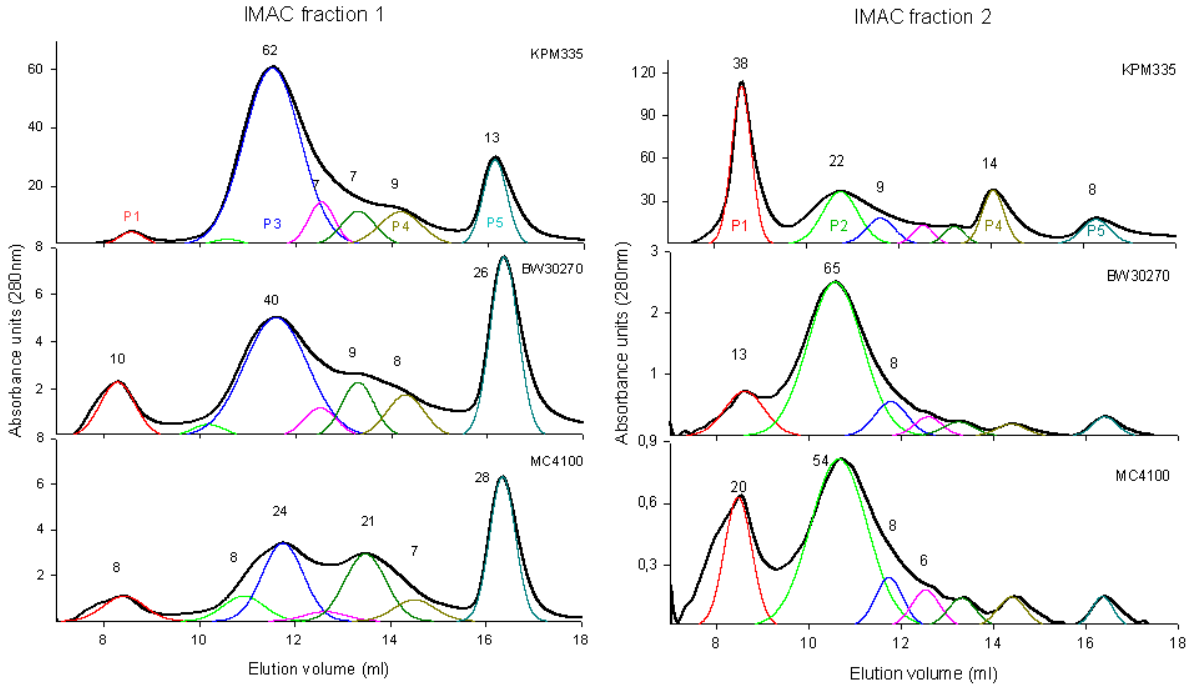


Figure 1. MORPHOMETRIC VARIABILITY OF T22-GFP-H6. SEC chromatograms of T22-GFP-H6 protein produced in KPM335, BW30270 or MC4100 and present on fraction 1 (left, low affinity) or fraction 2 (right, high affinity) of IMAC purification. Black lines indicate the average plot of the elution process. Numbers represent the percentage of protein amount in each oligomer population (only for nanoparticles represented over 5%), and line colours indicate regularly appearing peaks. Most representative oligomeric populations produced in KPM335 strain are identified as P1-P5 for further evaluation. A SEC calibration curve is depicted in the Supplementary Figure 2..

We included in the analysis the endotoxin-free *E. coli* strain KPM335, that is particularly interesting not only because of its interest in biomedicine as an endotoxin-free cell factory [25] but also because of the complex genetic modifications performed to remove endotoxic components from the cell wall. Such manipulation has resulted in the constitutive up or down regulation of several heat-shock genes involved in the quality control of recombinant proteins [23]. As observed (Figure 1), the resulting nanoparticles were distributed in numerous peaks, contrasting with the apparent morphometric

homogeneity determined by DLS over the pooled material (Supplementary Figure 1). The sizes of nanoparticle populations were highly coincident when comparing bacterial strains, and even when comparing IMAC fractions 1 and 2, indicative of high regularity in the oligomerisation states of GFP. In this regard, the differences observed were mostly lying on the relative proportions of these populations. For instance, T22-GFP-H6 oligomers from IMAC fraction 1 accumulated in SEC peaks P3 and P5, while the material present in the IMAC fraction 2 tended to majorly occur in SEC peaks P1 and P2. P5 appeared to correspond to the unassembled protein forms, namely the monomeric or probably dimeric building blocks.

Table 1: Summary of size and fluorescence emission of major populations of GFP variants. Coloured numbers indicate the protein populations, segregated by size and shape in SEC, as indicted in A

	SEC			DLS	SAXS	Specific fluorescence (µ.a/mg)
Protein	Population	Elution vol. (ml)	R_h	R_h	R_g	
GFP-H6 BL21 (DE3)	P5	15.5	2.9	2.5	2.6	754
A5G27-GFP-H6 BL21 (DE3)	P1	8.6	13.1	8.9	16.7	210
	P2	10.0	7.2	6.7	6.6	341
	P5	16.2	2.4	2.8	5.8	631
T22-GFP-H6 Origami B	P1	8.7	11.4	12.3	13.3	375
	P2	10.4	6.7	N/D	6.1	434
	P5	17.0	2.0	3.5	N/D	704
T22-GFP-H6 KPM335 IMAC fraction 1	P1	8.3	>15	28.0	N/D	N/D
	P3	11.5	5.5	5.1	4.9	796
	P4	14.2	3.7	4.0	4.3	961
	P5	16.1	2.5	N/D	3.8	857
T22-GFP-H6 KPM335 IMAC fraction 2	P1	8.6	13.1	11.8	13.1	N/D
	P2	10.7	6.3	5.9	5.1	336
	P4	14.0	3.7	3.7	4.3	934
	P5	16.3	4.9	2.5	N/D	666

Size (DLS and SAXS) and specific fluorescence (fluorometry) of the relevant protein populations separated by SEC were determined by independent techniques (Table 1), in an exhaustive analysis that also included A5G27-GFP-H6 from *E. coli* BL21 and T22-GFP-H6 produced in Origami B (both eluted in a single IMAC fraction). Again, a high coincidence with size data and elution peaks was observed, here also extended to inter protein pairwise comparisons. The unassembled P5 forms of GFP-H6, A5G27-GFP-H6 and T22-GFP-H6 usually ranged between 2 and 3 nm of radius, P4, when observed, around 4 nm, P2 between 6 and 7 nm and P1 between 11 and 13nm. The high size coincidence when comparing analytical approaches confirmed the robustness in the assembling pattern of the GFP oligomers that appeared to be highly regular despite the

nature of the cationic peptide at the amino terminus of the construct. Just a few obtained data were out of these ranges. In particular, T22-GFP-H6 nanoparticles from the IMAC elution fraction 1 of KPM335 showed an unusual high comparative size, which is coincident with previous observations [22], indicative of a differential organisation of the fully assembled building blocks. Interestingly, all protein fractions were fluorescent, but higher emission values were generically observed in the monomeric forms (P5) when comparing fully assembled nanoparticles and intermediate oligomers (P1 and P2, 50 % of less fluorescent than the building blocks). The range of fluorescence emission independently confirmed alternative conformational status of the oligomerised GFP. All these data allowed proposing, in summary, a variable organisation of GFP oligomers probably due to possible alternative protein-protein contact patterns that might be favoured in particular producing strains.

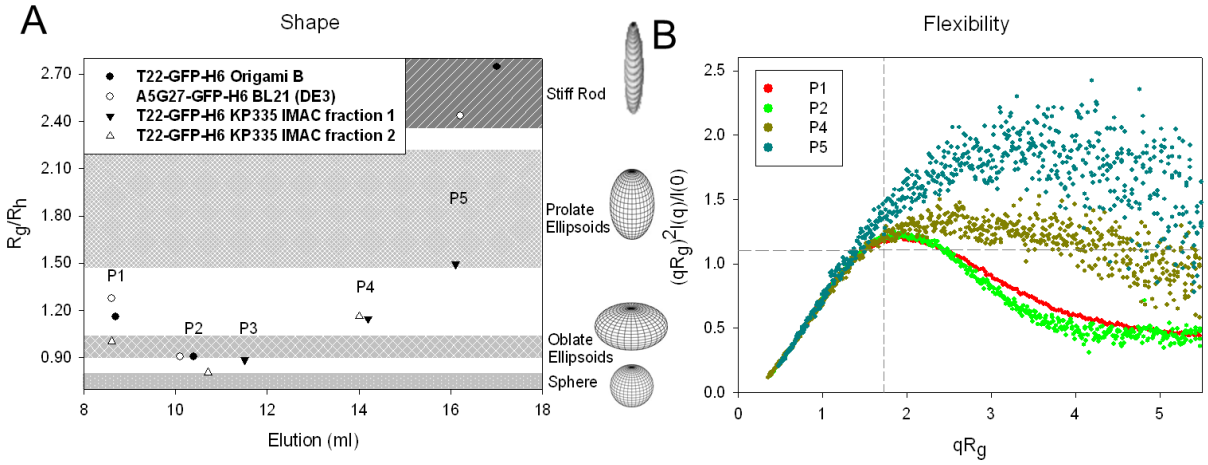


Figure 2. SHAPE AND FLEXIBILITY OF GFP-BASED OLIGOMERS. A. Nanoparticle shape based on the Rg/Rh ratio. Larger oligomers show an oblate ellipsoid shape in comparison with the prolate/stiff rod shape exhibited by monomers. B. Normalized Kratky plot from the scattering curve of T22-GFP-H6 KPM335 IMAC fraction2 oligomeric organisations. The peak position corresponding to a globular protein, which has a value of qRg=√3 with a maximum at 1.104, is indicated with grey-dashed lines as a reference. Nanoparticle compactness is directly proportional with size. Higher nanoparticles display a curve representative of compact structures whereas smaller nanoparticles show peptide flexibility

The different architectonic patterns adopted by GFP oligomers would necessarily be connected to distinct morphometries and biophysical properties of the materials, which have been so far generically identified as planar, toroid nanoparticles [19]. In this context, fine SAXS analyses revealed a broad range of shapes, from rod forms to spherical forms, depending on the SEC population to which they belong (Figure 2 A). Also, the molecular flexibility of the overhanging peptides from GFP-fusions was also variable (Figure 2 B), revealing a changeable potential of the building block to adopt alternative conformations that might limit, impair or favour specific oligomeric organisations. The best fitting of the SAXS profiles for the materials to different form factors was determined by using SasView, as shown in the Supplementary Figure 3. The alternative morphometries in SEC peaks identified by SAXS were fully assessed by high resolution

TEM and FESEM imaging (Figure 3), confirming the nano-architectonic variability of nanoparticles in a fully visual way.

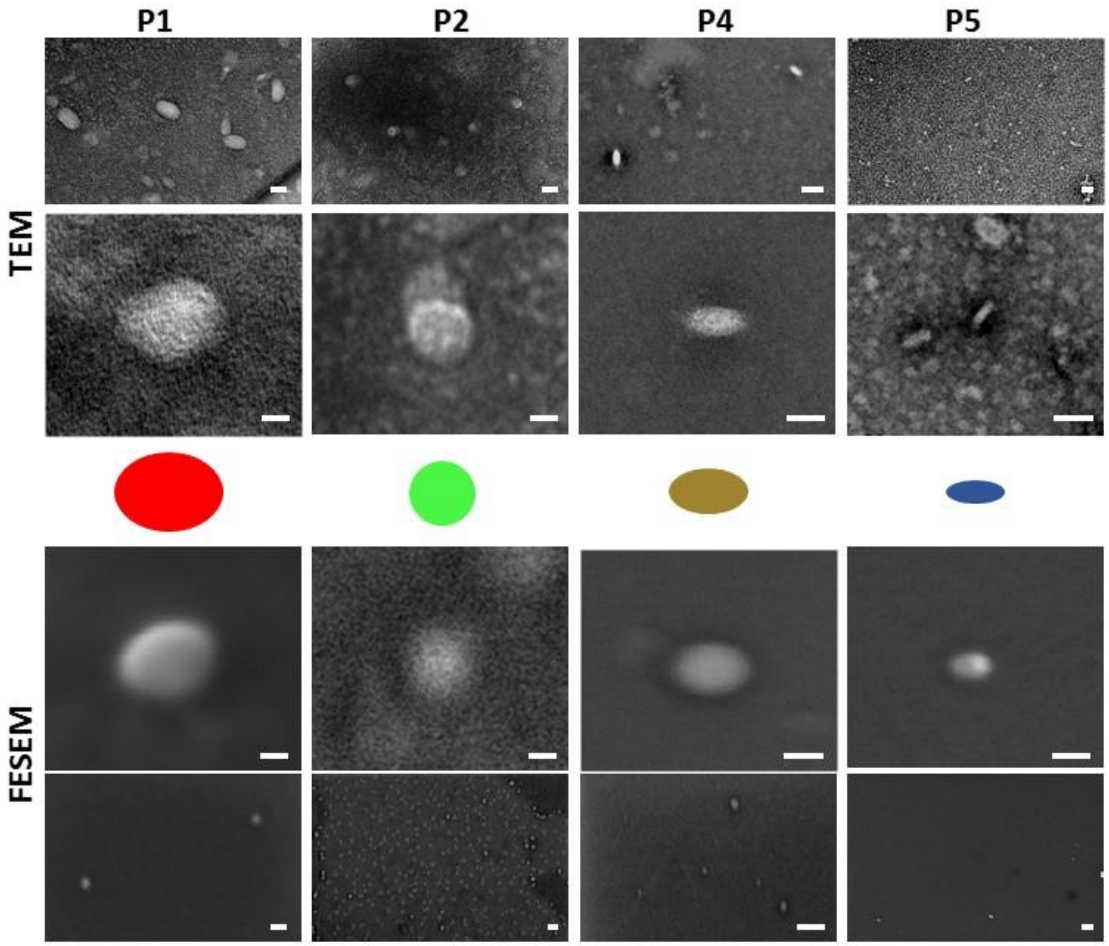


Figure 3. MORPHOMETRY AT ULTRASTRUCTURAL LEVEL OF GFP OLIGOMERS. Representative TEM and FESEM images of T22-GFP-H6 oligomers produced in the E. coli strains Origami B and KPM335, classified by their occurrence in SEC peaks. Bar sizes represent 20 nm in general views and 10 nm in detail images. Shapes within the panels correspond to the expected particle form and their colours to the peak in the plots from Figure 1

The protein nanoparticles studied here had been conceived as drug carriers for cancer treatments, what was lately encouraged by their good biodistribution when systemically administrated, upon which the material accumulated intracellularly in tumor and metastatic foci but not in liver, kidney spleen and other non-target organs [19-20]. Then, how the morphometry and other physical properties of the oligomeric populations might influence receptor-dependent cell penetration is a critical issue that was addressed in a CXCR4+cell culture model. As observed (Figure 4A), the unassembled populations of CXCR4-targeted nanoparticles were inefficient in penetrating target cells, although their uptake was clearly over the background values of the non-targeted, parental GFP-H6 (devoid of any receptor ligand). Noteworthy, cell penetrability was progressively gained with the complexity of the oligomers, indicating that the mere

presence of a cell ligand is not sufficient to support internalisation. This was fully in agreement with previous data obtained with assembled and disassembled IRPF-based protein nanoparticles [26] and with the role of multivalence in the cell binding process and further receptor-mediated internalisation [12]. In fact, monovalent building blocks are equally inefficient in reaching their target *in vivo* [19]. Although internalisation is significantly lower for discreet populations compared to larger structures, noticeably, the prevalence of peptide flexibility in P4 and P5 populations shown by the Normalized Kratky Plot (Figure 2 B) could also be involved in a favoured exposure of T22 to CXCR4 receptor, leading to major specific internalisation. Note also that curves from populations P4 and P5 (Figure 2B) decay at higher qRg values, which is indicative of domain flexibility.

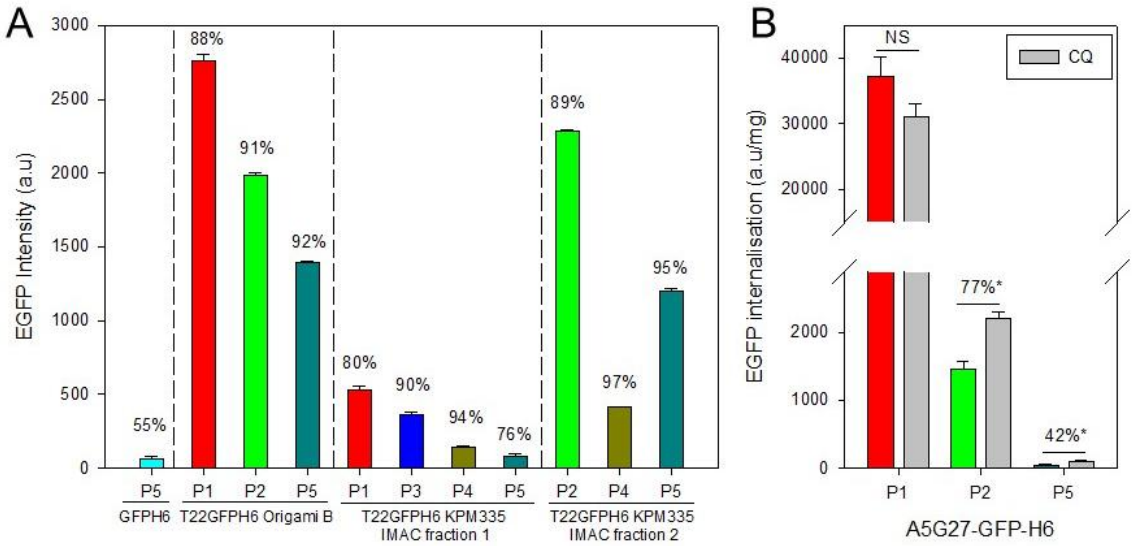


Figure 4. CELL INTERNALISATION OF PROTEIN OLIGOMERS. A. Internalisation in HeLa cells of T22-GFP-H6 oligomers produced in Origami B and KPM 335 strains, determined by the intracellular fluorescence, upon corrected by specific fluorescence (values are then representative of protein amounts). Numbers on top refer to the inhibition of protein entrance by means of a specific CXCR4 antagonist, AMD3100, and indicate the specificity of the internalisation process. B. Internalisation in MDA-MB-231 cells of A5G27-GFP-H6 oligomers. Numbers refer to the percentages of protein that escape from endosomal degradation (* for $p<0.05$) which is also indicated by the grey bars

Interestingly, T22-GFP-H6 nanoparticles produced in KPM335 and present in the IMAC fraction 1, were all poorly internalised (in agreement with previous observations using the material pool [22]), and showed also lower specificity. In these populations, both H6 and T22 tags might be both less available for intermolecular interactions, also fitting with the low affinity in IMAC. However, despite the influence that ligand conformation might have on cell uptake, particle size was found as the major determinant of cell penetrability into target cells, as larger oligomers, presenting enhanced multivalency, internalise more efficiently (compare to P4 and P5) (Figure 4 A and B). Further internalisation analysis with A5G27-GFP-H6 protein oligomers in presence of chloroquine showed an increase in protein lysosomal degradation with a decrease on oligomer size (Figure 4 B). Endosomal escape of larger oligomers was then more efficient compared to smaller protein assembles and together with the entrance specificity

suggest an unspecific internalisation route alternative to clathrin-mediated endocytosis pathway for P1 populations, which do not lead to late endosome avoiding subsequent protein degradation [27]. In this regard, the impact that nanoparticle geometry (size and shape) has on cell penetrability [28] and biodistribution [29] has been demonstrated. However, the influence of these parameters may vary depending of bulk material [30] and, so far, the effect on protein-based nanoparticles have not been closely examined. In receptor-targeted nanoparticles that are activated with overhanging peptides, uptake might be in addition modulated by the exposure and bioavailability of functional ligands on the particle's surface. In our system, these ligands are the tumor homing peptides A5G27 and T22, while the H6 tail has an important role in the endosomal escape upon internalisation [31].

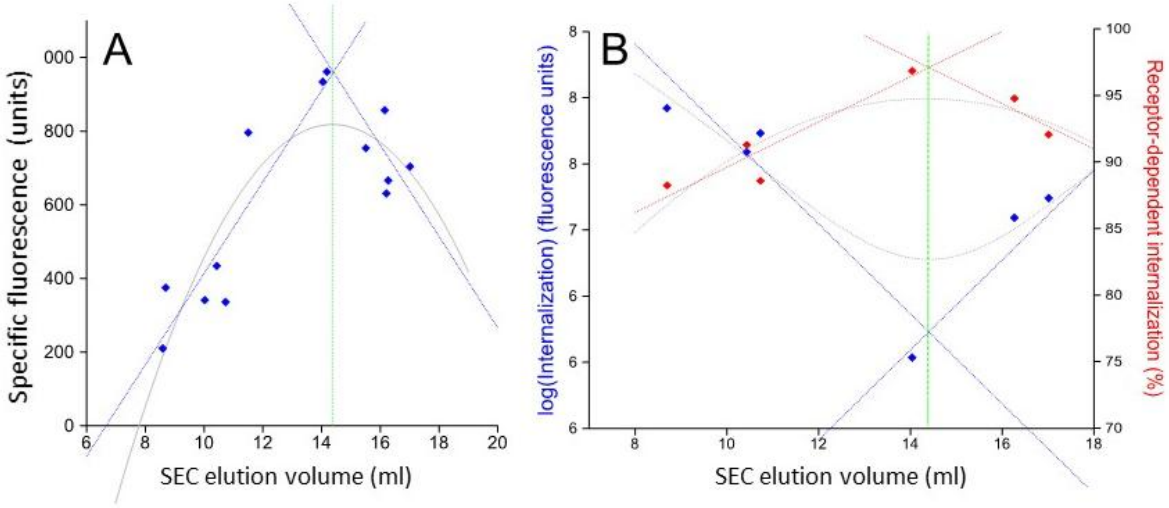


Figure 5. RELATIONSHIPS BETWEEN RELEVANT PARAMETERS OF T22-GFP-H6 NANOPARTICLE POPULATIONS. A. Regresion lines for specific fluorescence F(E) (blue), quadratic prediction F(E) (grey), and symmetry point (green).B. Regresion curves for protein internalisation (logarithm) I(E) (blue)and receptor-dependent cell internalisation R(E) (red) in terms of the SEC-elution volume. Quadratic predictions in grey and symmetry point in green

The variability in the specific GFP fluorescence emission when comparing all the nanoparticles studied here (Figure 1B) suggested alternative configurations of the material [22]. In this context, and for a global analysis of the potential relationships between physical and biological properties of the variant oligomers, we performed a numerical screening of dependences between both sets of parameters. Interestingly, highly fluorescent protein particles were regularly collected in the elution SEC peaks around 14 ml, (Figure 5A), formed by a population of nanoparticles that exhibit medium size. These materials are not particularly efficient as potential drug vehicles, since although receptor dependent penetrability shows a maximum value here, the cell penetrability (amount of internalised material) is contrarily low (Figure 5 B). Interestingly, GFP oligomers eluted at this stage (around 14 ml in the SEC elution volume, corresponding to the SEC P4), represent a functional and morphometric inflexion in the continuum of forms of the assembled protein. In this regard, the SEC P4

sample represents a virtual frontier between two set of nanoparticles, regarding the variability of the morphometric traits of the whole protein population (Figure 5 B).

Among the morphometric population spectrum derived from the cell factory, it is possible to identify and separate by SEC the supramolecular organisation of the nanoparticles that display optimal biological properties. In this regard, for this particular system, oligomers from P2 population (which adopt a spherical shape) exhibit higher cell internalisation and better receptor specificity (Figure 5 B), probably because of a combination of solvent exposure of T22, optimal size and appropriate multivalence of the ligand for cell surface interactions.

In summary, protein nanoparticles generated through the combination of a cationic, cell targeting peptide at the amino terminus of a core GFP and a histidine rich domain at the C-terminus exhibit a spectrum of oligomerisation forms previously eclipsed by the pooled analysis of the material (Figure 1A). The major oligomeric states of the nanoparticles are coincident when comparing proteins constructed with unrelated amino-terminal tags such as A5G27 and T22, and targeted to different cell surface receptors (CD44 and CXCR4, respectively), revealing very regular patterns in the oligomerisation process. These variants exhibit distinguishable biophysical properties including shape, size, fluorescence emission, cellular penetrability and receptor-dependent specificity, some of them critical when considering the applicability of these protein materials as drug carriers in systemic treatments. Interestingly, and considering that high molecular mass of building blocks in protein-only materials are exclusively produced by biological fabrication, the genetic background of the producing cell is dramatically influencing these abilities. Note for instance the poor penetrability of all nanoparticles produced in KPM335 released in the IMAC fraction 1 contrarily to the same material released in fraction 2 (Figure 4). This is probably due to the impact that altered protein folding machineries has in the conformation of the building block, what might be amplified upon assembly as complex oligomeric species. At the present stage of understanding of the protein quality control, it is not possible to rationally predict the assembling pattern and functional profile of a protein nanoparticle when produced in a defined bacterial strain. However, since the yield of building block production in bacteria is relatively high (11-23 mg/l of culture, before any media and process optimisation; not shown), it would be possible to separate, at the downstream stage, particularly convenient material fractions for high performance biological application, and then discard the less efficient. For the materials explored here, proteins eluted in SEC between 8 and 10 ml would appropriately combine high penetrability with high receptor specificity (Figure 5B), what is not the general case as these parameters tend to evolve in divergent fashions. Also, this set of materials show high particle sizes (Figure 1B) and geometries ranging from spheres to discoidal (oblate spheres) versions, far from the more elongated, poorly penetrating versions (Figure 2A and 4A). This is again stressing the regular shaped, high multimerisation forms as the most efficient architectonic patterns for receptor-mediated cell penetration.

Comparing with other types of nanostructures, protein-only nanoparticles are fully compatible with biological systems, providing a suitable platform for biomedical uses such as drug delivery and imaging probes. Far than chemical composition, size and shape are crucial factors determining the relationships between nanoparticles and increasingly complex biological systems (namely from cells to entire organisms) in critical aspects such as organ specificity and biodistribution [32-36], toxicity [37], and cell uptake and fate [35, 38-41]. Although checked in several types of nanostructures, mainly in crystalline nano- and micromaterials, the present study is the first evaluation of form (geometry) and function (cellular uptake) of protein nanoparticle populations. Although variability in biofabrication of self-assembling proteins might represent a priori a generic problem for reproducibility in *in vivo* applications, the extremely regularity and robustness of the oligomerisation patterns instead allows the proper downstream selection of advantageous variants regarding a particular set of applications, such as for instance, we illustrate here regarding cell penetrability and specificity in the receptor-dependent uptake of tumor-targeted protein nanoparticles. This controlling of physical and biological properties of fluorescent nanoparticles may be a crucial aspect to increase efficiency for therapeutic applications.

METHODS

Protein production

T22-GFP-H6 and A5G27-GFP-H6 are self-assembling modular proteins (Supplementary Figure 1) targeted to CXCR4 and CD44 respectively, through amino terminal peptides (T22, RRWCYRKCYKGYCYRKCR and A5G27, RLVSYNGIIFFLK) binding these cell surface receptors [20-21]. T22-GFP-H6 was produced in Escherichia coli Origami B (Novagen, [F−ompT hsdSB (rB− mB−) gal dcm lacY1 ahpC (DE3) gor522::Tn10 trxB (KanR, TetR)]), encoded in a pET22b-derived vector, and in the endotoxin-free KPM335 (msbA52, ΔgutQ, ΔkdsD, ΔlpxL, ΔlpxM, ΔpagP, ΔlpxP, ΔeptA, frr181), its parental BW30270 (CGSC#7925–MG1655; F−, rph+, fnr+) and the routine wild type MC4100 (F− [araD139]B/r, Del(argF-lac)169, flhD5301, Δ(fruK-yeiR)725(fruA25), relA1, rpsL150(strR), rbsR22, Del(fimB-fimE)632(::IS1), deoC1) from a pTrc99a-derivative plasmid. A5G27-GFP-H6 and the control GFP-H6 protein with an N-terminal random (non-cationic) peptide were produced in *E. coli* BL21 DE3 (Novagen) transformed with pET22b and pET21b plasmids respectively. All encoding gene sequences were optimised for expression according to the *E. coli* codon usage. Protein production was induced by the addition of 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) to in bacterial cultures that were further cultivated overnight at 20 ºC in LB medium with the selective antibiotic (at 16 ºC for A5G27-GFP-H6 producing cultures).

Protein purification

Proteins were purified as previously described [20]. Briefly, cell pellets were resuspended in Wash Buffer (10 mM TrisHCl, 500 mM NaCl, 10 mM Imidazol pH 8.0 in presence of protease inhibitors. Cell disruption was performed at 1,200 psi using a French Press and lysate was centrifuged at 15,000 g for 45 min. Protein was purified through the His-tag by Immobilized Metal Affinity Chromatography (IMAC) and protein separation was achieved using an Imidazole gradient up to 500 mM. Protein peak fractions were collected, dialysed against carbonate buffer (166 mM NaCO3H pH 7.4) and centrifuged to remove insoluble aggregates. Protein integrity was analysed by SDS electrophoresis followed by Coomassie Brilliant Blue staining and Western blotting using an anti-His monoclonal antibody (Santa Cruz Biotechnology, Inc., Heidelberg, Germany). Protein concentration was determined by an adapted Bradford's assay [42].

Gel filtration

Protein oligomers were analysed using Size-exclusion chromatography. Samples were loaded on a Superdex 200 Increase 10/300 (GE Healthcare, Piscataway, NJ, USA) column pre-equilibrated with carbonate buffer at 0.75 ml/min. Protein samples intended for small-angle X-ray scattering (SAXS) were prepared by collecting fractions corresponding to each oligomer population and concentrated to 5 mg/ml, using Amicon Centrifugal Filters 3K (Millipore). The supernatant was centrifuged at 15,000 g for 30 min to remove possible aggregates and stored at -80ºC. The hydrodynamic radius (Rh)

was obtained using a protein standard (GE Healthcare), and the relative amounts of oligomeric forms were calculated by Gaussian deconvolution of the obtained size-exclusion chromatograms using the Peakfit 4.12 software (Systat Software Inc.) and applying a residual method to resolve the overlapped peaks.

Dynamic light scattering (DLS) and fluorescence determination

The volume size distribution of nanoparticles was determined at 1 mg/ml in carbonate buffer by dynamic light scattering at 633 nm (Zetasizer Nano ZS, Malvern Instruments Limited, Malvern, UK). Green fluorescence was determined by a Varian Cary Eclipse fluorescence spectrophotometer (Agilent Technologies, Santa Clara, CA, USA) at detection wavelength of 510 nm and at 1 mg/ml protein concentration, by using an excitation wavelength of 450 nm.

Cell lines and protein internalisation

MDA-MB-231 cell line was maintained in RPMI 1640 supplemented with 10 % fetal bovine serum (FBS) and 6 mM GlutaMAX (Invitrogen), whereas HeLa (ATCC-CCL-2) cells were maintained in MEMα (GIBCO, Rockville, MD, USA) supplemented with 10 % FBS. Both cell lines were incubated at 37 ºC and 5 % CO2. To analyse protein internalisation, HeLa and MDA-MB-231 were cultured on 24-well plates at 3·104 cells/well and 8·104 cells/well respectively until reaching 70 % confluence. The medium was discarded and the cells were washed twice with PBS (Sigma-Aldrich, Steinheim, Germany). Cells were then incubated with OptiPRO™ serum-free medium supplemented with L-glutamine with recombinant proteins dissolved at convenient concentrations for 3 h (T22-GFP-H6) or 24 h (A5G27-GFP-H6). AMD3100, a specific CXCR4 antagonist, was added to the cells at 10 mM, 1 h before protein addition, to assess the specificity of protein internalisation. In parallel, to study protein endosomal escape, 50 mM chloroquine was added 3 h before protein addition. After protein incubation, the medium was removed and cells were washed with PBS. Cells were treated with 1 mg/ml trypsin for 15 min to remove protein bound to cell surface followed by the addition of complete medium. Then they were centrifuged at 1,200 rpm for 15 min and the pellet was resuspended in PBS. Protein internalisation was analysed using a FACS-Canto system (Becton Dickinson, Franklin Lakes, NJ, USA) using a 15 mW air-cooled argon ion laser at 488 nm excitation. Experiments were performed in duplicate.

SAXS measurements

SAXS profiles were recorded in the Non-Crystalline Diffraction (NCD) beamline at ALBA Synchrotron Light Source (Cerdanyola del Vallès, Barcelona, Spain), by using an imXPAD-S1400 photon-counting detector (ImXPAD, La Ciotat, France) placed at 5.9 m from the sample. Multiple frames of 0.5-2 seconds exposure time were collected at 12.4 keV energy ($\lambda = 1$ Å) without attenuation. Samples were measured in a Teflon cell with a path length of 3 mm and mica windows of 25 μm thickness.

SAXS data were processed by using the EMBL-Hamburg ATSAS software package [43]. The radius of gyration, Rg, was calculated from the pair-distances distribution function by using GNOM [44]. Prior to Rg calculation, similarity between frames was assessed using the Correlation Map (CorMap) test [45] in order to discard frames with radiation damage. The fitting of the scattering profiles after background subtraction to the different form factors was performed by using SasView 3.1.2.

The ratio between the radius of gyration Rg determined by SAXS and the hydrodynamic radius Rh determined by SEC, Rg/Rh, was used as an estimator of the geometry of the different nanoparticles.

Electron Microscopy

The near native state ultrastructure of NP was assessed with Field Emission Scanning Electron Microscopy (FESEM) and Transmission Electron Microscopy (TEM). For FESEM, drops of 3 µl ofsamples of P1 (Origami B and KPM335 IMAC fraction 2), P2 (Origami B and KPM335 IMAC fraction 2), P4 (KPM335 IMAC fraction 1 and 2), and P5 (Origami B and KPM335 IMAC fraction 2) were directly deposited on a silicon surface (Ted Pella Inc., Reading, CA, USA) for1 min, air dried and observed without coating in a FESEM Merlin (Zeiss, Oberkochen, Germany) operating at 2 kV. Images were acquired with a high resolution in-lens secondary electron detector.

For TEM, drops of 3 µl from the same 8 samples, at the same concentrations usedfor FESEM, were deposited for2 min on 400 mesh carbon coated copper grids, contrasted with 2 % uranyl acetate for2 min, air dried and observed with a transmission electron microscope JeolJEM-1400 (Jeol Ltd., Tokyo, Japan) operating at 80 kV. Images were acquired with a CCD Gatan ES1000W Erlangshen camera (Gatan, Abingdon, UK).

Mathematical Methods

Internalisation (I) and Receptor-dependent internalisation (R) are variables found depending on the specific fluorescence F and SEC elution volume E. According to Rueda and co-workers [22], the relationship I(F) is given by $\ln(I) = a + bF$. The receptor-dependent specific internalisation R(F) is also approximately linear $R = p + qF$. To interpret the plots of I(E) and R(E) shown in Fig. 5B it was previously necessary to analize the relationship between the specific Fluorescence F and the SEC-Elution volume E plotted in Fig. 5A, which presents two main features: first, the symmetry about the maximum specific fluorescence F(e0) (green line); second, the linear and opposite trends of the dots about this symmetry point. To find out the value e0, the data were approximated by a parabola (in grey), by providing the optimal SEC elution volume point at e0 =14.4 ml. Now it was possible to calculate the least squares approximation for the following relationship (blue lines),

$$F = m + n|E - e0|$$

Therefore, the compositions of functions I(F(E)) and R(F(E)) necessarily show the same behaviour represented in Fig. 5B. The symmetry point e0 matches the previous value. The regression curve in blue of Fig. 3B was

$$\ln(I) = a' + b'|E - e0|;\ a' = a + bm;\ b' = bn$$

and the curve in red was

$$R = p' + q'|E - e0|;\ p' = p + qm;\ q' = qn$$

The particular role of the optimal SEC elution volume e0 was also noticeable in terms of other parameters, such as the hydrodynamic diameter and the form.

ACKNOWLEDGMENT

We are indebted to MINECO (BIO2013-41019-P), AGAUR (2014SGR-132) and CIBER de Bioingeniería, Biomateriales y Nanomedicina (project NANOPROTHER) to AV, Marató de TV3 foundation (TV32013-132031) and AGAUR (2014 PROD 00055) to RM, Marató de TV3 foundation to EV (TV32013-133930) and FIS to EV and RM (PI15/00272, PI15/00378) for funding our research on protein-based therapeutics. This work benefitted from SasView software, originally developed by the DANSE project under NSF award DMR-0520547 (SasView, http://www.sasview.org/). EC was supported by the Spanish Nuclear Safety Council (Consejo de Seguridad Nuclear, CSN). Protein production has been partially performed by the ICTS "NANBIOSIS", more specifically by the Protein Production Platform of CIBER in Bioengineering, Biomaterials & Nanomedicine (CIBER-BBN)/ IBB, at the UAB SepBioES scientific-technical service (http://www.nanbiosis.es/unit/u1-protein-production-platform-ppp/). We are also indebted to Fran Cortés from the Cell Culture and Cytometry Units of the Servei de Cultius Cel·lulars, Producció d'Anticossos i Citometria (SCAC), and to the Servei de Microscòpia, both at the UAB. Strain KPM335 was kindly provided by Research Corporation Technologies, Tucson, AZ. AV received an ICREA ACADEMIA award. UU received a Sara Borrell postdoctoral fellowship from ISCIII

REFERENCES

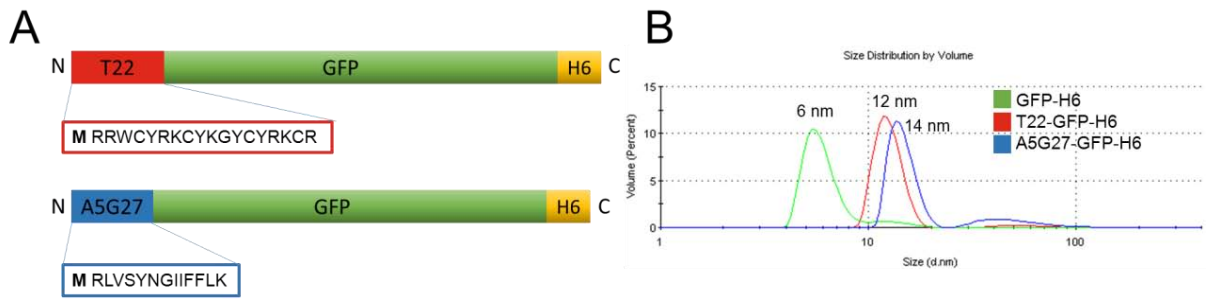
1. Sanchez-Garcia, L.; Martin, L.; Mangues, R.; Ferrer-Miralles, N.; Vazquez, E.; Villaverde, A., Recombinant pharmaceuticals from microbial cells: a 2015 update. Microb Cell Fact 2016, 15.
2. Marino-Ramirez, L.; Hu, J. C., Isolation and mapping of self-assembling protein domains encoded by the Saccharomyces cerevisiae genome using lambda repressor fusions. Yeast 2002, 19 (7), 641-650.
3. Engel, J.; Kammerer, R. A., What are oligomerisation domains good for? Matrix Biol 2000, 19 (4), 283-288.
4. Gradiar, H.; Jerala, R., Self-assembled bionanostructures: proteins following the lead of DNA nanostructures. J Nanobiotechnol 2014, 12.
5. Doll, T. A. P. F.; Dey, R.; Burkhard, P., Design and optimisation of peptide nanoparticles. J Nanobiotechnol 2015, 13.
6. Li, D.; Jones, E. M.; Sawaya, M. R.; Furukawa, H.; Luo, F.; Ivanova, M.; Sievers, S. A.; Wang, W. Y.; Yaghi, O. M.; Liu, C.; Eisenberg, D. S., Structure-Based Design of Functional Amyloid Materials. J Am Chem Soc 2014, 136 (52), 18044-18051.
7. Corchero, J. L.; Vazquez, E.; Garcia-Fruitos, E.; Ferrer-Miralles, N.; Villaverde, A., Recombinant protein materials for bioengineering and nanomedicine. Nanomedicine-Uk 2014, 9 (18), 2817-2828.
8. Ferrer-Miralles, N.; Rodriguez-Carmona, E.; Corchero, J. L.; Garcia-Fruitos, E.; Vazquez, E.; Villaverde, A., Engineering protein self-assembling in protein-based nanomedicines for drug delivery and gene therapy. Crit Rev Biotechnol 2015, 35 (2), 209-221.
9. Cranford, S.; Buehler, M. J., Materiomics: biological protein materials, from nano to macro. Nanotechnol Sci Appl 2010, 3, 127-48.
10. de la Rica, R.; Matsui, H., Applications of peptide and protein-based materials in bionanotechnology. Chem Soc Rev 2010, 39 (9), 3499-3509.
11. Webber, M. J.; Appel, E. A.; Meijer, E. W.; Langer, R., Supramolecular biomaterials. Nat Mater 2016, 15 (1), 13-26.
12. Unzueta, U.; Cespedes, M. V.; Vazquez, E.; Ferrer-Miralles, N.; Mangues, R.; Villaverde, A., Towards protein-based viral mimetics for cancer therapies. Trends Biotechnol 2015, 33 (5), 253-258.
13. Loo, Y.; Goktas, M.; Tekinay, A. B.; Guler, M. O.; Hauser, C. A. E.; Mitraki, A., Self-Assembled Proteins and Peptides as Scaffolds for Tissue Regeneration. Adv Healthc Mater 2015, 4 (16), 2557-2586.
14. Jain, R. K.; Joyce, P. B. M.; Molinete, M.; Halban, P. A.; Gorr, S. U., Oligomerisation of green fluorescent protein in the secretory pathway of endocrine cells. Biochem J 2001, 360, 645-649.
15. Tang, J. C. Y.; Rudolph, S.; Dhande, O. S.; Abraira, V. E.; Choi, S.; Lapan, S. W.; Drew, I. R.; Drokhlyansky, E.; Huberman, A. D.; Regehr, W. G.; Cepko, C. L., Cell type-specific manipulation with GFP-dependent Cre recombinase. Nat Neurosci 2015, 18 (9), 1334-+.
16. Kim, Y. E.; Kim, Y. N.; Kim, J. A.; Kim, H. M.; Jung, Y., Green fluorescent protein nanopolygons as monodisperse supramolecular assemblies of functional proteins with defined valency. Nat Commun 2015, 6.
17. Unzueta, U.; Ferrer-Miralles, N.; Cedano, J.; Xu, Z. K.; Pesarrodona, M.; Saccardo, P.; Garcia-Fruitos, E.; Domingo-Espin, J.; Kumar, P.; Gupta, K. C.; Mangues, R.; Villaverde, A.; Vazquez, E., Non-amyloidogenic peptide tags for the regulatable self-assembling of protein-only nanoparticles. Biomaterials 2012, 33 (33), 8714-8722.
18. Serna, N.; Cespedes, M. V.; Saccardo, P.; Xu, Z. K.; Unzueta, U.; Alamo, P.; Pesarrodona, M.; Sanchez-Chardi, A.; Roldan, M.; Mangues, R.; Vazquez, E.; Villaverde, A.; Ferrer-Miralles, N., Rational engineering of single-chain polypeptides into protein-only, BBB-targeted nanoparticles. Nanomed-Nanotechnol 2016, 12 (5), 1241-1251.
19. Cespedes, M. V.; Unzueta, U.; Tatkiewicz, W.; Sanchez-Chardi, A.; Conchillo-Sole, O.; Alamo, P.; Xu, Z. K.; Casanova, I.; Corchero, J. L.; Pesarrodona, M.; Cedano, J.; Daura, X.; Ratera, I.; Veciana, J.;

Ferrer-Miralles, N.; Vazquez, E.; Villaverde, A.; Mangues, R., *In vivo* Architectonic Stability of Fully de Novo Designed Protein-Only Nanoparticles. Acs Nano 2014, 8 (5), 4166-4176.

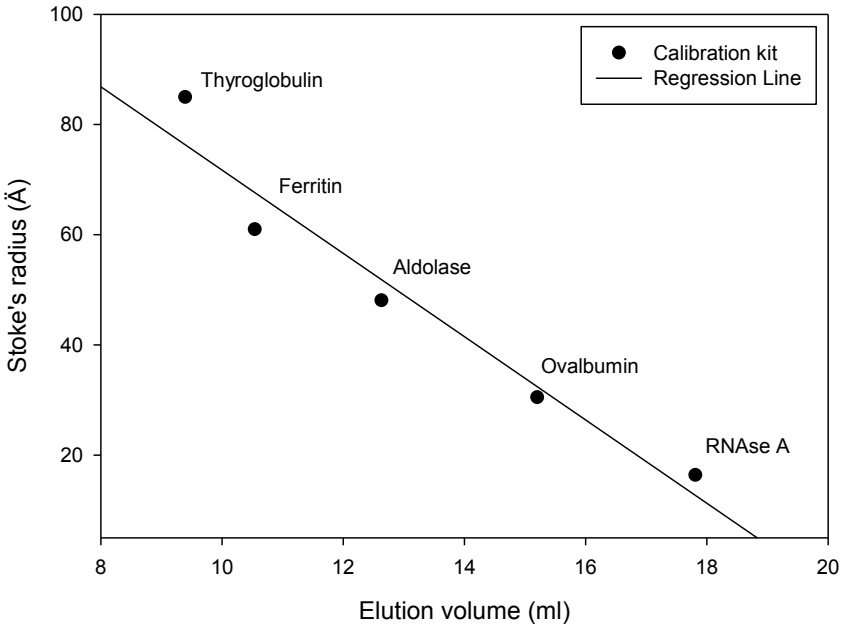
20. Pesarrodona, M.; Ferrer-Miralles, N.; Unzueta, U.; Gener, P.; Tatkiewicz, W.; Abasolo, I.; Ratera, I.; Veciana, J.; Schwartz, S.; Villaverde, A.; Vazquez, E., Intracellular targeting of CD44(+) cells with self-assembling, protein only nanoparticles. Int J Pharmaceut 2014, 473 (1-2), 286-295.
21. Unzueta, U.; Cespedes, M. V.; Ferrer-Miralles, N.; Casanova, I.; Cedano, J.; Corchero, J. L.; Domingo-Espin, J.; Villaverde, A.; Mangues, R.; Vazquez, E., Intracellular CXCR4(+) cell targeting with T22-empowered protein-only nanoparticles. Int J Nanomed 2012, 7, 4533-4544.
22. Rueda, F.; Cespedes, M. V.; Conchillo-Sole, O.; Sanchez-Chardi, A.; Seras-Franzoso, J.; Cubarsi, R.; Gallardo, A.; Pesarrodona, M.; Ferrer-Miralles, N.; Daura, X.; Vazquez, E.; Garcia-Fruitos, E.; Mangues, R.; Unzueta, U.; Villaverde, A., Bottom-Up Instructive Quality Control in the Biofabrication of Smart Protein Materials. Adv Mater 2015, 27 (47), 7816-7822.
23. Rueda, F.; Cespedes, M. V.; Sanchez-Chardi, A.; Seras-Franzoso, J.; Pesarrodona, M.; Ferrer-Miralles, N.; Vazquez, E.; Rinas, U.; Unzueta, U.; Mamat, U.; Mangues, R.; Garcia-Fruitos, E.; Villaverde, A., Structural and functional features of self-assembling protein nanoparticles produced in endotoxin-free Escherichia coli. Microb Cell Fact 2016, 15.
24. Cano-Garrido, O.; Cespedes, M. V.; Unzueta, U.; Saccardo, P.; Roldan, M.; Sanchez-Chardi, A.; Cubarsi, R.; Vazquez, E.; Mangues, R.; Garcia-Fruitos, E.; Villaverde, A., CXCR4(+)-targeted protein nanoparticles produced in the food-grade bacterium Lactococcus lactis. Nanomedicine-Uk 2016, 11 (18), 2387-2398.
25. Mamat, U.; Wilke, K.; Bramhill, D.; Schromm, A. B.; Lindner, B.; Kohl, T. A.; Corchero, J. L.; Villaverde, A.; Schaffer, L.; Head, S. R.; Souvignier, C.; Meredith, T. C.; Woodard, R. W., Detoxifying Escherichia coli for endotoxin-free production of recombinant proteins. Microb Cell Fact 2015, 14.
26. Xu, Z. K.; Unzueta, U.; Roldan, M.; Mangues, R.; Sanchez-Chardi, A.; Ferrer-Miralles, N.; Villaverde, A.; Vazquez, E., Formulating tumor-homing peptides as regular nanoparticles enhances receptor-mediated cell penetrability. Mater Lett 2015, 154, 140-143.
27. Medina-Kauwe, L. K.; Xie, J.; Hamm-Alvarez, S., Intracellular trafficking of nonviral vectors. Gene Ther 2005, 12 (24), 1734-1751.
28. Jiang, W.; Kim, B. Y. S.; Rutka, J. T.; Chan, W. C. W., Nanoparticle-mediated cellular response is size-dependent. Nat Nanotechnol 2008, 3 (3), 145-150.
29. Blanco, E.; Shen, H.; Ferrari, M., Principles of nanoparticle design for overcoming biological barriers to drug delivery. Nat Biotechnol 2015, 33 (9), 941-951.
30. Murugan, K.; Choonara, Y. E.; Kumar, P.; Bijukumar, D.; du Toit, L. C.; Pillay, V., Parameters and characteristics governing cellular internalisation and trans-barrier trafficking of nanostructures. Int J Nanomed 2015, 10, 2191-2206.
31. Ferrer-Miralles, N.; Corchero, J. L.; Kumar, P.; Cedano, J. A.; Gupta, K. C.; Villaverde, A.; Vazquez, E., Biological activities of histidine-rich peptides; merging biotechnology and nanomedicine. Microb Cell Fact 2011, 10.
32. Decuzzi, P.; Godin, B.; Tanaka, T.; Lee, S. Y.; Chiappini, C.; Liu, X.; Ferrari, M., Size and shape effects in the biodistribution of intravascularly injected particles. J Control Release 2010, 141 (3), 320-327.
33. Hirn, S.; Semmler-Behnke, M.; Schleh, C.; Wenk, A.; Lipka, J.; Schaffler, M.; Takenaka, S.; Moller, W.; Schmid, G.; Simon, U.; Kreyling, W. G., Particle size-dependent and surface charge-dependent biodistribution of gold nanoparticles after intravenous administration. Eur J Pharm Biopharm 2011, 77 (3), 407-416.
34. Osborne, O. J.; Lin, S. J.; Chang, C. H.; Ji, Z. X.; Yu, X. C.; Wang, X.; Lin, S.; Xia, T.; Nel, A. E., Organ-Specific and Size-Dependent Ag Nanoparticle Toxicity in Gills and Intestines of Adult Zebrafish. Acs Nano 2015, 9 (10), 9573-9584.
35. Toy, R.; Peiris, P. M.; Ghaghada, K. B.; Karathanasis, E., Shaping cancer nanomedicine: the effect of particle shape on the *in vivo* journey of nanoparticles. Nanomedicine-Uk 2014, 9 (1), 121-134.
36. Tan, J. F.; Shah, S.; Thomas, A.; Ou-Yang, H. D.; Liu, Y. L., The influence of size, shape and vessel geometry on nanoparticle distribution. Microfluid Nanofluid 2013, 14 (1-2), 77-87.

37. Powers, K. W.; Carpinone, P. L.; Siebein, K. N., Characterisation of nanomaterials for toxicological studies. Methods Mol Biol 2012, 926, 13-32.
38. Albanese, A.; Tang, P. S.; Chan, W. C. W., The Effect of Nanoparticle Size, Shape, and Surface Chemistry on Biological Systems. Annu Rev Biomed Eng 2012, 14, 1-16.
39. Chithrani, B. D.; Ghazani, A. A.; Chan, W. C. W., Determining the size and shape dependence of gold nanoparticle uptake into mammalian cells. Nano Lett 2006, 6 (4), 662-668.
40. Herd, H.; Daum, N.; Jones, A. T.; Huwer, H.; Ghandehari, H.; Lehr, C. M., Nanoparticle Geometry and Surface Orientation Influence Mode of Cellular Uptake. Acs Nano 2013, 7 (3), 1961-1973.
41. Shang, L.; Nienhaus, K.; Nienhaus, G. U., Engineered nanoparticles interacting with cells: size matters. J Nanobiotechnol 2014, 12.
42. Pesarrodona, M.; Fernandez, Y.; Foradada, L.; Sanchez-Chardi, A.; Conchillo-Sole, O.; Unzueta, U.; Xu, Z. K.; Roldan, M.; Villegas, S.; Ferrer-Miralles, N.; Schwartz, S.; Rinas, U.; Daura, X.; Abasolo, I.; Vazquez, E.; Villaverde, A., Conformational and functional variants of CD44-targeted protein nanoparticles bio-produced in bacteria. Biofabrication 2016, 8 (2).
43. Petoukhov, M. V.; Franke, D.; Shkumatov, A. V.; Tria, G.; Kikhney, A. G.; Gajda, M.; Gorba, C.; Mertens, H. D. T.; Konarev, P. V.; Svergun, D. I., New developments in the ATSAS program package for small-angle scattering data analysis. J Appl Crystallogr 2012, 45, 342-350.
44. Svergun, D. I., Determination of the Regularisation Parameter in Indirect-Transform Methods Using Perceptual Criteria. J Appl Crystallogr 1992, 25, 495-503.
45. Franke, D.; Jeffries, C. M.; Svergun, D. I., Correlation Map, a goodness-of-fit test for one-dimensional X-ray scattering spectra. Nat Methods 2015, 12 (5), 419-+.

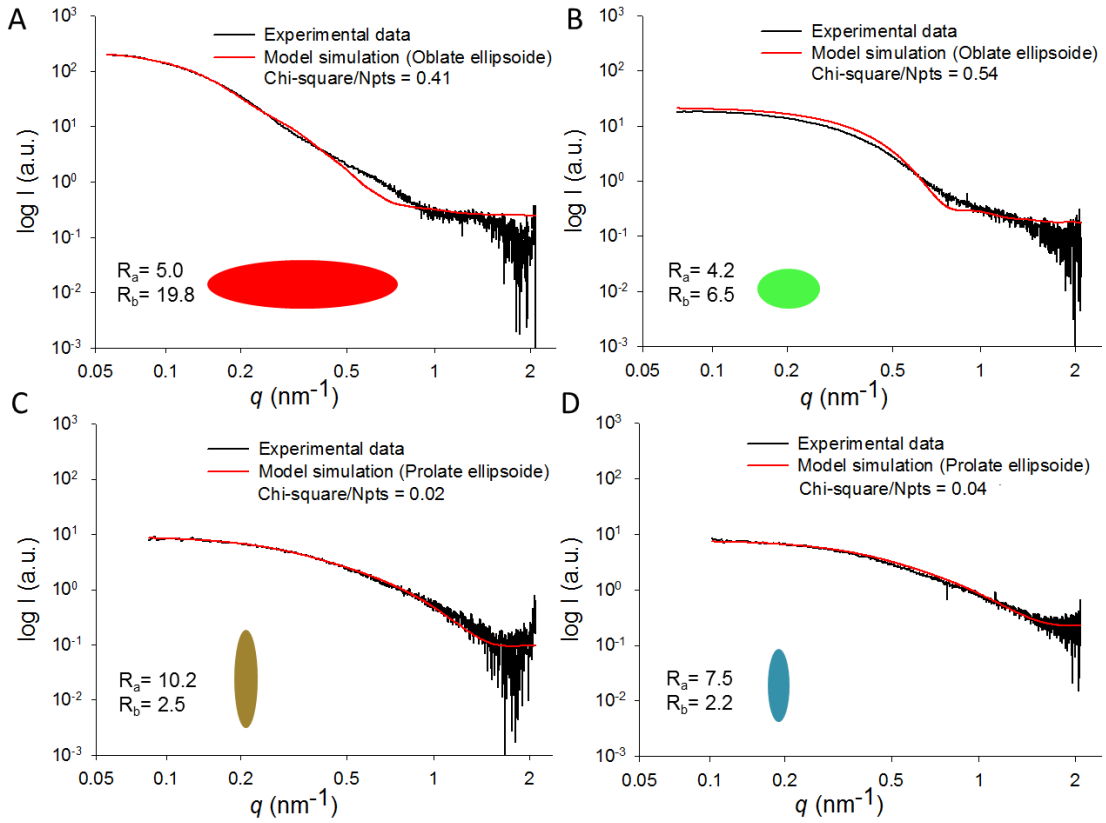
SUPPLEMENTARY INFORMATION



Supplementary Figure 1. BASIC FEATURES OF T22-GFP-H6 AND A5G27-GFP-H6 BUILDING BLOCKS. A. Modular design of T22-GFP-H6 and A5G27-GFP-H6, from amino (N) to carboxyl (C) termini. In boxes, the amino acid sequences of the cell receptor ligands that act simultaneously as architectonic tags. Further details of the amino acid sequences can be found elsewhere. The relative sizes of the protein cassettes are not accurate. B. Average size of pooled protein nanoparticles resulting from the spontaneous self-assembling of T22-GFP-H6 and A5G27-GFP-H6 building blocks and of unassembled protein control GFP-H6, determined by DLS.



Supplementary Figure 2. SEC CALIBRATION CURVE using a standard calibration kit of known Stoke's radius. A regression line is shown.



Supplementary Figure 3. MORPHOMETRIC CHARACTERISATION OF T22-GFP-H6 KPM335 IMAC FRACTION 2 OLIGOMERS. SAXS data P1 (A), P2 (B), P4 (C) and P5 (D) (black dots) were fitted to ellipsoidal form factor simulations (red line). The Chi-square values for the different fittings are displayed on the graphs. Radius A and radius B (vertical and horizontal radius respectively) calculated with SasView and the shaped figure are also shown. Shapes within the panels correspond to the particle form and their colours to the peak in the plots from Figure 1 A.

ANNEX II

Non-amyloidogenic peptide tags for the regulatable self-assembling of protein-only nanoparticles

Unzueta U, Ferrer-Miralles N, Cedano J, Zikung X, **Pesarrodona M**, Saccardo P, García-Fruitós E, Domingo-Espín J, Kumar P, Gupta KC, Mangues R, Villaverde A, Vazquez E

Biomaterials. 2012;33(33):8714-22.

ANNEX III

In vivo architectonic stability of fully de novo designed protein-only nanoparticles

Céspedes MV, Unzueta U, Tatkiewicz W, Sánchez-Chardi A, Conchillo-Solé O, Álamo P, Xu Z, Casanova I, Corchero JL, **Pesarrodona M**, Cedano J, Daura X, Ratera I, Veciana J, Ferrer-Miralles N, Vazquez E, Villaverde A, Mangues R.

ACS Nano. 2014, 27;8(5):4166-76.

ANNEX IV

Bottom-Up Instructive Quality Control in the Biofabrication of Smart Protein Materials

Rueda F, Céspedes MV, Conchillo-Solé O, Sánchez-Chardi A, Seras-Franzoso J, Cubarsi R, Gallardo A, **Pesarrodona M**, Ferrer-Miralles N, Daura X, Vázquez E, García-Fruitós E, Mangues R, Unzueta U, Villaverde A.

Advance Materials 2015 Dec; 27(47):7816-22.

ANNEX V

Structural and functional features of self-assembling protein nanoparticles produced in endotoxin-free Escherichia coli.

Rueda F, Céspedes MV, Sánchez-Chardi A, Seras-Franzoso J, **Pesarrodona M**, Ferrer-Miralles N, Vázquez E, Rinas U, Unzueta U, Mamat U, Mangues R, García-Fruitós E, Villaverde A.

Microbial Cell Factories 2016 Apr; 15:59

Rueda *et al. Microb Cell Fact (2016) 15:59*
DOI 10.1186/s12934-016-0457-z

Microbial Cell Factories

RESEARCH Open Access

Structural and functional features of self-assembling protein nanoparticles produced in endotoxin-free *Escherichia coli*

Fabián Rueda[1,2,3], María Virtudes Céspedes[3,4], Alejandro Sánchez-Chardi[5], Joaquín Seras-Franzoso[1,2,3,9], Mireia Pesarrodona[1,2,3], Neus Ferrer-Miralles[1,2,3], Esther Vázquez[1,2,3], Ursula Rinas[6,7], Ugutz Unzueta[3,4], Uwe Mamat[8], Ramón Mangues[3,4], Elena García-Fruitós[1,2,3,10] and Antonio Villaverde[1,2,3*]

Abstract

Background: Production of recombinant drugs in process-friendly endotoxin-free bacterial factories targets to a lessened complexity of the purification process combined with minimized biological hazards during product application. The development of nanostructured recombinant materials in innovative nanomedical activities expands such a need beyond plain functional polypeptides to complex protein assemblies. While *Escherichia coli* has been recently modified for the production of endotoxin-free proteins, no data has been so far recorded regarding how the system performs in the fabrication of smart nanostructured materials.

Results: We have here explored the nanoarchitecture and in vitro and in vivo functionalities of CXCR4-targeted, self-assembling protein nanoparticles intended for intracellular delivery of drugs and imaging agents in colorectal cancer. Interestingly, endotoxin-free materials exhibit a distinguishable architecture and altered size and target cell penetrability than counterparts produced in conventional *E. coli* strains. These variant nanoparticles show an eventual proper biodistribution and highly specific and exclusive accumulation in tumor upon administration in colorectal cancer mice models, indicating a convenient display and function of the tumor homing peptides and high particle stability under physiological conditions.

Discussion: The observations made here support the emerging endotoxin-free *E. coli* system as a robust protein material producer but are also indicative of a particular conformational status and organization of either building blocks or oligomers. This appears to be promoted by multifactorial stress-inducing conditions upon engineering of the *E. coli* cell envelope, which impacts on the protein quality control of the cell factory.

Keywords: Protein engineering, Recombinant proteins, Nanoparticles, Nanomedicine, Biomaterials, Biodistribution, *E. coli*, Endotoxin-free strains

Background

The physiological diversity of microorganisms makes them suitable platforms for the cost-effective fabrication of a diversity of substances and macromolecules, including emerging nanostructured materials such as metal deposits, polymeric granules, functional amyloids and viruses or virus-like protein assemblies [1]. Being highly versatile, environmentally friendly and fully scalable, microbial fabrication is in many aspects more convenient than chemical synthesis [2–4]. Microbial cells, namely bacteria and yeast, have been exploited as factories for the production of many protein drugs that have been approved for use in humans. Historically, *Escherichia coli* was the pioneering cell factory for protein production because of the well-known genetics and metabolism. However, the presence of endotoxins in the cell envelope of Gram-negative bacteria poses an up to now unsolvable obstacle in the use of this organism for drug production

*Correspondence: antoni.villaverde@uab.cat
[1] Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, Cerdanyola del Vallès, 08193 Barcelona, Spain
Full list of author information is available at the end of the article

[5]. Lipopolysaccharide (LPS) removal increases the complexity of protein purification processes, and it is recognized that LPS, even in low amounts, is a common contaminant of complex protein complexes such as bacteriophages [6] and also of plain recombinant proteins [7]. In fact, contaminant LPS and other cell envelope components are responsible for wrongly attributed biological effects of recombinant proteins tested in biomedical assays, and for poor biocompatibility (some examples can be found in [8–10]). In this context, Gram-positive bacteria, yeast and mammalian cells are endotoxin-free alternatives to *E. coli*, and their prevalence in protein drug production has steadily increased [11, 12], although not absent of problems.

Very recently [13], endotoxin-free *E. coli* strains have been developed by fine multi-step mutagenesis as a way to merge the versatility of *E. coli* as an advantageous protein production platform in the absence of endotoxic contaminants in its products. Those strains have been proved efficient in the production of diverse soluble protein species [13] and also of inclusion bodies intended as smart topologies [14] or as functional amyloids for intracellular protein release [15]. We wanted here to explore how the production, functionalities and nanoarchitecture of a complex, virus-like protein nanoparticle (T22-GFP-H6) designed as drug carrier [16] would be affected by the use of this particular fabrication platform. These particles are formed by a self-organizing, CXCR4-targeted polypeptide that accumulates, in the assembled form, in tumoral tissues of colorectal cancer animal models [16, 17]. Accumulation in tumor is promoted by the N-terminal peptide T22, which is a potent ligand of the cell surface cytokine receptor CXCR4 [16]. CXCR4 is overexpressed in colorectal cancer cells and linked to tumor aggressiveness [18]. The combination of the highly cationic T22 peptide with the C-terminal histidine tail in a modular protein promotes the self-assembly of the whole construct, favored by the bipolar charge distribution of the building block surface [19]. Recently, we have observed that strain-dependent genetic features of the producing bacteria impact on tumor targeting and on the whole-body biodistribution of the oligomeric construct upon its systemic administration [20]. Then, we were interested in dissecting the structure and activities of these tumor-homing nanoparticles when produced in endotoxin-free *E. coli* cells, whose complex genetic development might have resulted in relevant physiological shifts.

Methods

Strains, plasmids and media

T22-GFP-H6 is a tumor-targeted modular protein that self-assembles as toroid nanoparticles of about 12 nm of diameter. These oligomers are fully stable in vivo upon tail vein administration in colorectal cancer mice models [17] and accumulate intracellularly in primary tumor and in metastatic foci [16]. T22-GFP-H6 protein nanoparticles were produced in the KPM22 L11-derivative [21], endotoxin-free strain KPM335 (*msbA52*, Δ*gutQ*, Δ*kdsD*, Δ*lpxL*, Δ*lpxM*, Δ*pagP*, Δ*lpxP*, Δ*eptA*, *frr181*) and its parental K-12 strain BW30270 (CGSC#7925–MG1655; F^-, rph^+, fnr^+) with a wild-type LPS. The design of KPM335, based on the incorporation of non-reverting deletions of seven genes to disrupt Kdo biosynthesis and modifications of lipid IV_A, prevents the strain from regaining the potential to synthesize normal LPS or endotoxically active lipid IV_A derivatives through acquiring mutations. Also, *E. coli* K-12 MC4100 ([*araD139*], (*argF-lac*)169, λ—*relA1*, *rpsL150*, *rbsR22*, *flb5301*, *deoC1*, *pstF25* $Strep^R$) and BL21 Origami B [F^- *ompT* $hsdS_B(r_B^- m_B^-)$ *gal dcm lacY1 ahpC* (DE3) *gor522*::Tn*10 trxB* (Kan^R, Tet^R)] (Novagen, Darmstadt, Germany) were used for protein production as controls. KPM335, BW30270 and MC4100 were transformed with plasmid pTrc99a, whereas Origami B was transformed with a pET22b-derived plasmid. All expression plasmids contained the T22-GFP-H6 DNA sequence optimized by the codon usage of *E. coli* and have been described previously [20]. Bacteria were always grown in Lysogeny Broth (LB) rich media [22]. Part of the protein was produced with the assistance of the Protein Production Platform of CIBER-BBN/IBB, at the UAB (http://www.nanbiosis.es/unit/u1-protein-production-platform-ppp/).

Competent cells and transformation

Cultures were grown overnight at 37 °C with shaking at 250 rpm and used as a 1/100 inoculum in 50 ml of LB. After reaching an optical density (OD_{550}) between 0.2 and 0.4, MC4100, BW30270 and Origami B cultures were centrifuged (4000×*g*) at 4 °C for 15 min. Pellets were resuspended in 12.5 ml of cold and sterile 50 mM $CaCl_2$ and incubated for 45 min in an ice bath. Cells were centrifuged again as described above and resuspended in 1.25 ml of cold and sterile 50 mM $CaCl_2$ in glycerol (15 % v/v) to prepare aliquots of 200 µl and stored at −80 °C. To transform the cells, 40 ng of plasmid DNA were added to the competent cells. The mixtures were incubated on ice for 30–60 min, warmed up to 42 °C for 45 s and placed on ice for 30 s. After incubation, 800 µl of LB media were added, and transformed cells were incubated at 37 °C for 1 h. Finally, the cells were plated on LB-agar plates containing the corresponding antibiotic. In the case of KMP335 strain, cells were grown to an OD_{550} between 0.2 and 0.4, placed on ice for 20 min and sedimented by centrifugation (3100×*g*, 4 °C, 20 min). The pellets were centrifuged and resuspended successively in 40, 20 and

10 ml of H_2O (ice-cold and sterile), followed by a wash with 5 ml of 10 % glycerol (ice-cold and sterile). The final pellet was resuspended in 1 ml of 10 % glycerol (ice-cold and sterile) to prepare 50-µl aliquots for storage at −80 °C. Electrocompetent KPM335 cells were transformed by electroporation in pre-chilled 0.2 cm electroporation cuvettes using 50 µl of competent cells and 40 ng of plasmid DNA. Cells were pulsed using a Gene Pulser MX cell electroporator (Bio-Rad, Hercules, CA, USA) at 25 µF, 200 Ω, 2500 V and 4.7–4.8 ms. Immediately after the pulse, 800 µl of LB medium were added, and the mixture was incubated at 37 °C for 1 h. Cells were plated on LB-agar plates containing ampicillin (100 µg/ml) and incubated at 37 °C overnight.

Protein production and purification

Overnight cultures were inoculated in 500 ml of LB media with 100 µg/ml ampicillin in 2-L flasks. Streptomycin (at 30 µg/ml) was also added to MC4100 cultures, and tetracycline (100 µg/ml) and kanamycin (33 µg/ml) to Origami B cultures. Each culture was grown at 37 °C and 250 rpm to an OD_{550} of about 0.5 before isopropyl-β-D-thiogalactoside (IPTG) was added to 1 mM to induce recombinant gene expression. T22-GFP-H6 was produced overnight at 20 °C with shaking (250 rpm) in all cases.

After overnight production, cultures were centrifuged (3280×*g*, 4 °C, 40 min), and the cell pellets were resuspended in 20 mM Tris–HCl, pH 8.0, containing 500 mM NaCl, 20 mM imidazole (buffer A), and an EDTA-free protease inhibitor cocktail (Complete EDTA-free, Roche Diagnostics, Indianapolis, IN, USA). Intra-cellular protein was extracted by cell disruption using a French press (Thermo FA-078A) at 1100 psi, and purified by His-tag affinity chromatography using a 1-ml HiTrap Chelating HP column (GE Healthcare, Piscataway, NJ, USA) in an ÄKTA purifier FPLC (GE Healthcare). Protein separation was achieved with a linear imidazole gradient from 20 mM to 500 mM, by diluting 20 mM Tris–HCl, pH 8.0, plus 500 mM NaCl and 500 mM imidazole (buffer B) in buffer A, and fractions were collected and dialyzed against 166 mM $NaHCO_3$, pH 7.4. The amount of protein was determined by Bradford's assay [23], and the integrity of the protein was analyzed by sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE), followed by Western blot analysis as described below.

Western blot analysis

Separated fractions were subjected to 10 % SDS-PAGE according to Laemmli's method [24]. Protein was transferred to nitrocellulose membranes (GE Healthcare, Piscataway, NJ, USA) that were blocked overnight with 5 % skim milk in PBS. After blocking, membranes were incubated with anti-GFP antibody (1:500, sc-8334, Santa Cruz Biotechnology, Santa Cruz, CA, USA), followed by incubation with a secondary HRP-conjugated anti-rabbit IgG (H + L) antibody (Bio-Rad, Hercules, CA, USA) at a dilution of 1:2000. Protein bands were visualized with a solution of 25 % cold methanol, 0.2 % H_2O_2 and 0.65 mg/ml of 4-chloronaftol in PBS, and images were obtained using a GS800 Calibrated Densitometer scanner (Bio-Rad).

Fluorescence intensity

Fluorescence intensity of T22-GFP-H6 was determined in a Varian Cary Eclipse fluorescence spectrometer (Agilent Technologies, Mulgrave, Australia) at excitation and emission wavelengths of 450 and 510 nm, respectively.

Particle size and zeta potential

Size distribution and zeta potential of nanoparticles were measured by dynamic light scattering (DLS) at 633 nm (Zetasizer Nano ZS, Malvern Instruments Limited, Malvern, Worcestershire, UK). Samples were analyzed by triplicate averaging of fifteen single measurements.

Mass spectrometry (MS) analysis

Purified protein was diluted 1:2 with H_2O MilliQ and dialyzed against 50 mM of NH_4HCO_3 for 1.5 h. After dialysis, protein was mixed with a matrix of 2.6 dihydroxyacetophenone (1:1), and 1 µl of the mixture was deposited on a ground steel plate. Samples were analyzed using a lineal method in an UltrafleXtreme MALDI-TOF instrument (Bruker Daltonics, Bremen, Germany) with ion acceleration of 25 kV.

Circular dichroism (CD)

CD measurements were carried out in a Jasco J-715 spectropolarimeter at 25 °C and 10 µM protein in 160 mM $NaHCO_3$, pH 7.4. CD and high-tension (HT) voltage spectra were obtained over a wavelength range of 200–250 nm at a scan rate of 50 nm/min, a response time of 0.5 s, and a bandwidth of 0.5 nm.

Electron microscopy

To visualize the ultrastructure of protein nanoparticles by transmission electron microscopy (TEM), 2 µl of purified protein (0.2 mg/ml) were placed on carbon-coated copper grids for 1 min. Excess of sample was bloated and 2 µl of 1 % w/v uranyl acetate were added for negative staining. Samples were immediately visualized in a TEM Jeol JEM-1400 (Jeol Ltd., Tokyo, Japan) equipped with a Gatan CCD Erlangshen ES1000 W camera (Gatan Inc, Abingdon, UK) and operating at 80 kV. Ultrastructural analyses of nanoparticle morphology were complemented with imaging in a nearly native state with a

Field Emission Scanning Electron Microscope (FESEM). For that, protein samples were directly deposited over silicon wafers (Ted Pella, Reading, CA, USA), air dried and observed with a high resolution *in-lens* secondary electron detector through a FESEM Zeiss Merlin (Zeiss, Oberkochen, Germany), operating at 2 kV.

Analysis of cell internalization

HeLa cells were cultured at 37 °C in a humidified atmosphere with 5 % CO_2 for 24 h and plated in treated 24 well plates (Nunclon surface, Nunc 150628) with MEM-α medium supplemented with 10 % fetal bovine serum (FBS) and 2 mM Glutamax (Gibco, Rockville, MD, USA). Later, medium was removed and the cells were washed with Dulbecco's phosphate-buffered saline (DPBS). Then, 25 nM of T22-GFP-H6 in 250 µl Optipro supplemented with 2 mM L-glutamine was added. After incubation at 37 °C for 1 h, cells were treated with trypsin in DPBS (1 mg/ml) for 15 min, centrifuged at 300×*g* for 15 min, and pellets were resuspended in 300 µl DPBS. Samples were analyzed by flow cytometry using a FACSCanto system (Becton–Dickinson, Franklin Lakes, NJ, USA) with a 15 W air-cooled argon-ion laser at 488 nm excitation for GFP.

Confocal microscopy

HeLa cells were plated on a MatTek culture dish (MatTek Corporation, Ashland, MA, USA) at 150,000 cells/plate and incubated at 37 °C for 24 h. After incubation, medium was removed to wash the cells twice with DPBS. Optipro medium supplemented with 2 mM L-glutamine and 25 nM T22-GFP-H6 were added, and cells were incubated at 37 °C for 1 h. To visualize the samples, plasma membranes and the nucleus were labeled using 2.5 mg/ml CellMask Deep Red and 0.2 mg/ml Hoechst 33342 (Molecular Probes, Eugene, OR, USA), respectively. A confocal laser scanning microscope TCS-SP5 (Leica Microsystems, Heidelberg, Germany) was used to analyze the samples, and the images were processed through the Imaris Bitplane software (Bitplane, Zurich, Switzerland).

Determination of in vivo biodistribution

This study was approved by the Institutional Animal Ethics Committee. We implanted CXCR4-overexpressing SP5 human colorectal tumor line to generate subcutaneous colorectal tumors in swiss nude mice (Charles River, France) as previously described [16, 17]. When tumors reached ca. 500 mm^3, mice were randomly allocated to Origami B, MC4100, KPM335, BW30270 or buffer-treated groups (N = 3–5/group). The experimental mice received a single intravenous bolus of the corresponding nanoparticle (500 µg in carbonate buffer, pH 7.4), whereas control mice received only buffer. Five hours after the administration, we measured ex vivo the fluorescence emitted by the nanoparticles accumulated in the whole and slice sectioned tumor and normal tissues (kidney, lung, and heart, liver and brain) using IVIS® Spectrum equipment (Perkin Elmer, Waltham, MA, USA). The fluorescence signal was digitalized, and after subtracting the autofluorescence, it was displayed as a pseudocolor overlay and expressed as Radiant efficiency. Data were corrected by the specific fluorescence emitted by the different nanoparticles. Signal difference between groups was determined by applying the non-parametric Mann–Whitney test.

Histological and immunohistochemical evaluation

At necropsy, tumors were fixed with 4 % formaldehyde in PBS for 24 h and embedded in paraffin for histological and immunohistochemical evaluation. 4 µm sections, were processed as previously described [16, 17] and haematoxylin and eosin (H&E) stained for histological analysis, that was performed by two independent observers. We assessed CXCR4 membrane expression and nanoparticle cell internalization in tumor and normal tissues by IHC using primary anti-CXCR4 (1:300; Abcam, Cambridge, UK) or anti-GFP (1:100; Santa Cruz Biotechnology, CA, USA), and secondary HRP conjugated antibody, followed by chromogenic detection. We quantified the percentage of CXCR4-expressing cells in relation to the total cell number and their staining intensity, scoring each sample from 0 to 3 (where 3 is the maximal intensity) and multiplying both parameters to obtain its H-score. Representative pictures were taken using CellɔB software (Olympus Soft Imaging v3.3, Tokyo, Japan) at 400×.

Quantitative real-time PCR

For isolation of total RNA, triplicate cultures of *E. coli* strains BW30270 and KPM335 were grown aerobically with shaking (250 rpm) at 37 °C in LB medium to the mid-exponential growth phase at an OD_{600} of 0.6–0.7. The bacterial cells were harvested in the presence of RNAprotect Bacteria Reagent for immediate RNA stabilization prior to RNA isolation using the RNeasy Mini Kit and on-column DNase treatment in accordance with the recommendations of the manufacturer (Qiagen). RNA concentrations were determined using a NanoDrop 2000 spectrophotometer (Thermo Scientific), and the integrity of the RNAs was verified with RNA Nano Chips in an Agilent 2100 Bioanalyzer. The Maxima First Strand cDNA Synthesis Kit was used to generate cDNA from 0.3 µg RNA following the manufacturer´s instructions (Thermo Scientific, Waltham, MA, USA). The cDNA levels of chaperone genes were then analyzed by quantitative real-time PCR (RT-qPCR) using a LightCycler 480 Instrument II with LightCycler 480 SYBR Green I Master

reaction mixes according to the instructions of the manufacturer (Roche Diagnostics). RT-qPCR was performed with an initial denaturation at 95 °C for 5 min, followed by 45 cycles of 10 s at 95 °C, 10 s at 60 °C, and 10 s at 72 °C. Each sample was analyzed in duplicate using 1:15 and 1:50 dilutions of the cDNAs and the gene-specific primers listed in Table 1. The primer pairs (Table 1) were designed following the rules of highest possible sensitivity and efficiency with the Primer3 2.3.4 plugin of the Geneious v9.0.5 software (Biomatters, Auckland, New Zealand). Relative expression levels of the target genes were normalized to the internal reference gene *ihfB* for the β-subunit of the integration host factor using the comparative $2^{-\Delta\Delta Ct}$ method [25].

Results and discussion

The self-assembling protein T22-GFP-H6 (Fig. 1a) was produced in the endotoxin-free strain KPM335 and its parental BW30270, as well as in MC4100 and Origami B strains, used in previous studies for the routine fabrication of T22-GFP-H6 nanoparticles. While the generation time of KPM335 has increased by about 50 % in comparison to the parental wild-type strain, no signs of instability, lysis, foaming or cell death could be observed.

Table 1 Primers used for RT-qPCR

Gene	Primer	Sequence	Amplicon length (bp)
tig	ECtig353F	AGGGTCTGGAAGCGATCGAAG	301
	ECtig653R	GGGAAGGTCACGTCGATGGT	
ibpA	ECibpA85F	CAGAGTAATGGCGGCTACCCT	299
	ECibpA383R	GCTTCCGGAATCACGCGTTC	
hscB	EChscB191F	CGTTAATGCGCGCGGAATATTTG	301
	EChscB491R	AGTTGTTCGGCACTGCTTCG	
hchA	EChchA398F	TCGCGGATGTTGTTGCCAG	300
	EChchA697R	CGCCGAAGTACCAGGTGAGA	
grpE	ECgrpE278F	ACGAATTGCTGCCGGTGATTG	301
	ECgrpE578R	GCTACAGTAACCATCGCCGC	
groL	ECgroL583F	TTCGACCGTGGCTACCTGTC	301
	ECgroL883R	GGGTTGCGATATCCTGCAGC	
dnaK	ECdnaK504F	CATCAACGAACCGACCGCAG	300
	ECdnaK803R	TTTTCTGCCGCTTCTTTCAGGC	
clpB	ECclpB426F	AATGCGTGGAGGTGAAAGCG	300
	ECclpB725R	ATATCCAGCGCCAGTACCCG	
clpA	ECclpA287F	AGTCCTCCGGTCGCAATGAG	300
	ECclpA586R	CCAGCTCCTTCTCACGACCA	
cbpA	ECcbpA477F	GCTGAATGTGAAGATCCCGGC	300
	ECcbpA776R	ACCAGACCTTTGCCTTTAACGC	
ihfB[a]	ECihfB9F	GTCAGAATTGATAGAAAGACTTGCCACC	258
	ECihfB266R	CGATCGCGCAGTTCTTTACCAG	

[a] Used for normalization of gene expression

In all cases, T22-GFP-H6 appeared as a stable full-length product showing the expected molecular mass of about 30 kDa (Fig. 1b). The steadily observed double band was probably due to different conformational species of the protein. When purified by His affinity chromatography from bacterial cell extracts, T22-GFP-H6 was eluted in two separated peaks (P1 and P2) irrespective of the strain used for production (Fig. 1c). P2 has been previously observed as the one containing protein materials more efficient in cell penetration assays [20]. MS of these P2 samples revealed the expected molecular masses of all protein variants with minor signs of proteolytic instability (more prevalent in MC4100 through the occurrence of a degradation fragment of 28.4 kDa; Fig. 1d), also confirming that the double band observed in blots (Fig. 1b) was indeed due to conformational variability instead to significant occurrence of truncated forms. Both protein yield and emitted fluorescence were significantly lower in KPM335 than in other strains (Fig. 2a), as well as the specific fluorescence emission of T22-GFP-H6 (Fig. 2b). While a moderate protein yield in recombinant KPM335 was not unexpected according to previous reports referring to different produced protein species [13], a specific fluorescence emission lower than that observed in its parental BW30270 (and in other reference strains) was indicative of a differential conformation of T22-GFP-H6 monomers or an alternate configuration of the resulting nanoparticles.

The formation of nanoparticles was firstly confirmed by DLS, revealing a regular size of the material ranging between 10 and 15 nm, depending on the cell factory (Fig. 3a). Interestingly, and according to the above hypothesis of a particular organization of T22-GFP-H6 as produced in KPM335, P1 nanoparticles produced in this strain were particularly large, reaching around 75 nm in diameter (Fig. 3a). This variability did not result in appreciable modifications in the superficial charge of protein nanoparticles (Fig. 3b). The observed deviations in the size of protein materials were fully confirmed by TEM and FESEM morphometric analyses (Fig. 4).

When analyzing penetrability into $CXCR4^+$ HeLa cells in culture, a convenient in vitro model for T22-empowered nanoparticles [16], P2 materials resulted generically more efficient than P1's (Fig. 5a). The exception found in Origami B cells is in agreement with previous observations [20], and it might be related to the less reducing cytoplasm in this bacterial strain and caused by a bottom-up impact of protein conformation on microscopic functioning of the particles in cell interfaces, associated to favored di-sulfide bridge formation in this factory. Cell penetrability of P2 particles was comparatively screened again by confocal microscopy and the enhanced performance of those produced in KPM335 fully confirmed

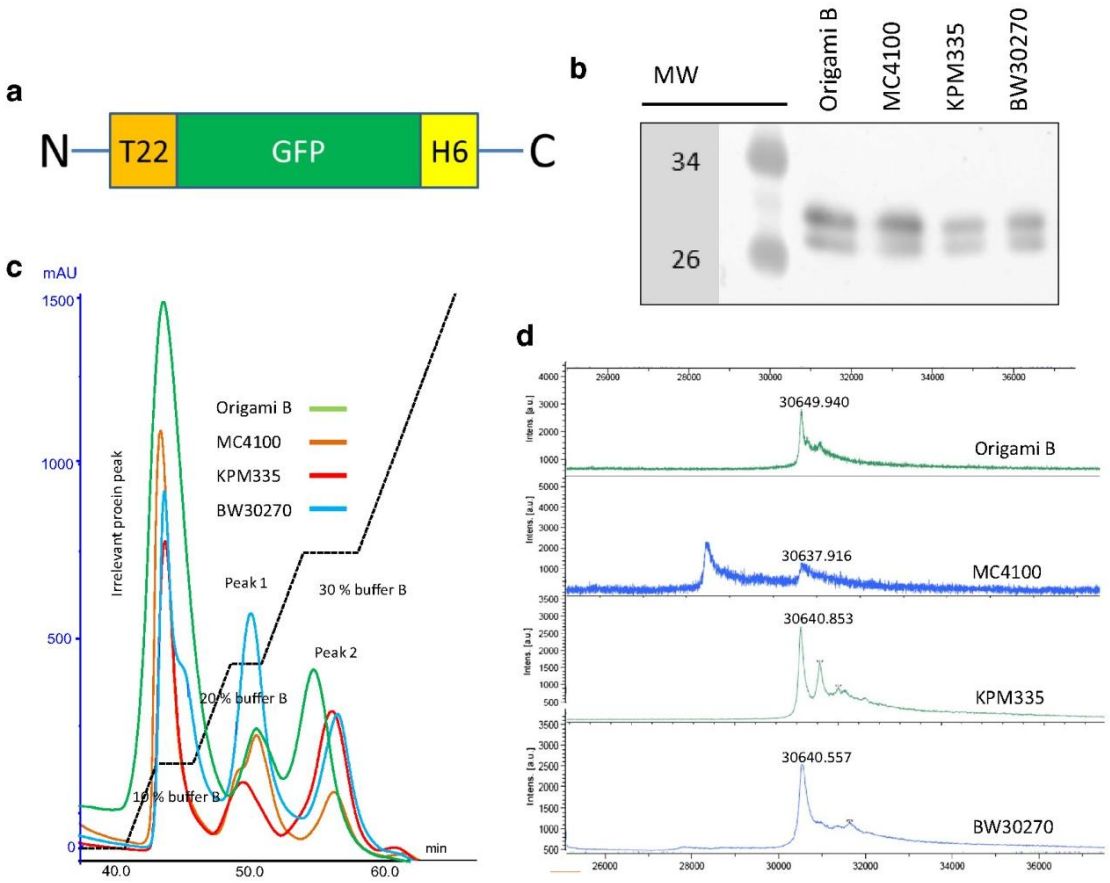


Fig. 1 Protein production and purification. **a** Scheme of the modular T22-GFP-H6 building block. Sizes of boxes are only approximate and do not precisely correspond to the length of primary aa sequence. **b** Western blot of crude cell extracts of T22-GFP-H6-producing strains. MW indicates the molecular weight of selected markers. **c** His-tag affinity chromatography of T22-GFP-H6 produced in the different *E. coli* strains. Purification was performed using an imidazole concentration gradient. Buffer B contains 500 mM imidazole. **d** Maldi-TOF identification of T22-GFP-H6 produced in different *E. coli* strains and purified by affinity chromatography (P2)

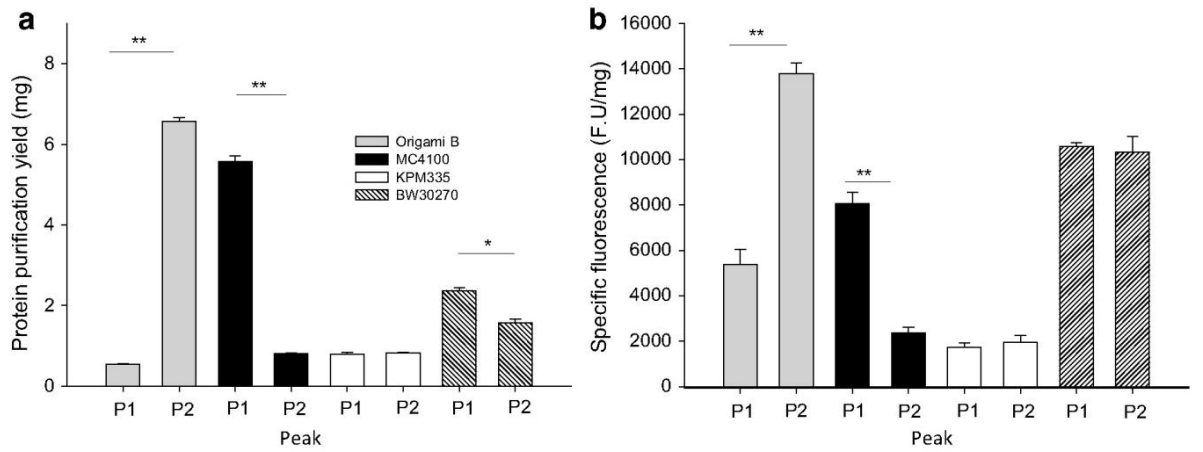


Fig. 2 Quantitative analyses of T22-GFP-H6 production levels and activity. **a** Yield of T22-GFP-H6 produced in *E. coli* strains Origami B, MC4100, KPM335 and BW30270, upon purification, from 500 ml of original protein producing cultures. *P1* and *P2* indicate the main elution peaks 1 and 2. **b** Specific fluorescence of T22-GFP-H6 calculated by using the above data. Symbols mean significant differences: **p < 0.001; *p < 0.05

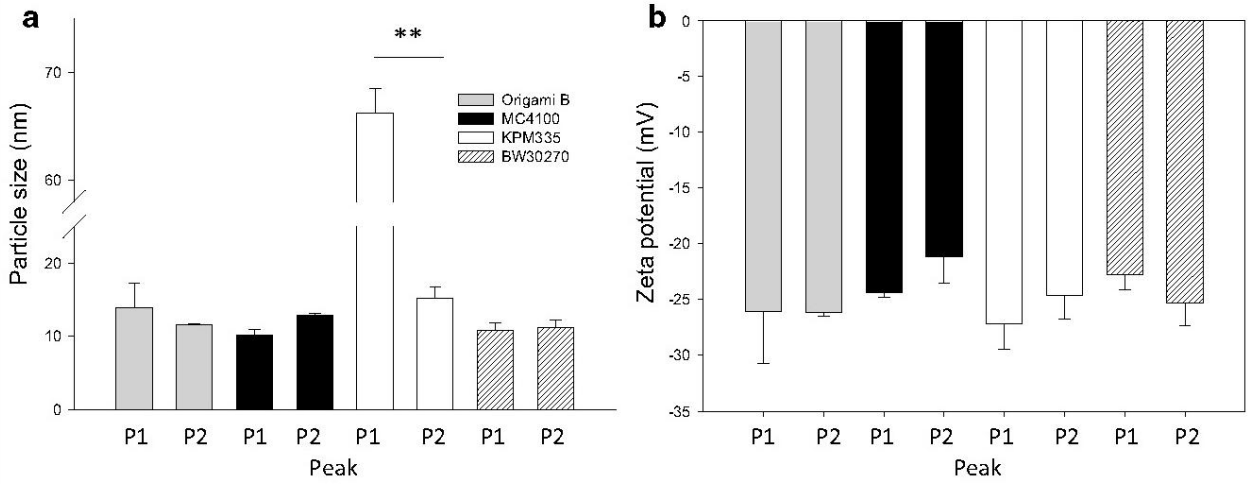


Fig. 3 Physical properties of protein nanoparticles. **a** Size of the self-assembled nanoparticles purified in two elution peaks and measured by DLS. **b** Surface charge distribution of self-assembled nanoparticles measured as zeta potential. Symbols mean significant differences: **p < 0.001; *p < 0.05

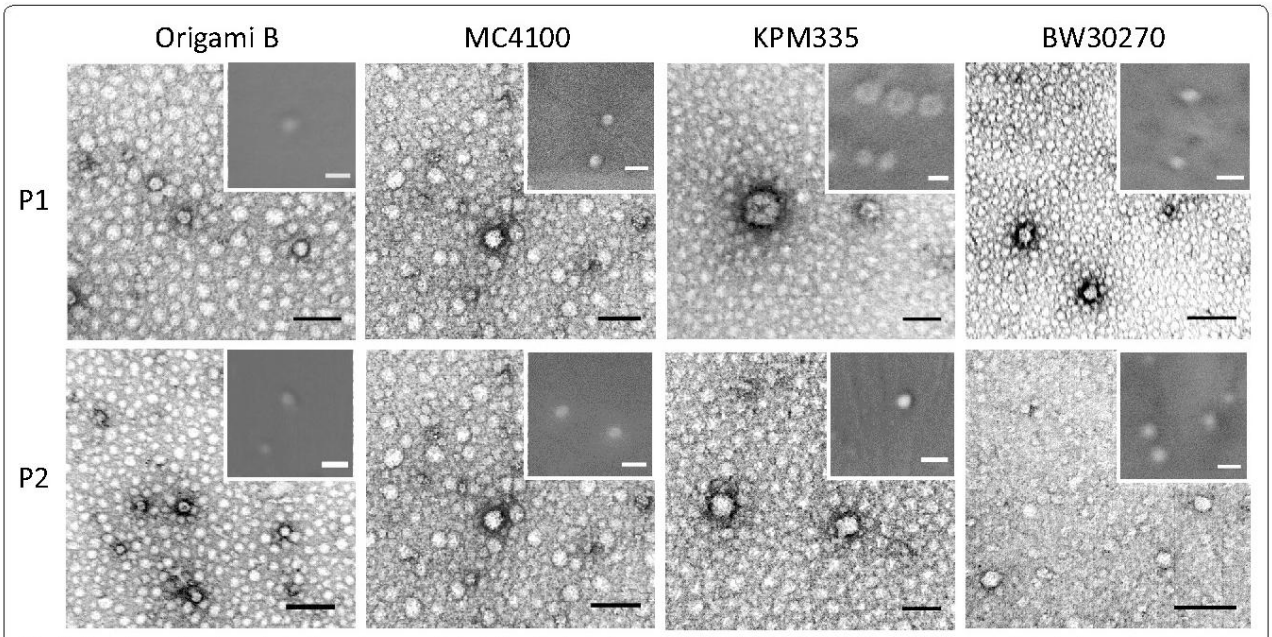


Fig. 4 Morphometric characterization of nanoparticles ultrastructure. Representative TEM and FESEM (insets) images of all strains and main elution peaks 1 and 2. *Scale bars* represent 50 nm

(Fig. 5b). These nanoparticles accumulated intracellularly in the perinuclear region, showing moderate nuclear penetrability (Fig. 5c).

The performance of P2 nanoparticles, those showing higher cell penetrability in vitro, was evaluated in vivo through analyzing their biodistribution and accumulation in CXCR4-overexpressing tumor tissues, in a subcutaneous mouse model of colorectal cancer. Upon systemic administration, all protein variants targeted and accumulated in the primary tumor, albeit to different extents (Fig. 6a–c), with materials produced in MC4100 being the most efficient. Nanoparticles produced in KPM335 showed results similar in efficacy to those manufactured in Origami B, the paradigmatic cell factory for T22-GFP-H6 [17], and slightly better than the parental BW30270 (although differences were not significant). In any case, the higher cell penetrability of these nanoparticles determined in cultured cell (Fig. 5a)

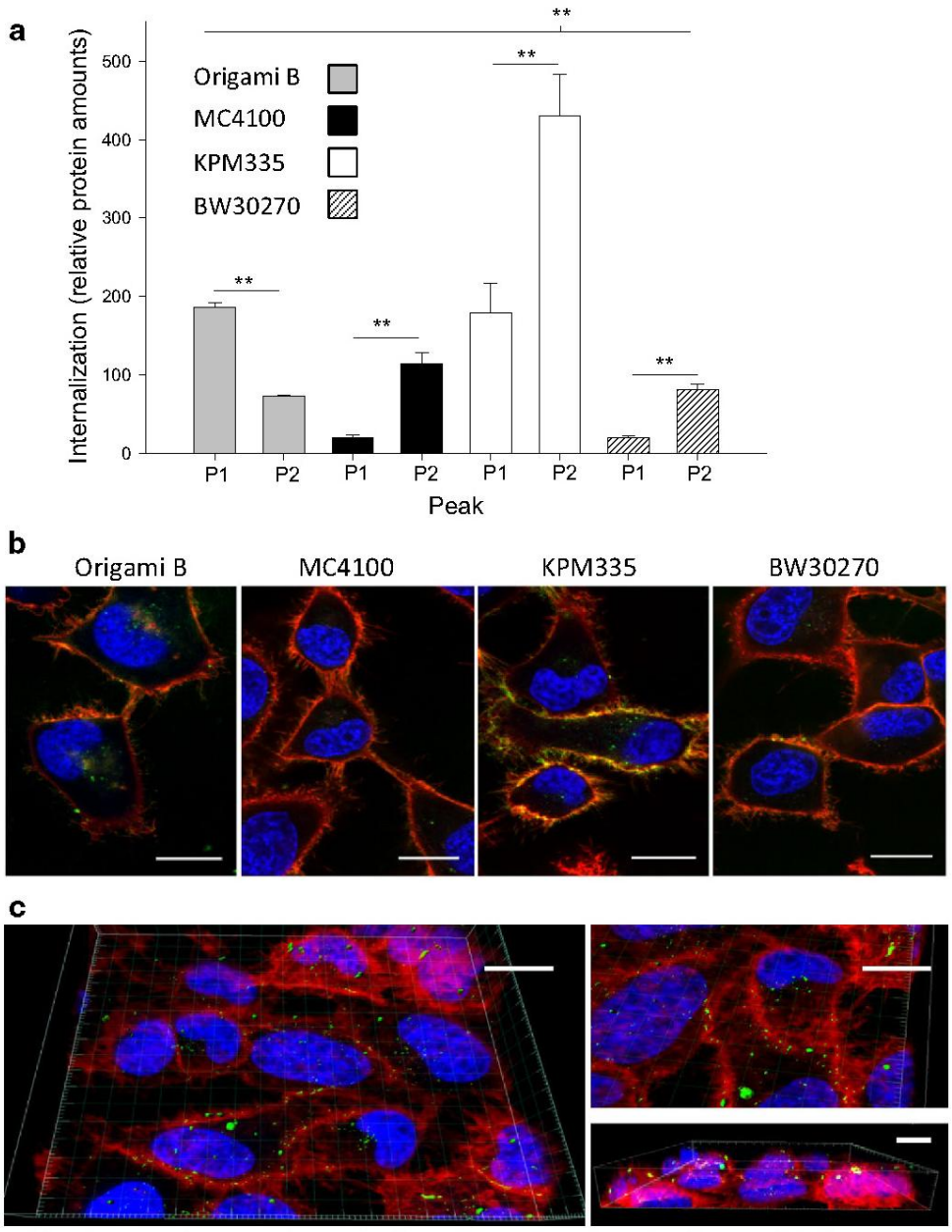
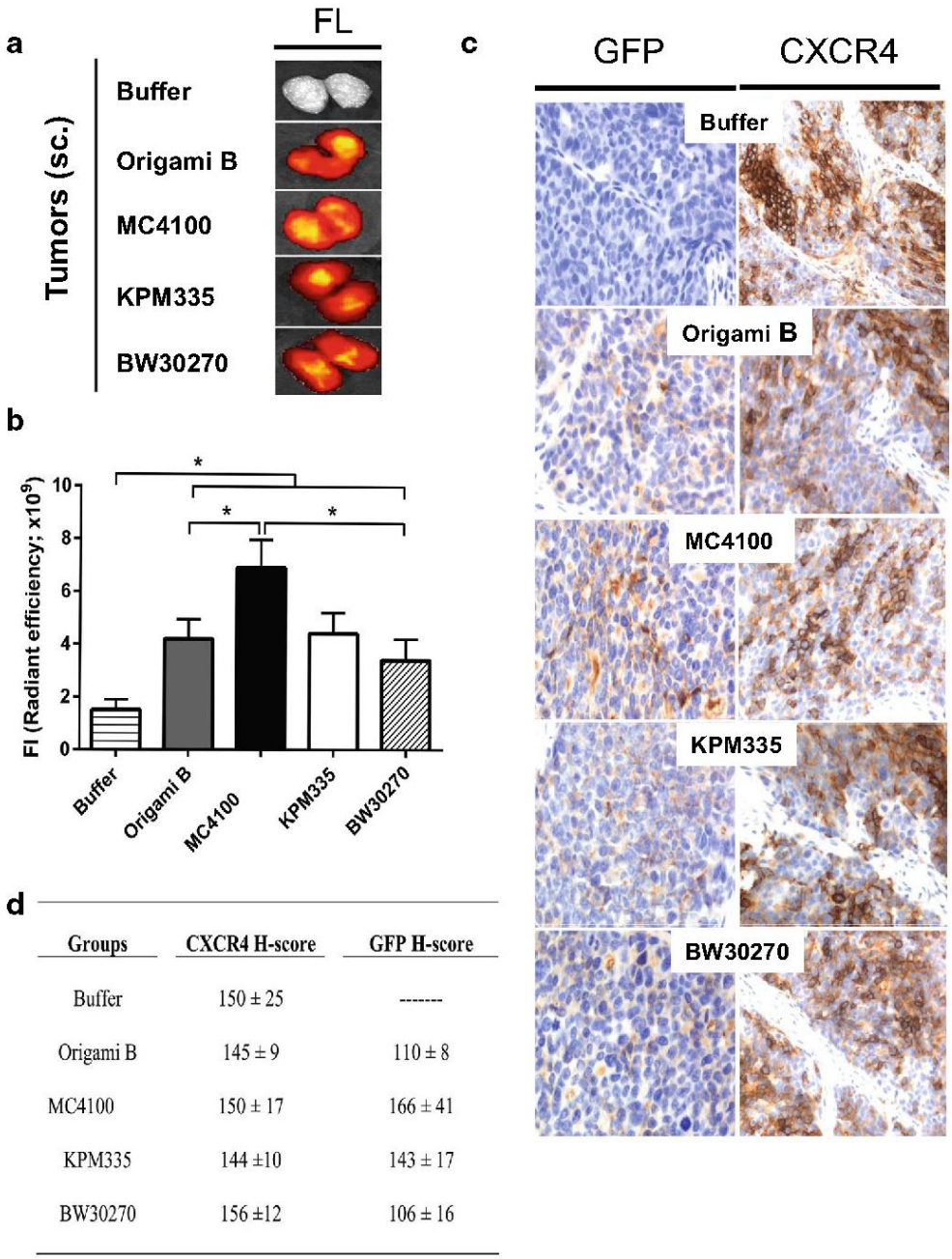


Fig. 5 Cell penetrability of nanoparticles. **a** Internalization of self-assembling nanoparticles analyzed by intracellular fluorescence of HeLa cells incubated with 25 nM of T22-GFP-H6 for 1 h. Internalize fluorescence (determined by flow cytometry) is corrected by specific fluorescence (determined by fluorimetry). The *quotient* represents relative protein amounts. **b** Internalization of nanoparticles visualized by confocal microscopy. Nucleus (*blue*) of HeLa cells were labeled using Hoechst 33342, membranes (*red*) were labeled with CellMask Deep Red. Cells were incubated with 25 nM of T22-GFP-H6 (*green*) for 1 h. *Bars* indicate 15 μm. **c** Three dimensional 3D reconstruction of T22-GFP-H6 nanoparticles (P2) from KPM335 into HeLa cells. Images were obtained using 35 layers per image. Three different fields of confocal microscopy as well as three angles were viewed through Imaris Bitplane Software. *Scale bars* indicate 15 μm and symbols mean significant differences: $^{**}p < 0.001$; $^{*}p < 0.05$

was not reflected by an enhanced tumor targeting in vivo. As expected, accumulation of protein nanoparticles was coincident with the overexpression of CXCR4 in target tissues (Fig. 6d, e), and none of the protein materials was observed to accumulate in the evaluated normal non-target organs (namely kidney, lung, heart, liver, spleen and pancreas; Fig. 7a, b). The absence of a signal in kidney was indicative of a high stability in vivo of the nanoparticles that keep assembled in oligomers of size over the cut-off of renal clearance (~8 nm).



d

Groups	CXCR4 H-score	GFP H-score
Buffer	150 ± 25	-------
Origami B	145 ± 9	110 ± 8
MC4100	150 ± 17	166 ± 41
KPM335	144 ±10	143 ± 17
BW30270	156 ±12	106 ± 16

Fig. 6 In vivo biodistribution of nanoparticles in tumor tissue. **a** Protein accumulation determined by ex vivo imaging of GFP-emitted fluorescence by representative subcutaneous colorectal cancer tumors grown in mice (measurements performed 5 h after the intravenous administration of a 500 µg nanoparticle dose). **b** Quantitation of GFP-emitted fluorescence in tumors of the compared groups, expressed as total radiant efficiency (photon/s/cm^2/sr/µW/cm^2). **c** Representative micrographs of nanoparticle internalization, as detected by IHC with an anti-GFP antibody, in the tumor cell cytosol and CXCR4 expression observed in tumor cells before nanoparticle injection in representative tumors. **d** Mean ± S.E. of CXCR4 expression H-score in tumor tissue (measured prior to nanoparticle administration) and GFP H-score (a measure of nanoparticle internalization in tumor cells) of mice belonging to the Origami B, MC4100, KPM335, BW30270 or buffer-treated groups. Note the similar level of CXCR4 expression in tumors among groups and the correlation between tumor-emitted fluorescence and the amount of nanoparticle internalized in tumor cells

Although the final performance of KPM335 materials in vivo was excellent and indistinguishable from that of Origami B-derived nanoparticles, the larger nanoparticle size observed in the P1 fraction (Figs. 3a, 4), and particularly the higher cell penetrability of the P2 fraction in vitro were indicative of either a different conformation

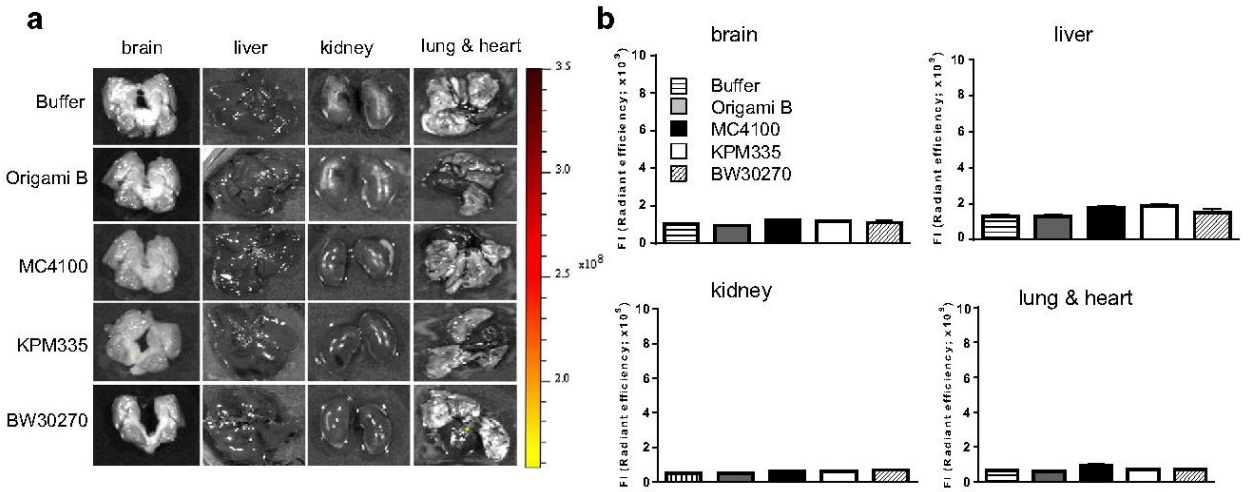


Fig. 7 Nanoparticle biodistribution in normal tissues. **a** Representative ex vivo images of T22-GFP-H6 nanoparticles uptake in mouse brain, liver, kidney, lung or heart tissue after injection as measured by GFP-emitted fluorescence. **b** Quantitation of GFP fluorescence expressed as radiant efficiency. Note the negligible level of detected fluorescence, being undistinguishable from the background fluorescence observed in vehicle-treated mice. This was in contrast to the high level of nanoparticle accumulation observed in tumor tissues in the same animals (see Fig. 6)

of T22-GFP-H6 building blocks, a diversity in the architecture of nanoparticles produced in this particular strain, or both. Conformational protein variations were indeed identified by CD of P2 protein fractions (Fig. 8a), distinguishing KPM335- and BW30270-derived materials as the extremes of a conformational spectrum. Since BW30270 is the endotoxin-containing, parental strain of the endotoxin-free KPM335 [13], the results suggest that genetic tailoring associated with LPS modification might have affected cell properties related to protein metabolism.

In a previous study [20], we have demonstrated that mutations in the *E. coli* genome that affect protein quality control activities (chaperones and proteases) impact on the architecture and function of produced protein nanoparticles [20]. Also, the growth of cultured KPM335 is slightly slower than that of the parental strain BW30270 [15]. Reduced bacterial growth, as promoted, for instance, by minimal defined media or by entering into the stationary phase, results in the modification of intracellular levels of chaperones such as increased levels of ClpB and IbpA and reduced levels of GrpE, HtpG and Hsc [26]. In this context, we wondered if the LPS engineering of KPM335 might have resulted in a physiological, indirect impact on protein processing by a modified cell factory through modification of chaperone balance. Transcriptomic analysis of main folding-assistant genes in KPM335 revealed significant differences regarding the parental BW30270 (Fig. 8b), and in some cases a pattern of expression that might be comparable to that found in proteomics of wild-type bacteria at the end of the exponential growth phase (Fig. 8b). In particular, *tig* transcript levels are lower being the production of trigger factor associated with ribosomal protein expression and growth rate [26]. On the other hand, some heat shock protein genes are upregulated, in particular those involved in inclusion body disaggregation (*clpB* and *ibpA*) [27, 28] indicating enhanced aggregation in the endotoxin-free strain when comparing with the parental background. In this regard, the formation of inclusion bodies by an aggregation-prone protein is slightly favoured in KPM335 when compared to BW30270, as measured by the percentage of insoluble recombinant protein over the total protein yield (86.7 ± 1.2 versus 81.649 ± 1.4 %; data calculated from [15]). Also, ClpB and IbpA also respond, independently of protein production, to growth rate. They are present in higher amounts in stationary phase at least when cells are growing in complex medium and the growth rate declines gradually [26]. On the other hand, some heat shock protein genes are downregulated (e.g. *grpE*), that is again in agreement with the endotoxin-free strain growing slower, as there is a slight tendency that these heat shock proteins do not change with growth rate or they are present in lower amounts when the cells grow slower or approach stationary phase [26].

Conclusions

In summary, the set of seven different mutations in LPS biosynthesis genes of KPM335 promotes a reduction of the cell growth rate that impact, intrinsically, on the

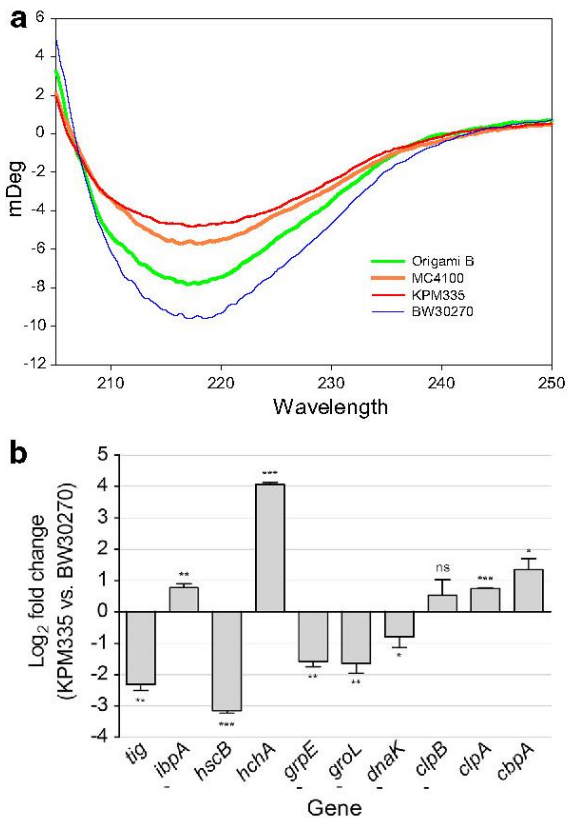


Fig. 8 Conformational protein status and activity of the cell's quality control. **a** Smoothed CD spectra from 250 to 205 nm of T22-GFP-H6 produced by different *E. coli* strains and purified at higher imidazol concentration (P2). Beta sheet structure signal is detected at 218 nm. **b** Expression of chaperone genes in *E. coli* KPM335 compared to their expression levels in *E. coli* BW30270. Relative changes in expression of target genes as quantified by RT-qPCR using the comparative $2^{-\Delta\Delta Ct}$ method [25] and the *ihfB* gene for normalization were converted into log2 values to display the fold changes in log2 scale. The values represent the means and standard deviations based on duplicate RT-qPCR runs for each of the three independent biological replicates per strain. The statistical significance of the differential expression patterns was analyzed using the paired t test. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ns, not significant. Genes labelled with a dash are those whose expression change (either up- or down-regulation) is coincident with the proteomic analysis of a conventional *E. coli* strain when entering into the stationary phase (according to data from [26]). The *tig, hscB, hchA* and *cbpA* gene products were not monitored in this previous study

expression of several among the main chaperones of the *E. coli* quality control network. This may have indirect influences on the conformational status of recombinant aggregation-prone protein materials and their building blocks. Although a more complete systems level analysis of KPM335 performance as a cell factory would be desirable, the set of data presented here indicates that LPS engineering secondarily affects the arm of the cells' biosynthetic and maintenance machineries dealing with protein conformation. While this fact does not negatively affect the final quality (tumor-homing) and organization (self-assembling) of complex smart materials such as CXCR4-targeted protein nanoparticles, it indeed influences some of their relevant properties at the macroscopic level.

Authors' contributions

FR performed most of the experiments regarding protein production and in vitro characterization. MVC performed the in vivo experiments and its linked analysis while AS-C executed the electron microscopy observations, JS-F parts of confocal microscopy and flow cytometry, MP the CD analysis, UU protein purification and UM the quantitative PCR. NF-M and EV designed most of the experimental set-up, UR analysed and discussed the proteomics and transcriptomics data, RM coordinated the in vivo experimental while EG-F and AV conceived the whole study. All authors have participated in the preparation of the manuscript (text and figures), that has been mostly written by AV. All authors read and approved the final manuscript.

Author details

[1] Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, Cerdanyola del Vallès, 08193 Barcelona, Spain. [2] Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Bellaterra, Cerdanyola del Vallès, 08193 Barcelona, Spain. [3] CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, Cerdanyola del Vallès, 08193 Barcelona, Spain. [4] Biomedical Research Institute Sant Pau (IIB-SantPau) and Josep Carreras Leukemia Research Institute, Hospital de la Santa Creu i Sant Pau, 08025 Barcelona, Spain. [5] Servei de Microscòpia, Universitat Autònoma de Barcelona, Bellaterra, Cerdanyola del Vallès, 08193 Barcelona, Spain. [6] Leibniz University of Hannover, Technical Chemistry & Life Science, Hannover, Germany. [7] Helmholtz Centre for Infection Research, Inhoffenstraße 7, Brunswick, Germany. [8] Division of Structural Biochemistry, Priority Area Asthma and Allergy, Research Center Borstel, Airway Research Center North (ARCN), Member of the German Center for Lung Research (DZL), 23845 Borstel, Germany. [9] Present Address: Cibbim-Nanomedicine, Hospital Vall d'Hebron, Passeig de la Vall d'Hebron, 119-129, 08035 Barcelona, Spain. [10] Present Address: Department of Ruminant Production, Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Torre Marimon, Caldes de Montbui, 08140 Barcelona, Spain.

Acknowledgements

Protein production has been partially performed by the ICTS "NANBIOSIS", more specifically by the Protein Production Platform of CIBER in Bioengineering, Biomaterials & Nanomedicine (CIBER-BBN)/IBB, at the UAB (http://www.nanbiosis.es/unit/u1-protein-production-platform-ppp/). We are indebted to MINECO BIO2013-41019-P, AGAUR (2014SGR-132) and CIBER de Bioingeniería, Biomateriales y Nanomedicina (project NANOPROTHER) to AV, Plan Estatal de I+D+I 2013-2016 del Instituto de Salud Carlos III (FEDER co-funding) FIS PI12/00327 to EV, Marató TV3 416/C/2013-2030 to RM for funding our research. We thank the CIBER-BBN Nanotoxicology Unit for fluorescent in vivo follow-up using the IVIS equipment. The authors thank Manuel Hein and Dörte Grella (Research Center Borstel) and also Fran Cortes the Cell Culture and Citometry Units of the Servei de Cultius Cel·lulars, Producció d'Anticossos i Citometria (SCAC), Servei de Proteómica and to the Servei de Microscòpia, for technical assistance, and the Soft Materials Service (ICMAB-CSIC/CIBER-BBN). FR and MP acknowledge financial support from "Francisco Jose de Caldas" Scholarship program of COLCIENCIAS (Colombia) and Universitat Autònoma de Barcelona through pre-doctoral fellowships respectively. UU received a Sara Borrell postdoctoral fellowship from ISCIII. AV received an ICREA ACADEMIA award. Strain KPM335 was kindly provided by Research Corporation Technologies, Tucson, AZ.

Competing interests

UU, EV, MVC, AV, NF-M and RM are co-authors of a patent (WO2012095527) covering the use of T22 as an intracellular targeting agent. No other potential competing interests have been identified.

Received: 21 January 2016 Accepted: 28 March 2016
Published online: 08 April 2016

References

1. Rodriguez-Carmona E, Villaverde A. Nanostructured bacterial materials for innovative medicines. Trends Microbiol. 2010;18:423–30.
2. Villaverde A. Nanotechnology, bionanotechnology and microbial cell factories. Microb Cell Fact. 2010;9:53.
3. Chen GQ. New challenges and opportunities for industrial biotechnology. Microb Cell Fact. 2012;11:111.
4. Vazquez E, Villaverde A. Microbial biofabrication for nanomedicine: biomaterials, nanoparticles and beyond. Nanomedicine (Lond). 2013;8:1895–8.
5. Taguchi S, Ooi T, Mizuno K, Matsusaki H. Advances and needs for endotoxin-free production strains. Appl Microbiol Biotechnol. 2015;99:9349–60.
6. Dor-On E, Solomon B. Targeting glioblastoma via intranasal administration of Ff bacteriophages. Front Microbiol. 2015;6:530.
7. Wakelin SJ, Sabroe I, Gregory CD, Poxton IR, Forsythe JL, Garden OJ, et al. "Dirty little secrets"—endotoxin contamination of recombinant proteins. Immunol Lett. 2006;106:1–7.
8. Gao B, Tsan MF. Endotoxin contamination in recombinant human heat shock protein 70 (Hsp70) preparation is responsible for the induction of tumor necrosis factor alpha release by murine macrophages. J Biol Chem. 2003;278:174–9.
9. Beenken-Rothkopf LN, Karfeld-Sulzer LS, Zhang X, Kissler H, Michie SA, Kaufman DB, et al. Protein polymer hydrogels: effects of endotoxin on biocompatibility. J Biomater Appl. 2013;28:395–406.
10. Pepys MB, Hawkins PN, Kahan MC, Tennent GA, Gallimore JR, Graham D, et al. Proinflammatory effects of bacterial recombinant human C-reactive protein are caused by contamination with bacterial products, not by C-reactive protein itself. Circ Res. 2005;97:e97–103.
11. Corchero JL, Gasser B, Resina D, Smith W, Parrilli E, Vazquez F, et al. Unconventional microbial systems for the cost-efficient production of high-quality protein therapeutics. Biotechnol Adv. 2013;31:140–53.
12. Sanchez-Garcia L, Martin L, Mangues R, Ferrer-Miralles N, Vazquez E, Villaverde A. Recombinant pharmaceuticals from microbial cells: a 2015 update. Microb Cell Fact. 2016;15:33.
13. Mamat U, Wilke K, Bramhill D, Schromm AB, Lindner B, Kohl TA, et al. Detoxifying *Escherichia coli* for endotoxin-free production of recombinant proteins. Microb Cell Fact. 2015;14:57.
14. Seras-Franzoso J, Tatkiewicz WI, Vazquez E, Garcia-Fruitos E, Ratera I, Veciana J, et al. Integrating mechanical and biological control of cell proliferation through bioinspired multieffector materials. Nanomedicine (Lond). 2015;10:873–91.
15. Rueda F, Cano-Garrido O, Mamat U, Wilke K, Seras-Franzoso J, Garcia-Fruitos E, et al. Production of functional inclusion bodies in endotoxin-free Escherichia coli. Appl Microbiol Biotechnol. 2014;98:9229–38.
16. Unzueta U, Cespedes MV, Ferrer-Miralles N, Casanova I, Cedano J, Corchero JL, et al. Intracellular CXCR4(+) cell targeting with T22-empowered protein-only nanoparticles. Int J Nanomedicine. 2012;7:4533–44.
17. Cespedes MV, Unzueta U, Tatkiewicz W, Sanchez-Chardi A, Conchillo-Sole O, Alamo P, et al. In vivo architectonic stability of fully de novo designed protein-only nanoparticles. ACS Nano. 2014;8:4166–76.
18. Kim J, Takeuchi H, Lam ST, Turner RR, Wang HJ, Kuo C, et al. Chemokine receptor CXCR4 expression in colorectal cancer patients increases the risk for recurrence and for poor survival. J Clin Oncol. 2005;23:2744–53.
19. Unzueta U, Ferrer-Miralles N, Cedano J, Zikung X, Pesarrodona M, Saccardo P, et al. Non-amyloidogenic peptide tags for the regulatable self-assembling of protein-only nanoparticles. Biomaterials. 2012;33:8714–22.
20. Rueda F, Cespedes MV, Conchillo-Sole O, Sanchez-Chardi A, Seras-Franzoso J, Cubarsi R, et al. Bottom-up instructive quality control in the biofabrication of smart protein materials. Adv Mater. 2015;27:7816–22.
21. Mamat U, Meredith TC, Aggarwal P, Kuhl A, Kirchhoff P, Lindner B, et al. Single amino acid substitutions in either YhjD or MsbA confer viability to 3-deoxy-d-*manno*-oct-2-ulosonic acid-depleted *Escherichia coli*. Mol Microbiol. 2008;67:633–48.
22. Sambrook J, Fritsch E, Maniatis T. Molecular Cloning, A Laboratory Manual, Cold Spring Harbor. New York: Cold Spring Harbor Laboratory Press; 1989.
23. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem. 1976;72:248–54.
24. Laemmli UK. Cleavage of Structural Proteins During Assembly of Head of Bacteriophage-T4. Nature. 1970;227:680.
25. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25:402–8.
26. Li Z, Nimtz M, Rinas U. The metabolic potential of *Escherichia coli* BL21 in defined and rich medium. Microb Cell Fact. 2014;13:45.
27. Weibezahn J, Bukau B, Mogk A. Unscrambling an egg: protein disaggregation by AAA + proteins. Microb Cell Fact. 2004;3:1.
28. Rinas U, Hoffmann F, Betiku E, Estape D, Marten S. Inclusion body anatomy and functioning of chaperone-mediated in vivo inclusion body disassembly during high-level recombinant protein production in *Escherichia coli*. J Biotechnol. 2007;127:244–57.

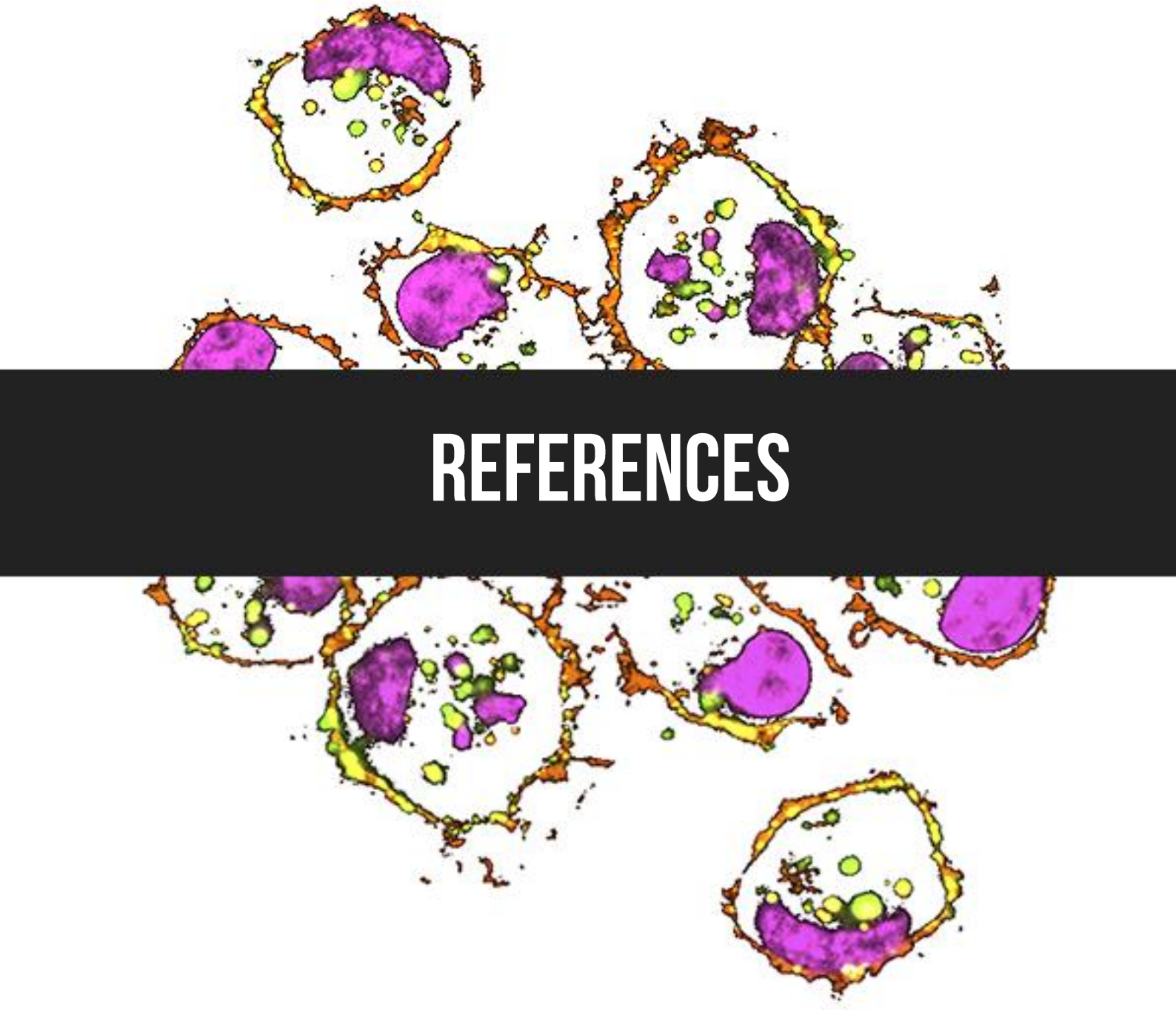
ANNEX VI

Other publications:

- **Pesarrodona M,** Unzueta U, Vázquez E. **Dialysis: a characterisation method of aggregation tendency**. Methods Mol Biol. 2015;1258:321-30.

- Xu, Z., Céspedes, M.V., Unzueta, U., Álamo, P., **Pesarrodona, M**., Mangues, R., Vázquez, E., Villaverde, A., Ferrer-Miralles, N. **Targeting low-density lipoprotein receptors with protein-only nanoparticles.** J Nanopart Res. 2015; 17: 150.

- Serna N, Céspedes MV, Saccardo P, Xu Z, Unzueta U, Álamo P, **Pesarrodona M**, Sánchez-Chardi A, Roldán M, Mangues R, Vázquez E, Villaverde A, Ferrer-Miralles N. **Rational engineering of single-chain polypeptides into protein-only, BBB-targeted nanoparticles.** Nanomedicine. 2016;12(5):1241-51.



REFERENCES

1 Wagner, V., Dullaart, A., Bock, A. K. & Zweck, A., The emerging nanomedicine landscape. *Nat.Biotechnol.* **2006,** *24* (10), 1211-1217.

2 Bawa, R., Bawa, S. R., Maebius, S. B., Flynn, T. & Wei, C., Protecting new ideas and inventions in nanomedicine with patents. *Nanomedicine : nanotechnology, biology, and medicine* **2005,** *1* (2), 150-158.

3 Tinkle, S. *et al.*, Nanomedicines: addressing the scientific and regulatory gap. *Ann.N.Y.Acad.Sci.* **2014,** *1313*, 35-56.

4 Bobo, D., Robinson, K. J., Islam, J., Thurecht, K. J. & Corrie, S. R., Nanoparticle-Based Medicines: A Review of FDA-Approved Materials and Clinical Trials to Date. *Pharm.Res.* **2016**.

5 Knudsen, K. B. *et al.*, In vivo toxicity of cationic micelles and liposomes. *Nanomedicine : nanotechnology, biology, and medicine* **2015,** *11* (2), 467-477.

6 Naahidi, S. *et al.*, Biocompatibility of engineered nanoparticles for drug delivery. *J Control Release* **2013,** *166* (2), 182-194.

7 Frohlich, E., Cellular targets and mechanisms in the cytotoxic action of non-biodegradable engineered nanoparticles. *Curr Drug Metab* **2013,** *14* (9), 976-988.

8 Vazquez, E., Mangues, R. & Villaverde, A., Functional recruitment for drug delivery through protein-based nanotechnologies. *Nanomedicine (London, England)* **2016,** *11* (11), 1333-1336.

9 Craik, D. J., Fairlie, D. P., Liras, S. & Price, D., The future of peptide-based drugs. *Chem.Biol.Drug Des* **2013,** *81* (1), 136-147.

10 Salatin, S., Jelvehgari, M., Maleki-Dizaj, S. & Adibkia, K., A sight on protein-based nanoparticles as drug/gene delivery systems. *Ther Deliv* **2015,** *6* (8), 1017-1029.

11 Ferlay, J. S., I; Ervik, M; Dikshit, R; Eser, S; Mathers, C; Rebelo, M; Parkin, DM; Forman, D; Bray, F. *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]*, (2013).

12 Sharp, P. A. & Langer, R., Research agenda. Promoting convergence in biomedical science. *Science* **2011,** *333* (6042), 527.

13 Chou, L. Y., Ming, K. & Chan, W. C., Strategies for the intracellular delivery of nanoparticles. *Chem.Soc.Rev.* **2011,** *40* (1), 233-245.

14 Schroeder, A. *et al.*, Treating metastatic cancer with nanotechnology. *Nat.Rev.Cancer* **2012,** *12* (1), 39-50.

15 Maeda, H., Macromolecular therapeutics in cancer treatment: the EPR effect and beyond. *J.Control Release* **2012,** *164* (2), 138-144.

16 Rink, J. S., Plebanek, M. P., Tripathy, S. & Thaxton, C. S., Update on current and potential nanoparticle cancer therapies. *Curr Opin Oncol* **2013,** *25* (6), 646-651.

17 Srinivasan, M., Rajabi, M. & Mousa, S. A., Multifunctional Nanomaterials and Their Applications in Drug Delivery and Cancer Therapy. *Nanomaterials* **2015,** *5*, 1690-1703.

18 Friedman, A. D., Claypool, S. E. & Liu, R., The smart targeting of nanoparticles. *Curr.Pharm.Des* **2013,** *19* (35), 6315-6329.

19 Glasgow, M. D. & Chougule, M. B., Recent Developments in Active Tumor Targeted Multifunctional Nanoparticles for Combination Chemotherapy in Cancer Treatment and Imaging. *J.Biomed.Nanotechnol.* **2015,** *11* (11), 1859-1898.

20 Seleci, M., Ag Seleci, D. & Joncyzk, R., Smart multifunctional nanoparticles in nanomedicine. *BioNanoMaterials* **2016,** *17 (1-2)*, 33-41.

21 Peer, D., Harnessing RNAi nanomedicine for precision therapy. *Mol.Cell Ther.* **2014,** *2*, 5.

22 Mao, Z. W., Zhou, X. Y. & Gao, C. Y., Influence of structure and properties of colloidal biomaterials on cellular uptake and cell functions. *Biomater Sci-Uk* **2013,** *1* (9), 896-911.
23 Ferrer-Miralles, N. *et al.*, Biological activities of histidine-rich peptides; merging biotechnology and nanomedicine. *Microb.Cell Fact.* **2011,** *10*, 101.
24 Kumar, S., Harrison, N., Richards-Kortum, R. & Sokolov, K., Plasmonic nanosensors for imaging intracellular biomarkers in live cells. *Nano.Lett.* **2007,** *7* (5), 1338-1343.
25 Uherek, C., Fominaya, J. & Wels, W., A modular DNA carrier protein based on the structure of diphtheria toxin mediates target cell-specific gene delivery. *The Journal of biological chemistry* **1998,** *273* (15), 8835-8841.
26 Liang, W. L. & Lam, J. K. W., Endosomal Escape Pathways for Non-Viral Nucleic Acid Delivery Systems. *Molecular Regulation of Endocytosis* **2012**, 429-456.
27 Kang, B., Mackey, M. A. & El-Sayed, M. A., Nuclear targeting of gold nanoparticles in cancer cells induces DNA damage, causing cytokinesis arrest and apoptosis. *J.Am.Chem.Soc.* **2010,** *132* (5), 1517-1519.
28 Field, L. D., Delehanty, J. B., Chen, Y. & Medintz, I. L., Peptides for specifically targeting nanoparticles to cellular organelles: quo vadis? *Acc.Chem.Res.* **2015,** *48* (5), 1380-1390.
29 Siegel, R. L., Miller, K. D. & Jemal, A., Cancer statistics, 2016. *CA Cancer J.Clin.* **2016,** *66* (1), 7-30.
30 Merlot, A. M., Kalinowski, D. S. & Richardson, D. R., Unraveling the mysteries of serum albumin-more than just a serum protein. *Front Physiol* **2014,** *5*, 299.
31 Ferrara, N., Hillan, K. J., Gerber, H.-P. & Novotny, W., Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer. *Nat Rev Drug Discov* **2004,** *3* (5), 391-400.
32 Cunningham, D. *et al.*, Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. *N Engl J Med* **2004,** *351* (4), 337-345.
33 Ricardo, S. *et al.*, Breast cancer stem cell markers CD44, CD24 and ALDH1: expression distribution within intrinsic molecular subtype. *J.Clin.Pathol.* **2011,** *64* (11), 937-946.
34 Zoller, M., CD44: can a cancer-initiating cell profit from an abundantly expressed molecule? *Nat.Rev.Cancer* **2011,** *11* (4), 254-267.
35 Sampieri, K. & Fodde, R., Cancer stem cells and metastasis. *Semin.Cancer Biol.* **2012,** *22* (3), 187-193.
36 Goodison, S., Urquidi, V. & Tarin, D., CD44 cell adhesion molecules. *Mol.Pathol.* **1999,** *52* (4), 189-196.
37 Zhang, S. S. *et al.*, CD133(+)CXCR4(+) colon cancer cells exhibit metastatic potential and predict poor prognosis of patients. *BMC.Med.* **2012,** *10*, 85.
38 Unzueta, U. *et al.*, Intracellular CXCR4(+) cell targeting with T22-empowered protein-only nanoparticles. *Int.J.Nanomedicine.* **2012,** *7*, 4533-4544.
39 Foekens, J. A. *et al.*, Prognostic value of CD44 variant expression in primary breast cancer. *Int.J.Cancer* **1999,** *84* (3), 209-215.
40 Choi, K. Y., Saravanakumar, G., Park, J. H. & Park, K., Hyaluronic acid-based nanocarriers for intracellular targeting: interfacial interactions with proteins in cancer. *Colloids Surf.B Biointerfaces.* **2012,** *99*, 82-94.
41 Wang, L. *et al.*, CD44 antibody-targeted liposomal nanoparticles for molecular imaging and therapy of hepatocellular carcinoma. *Biomaterials* **2012,** *33* (20), 5107-5114.

42 Qian, C. *et al.*, Suppression of pancreatic tumor growth by targeted arsenic delivery with anti-CD44v6 single chain antibody conjugated nanoparticles. *Biomaterials* **2013,** *34* (26), 6175-6184.

43 Zhang, D. *et al.*, A CD44 specific peptide developed by phage display for targeting gastric cancer. *Biotechnol.Lett.* **2015,** *37* (11), 2311-2320.

44 Zhang, D. *et al.*, Screening and Identification of a Phage Display Derived Peptide That Specifically Binds to the CD44 Protein Region Encoded by Variable Exons. *J.Biomol.Screen.* **2016,** *21* (1), 44-53.

45 Jalkanen, S. & Jalkanen, M., Lymphocyte CD44 binds the COOH-terminal heparin-binding domain of fibronectin. *J Cell Biol* **1992,** *116* (3), 817-825.

46 Yasuda, T. *et al.*, COOH-terminal heparin-binding fibronectin fragment induces nitric oxide production in rheumatoid cartilage through CD44. *Rheumatology (Oxford)* **2004,** *43* (9), 1116-1120.

47 Hibino, S. *et al.*, Laminin alpha5 chain metastasis- and angiogenesis-inhibiting peptide blocks fibroblast growth factor 2 activity by binding to the heparan sulfate chains of CD44. *Cancer Res* **2005,** *65* (22), 10494-10501.

48 Hibino, S. *et al.*, Identification of an active site on the laminin alpha5 chain globular domain that binds to CD44 and inhibits malignancy. *Cancer Res* **2004,** *64* (14), 4810-4816.

49 Park, H.-Y., Lee, K.-J., Lee, S.-J. & Yoon, M.-Y., Screening of peptides bound to breast cancer stem cell specific surface marker CD44 by phage display. *Mol Biotechnol* **2012,** *51* (3), 212-220.

50 Congote, L. F., Sadvakassova, G., Dobocan, M. C., Difalco, M. R. & Kriazhev, L., Biological activities and molecular interactions of the C-terminal residue of thrombospondin-4, an epitome of acidic amphipathic peptides. *Peptides* **2010,** *31* (4), 723-735.

51 Sadvakassova, G., Dobocan, M. C. & Congote, L. F., Osteopontin and the C-terminal peptide of thrombospondin-4 compete for CD44 binding and have opposite effects on CD133+ cell colony formation. *BMC Res Notes* **2009,** *2*, 215.

52 Alves, C. S., Yakovlev, S., Medved, L. & Konstantopoulos, K., Biomolecular characterisation of CD44-fibrin(ogen) binding: distinct molecular requirements mediate binding of standard and variant isoforms of CD44 to immobilized fibrin(ogen). *The Journal of biological chemistry* **2009,** *284* (2), 1177-1189.

53 Choi, J. *et al.*, Comparison of cytotoxic and inflammatory responses of photoluminescent silicon nanoparticles with silicon micron-sized particles in RAW 264.7 macrophages. *J.Appl.Toxicol.* **2009,** *29* (1), 52-60.

54 Albanese, A., Tang, P. S. & Chan, W. C., The effect of nanoparticle size, shape, and surface chemistry on biological systems. *Annu.Rev.Biomed.Eng* **2012,** *14*, 1-16.

55 Baetke, S. C., Lammers, T. & Kiessling, F., Applications of nanoparticles for diagnosis and therapy of cancer. *Br.J.Radiol.* **2015,** *88* (1054), 20150207.

56 Medina-Kauwe, L. K., Xie, J. & Hamm-Alvarez, S., Intracellular trafficking of nonviral vectors. *Gene Ther.* **2005,** *12* (24), 1734-1751.

57 Langston Suen, W. L. & Chau, Y., Size-dependent internalisation of folate-decorated nanoparticles via the pathways of clathrin and caveolae-mediated endocytosis in ARPE-19 cells. *J.Pharm.Pharmacol.* **2014,** *66* (4), 564-573.

58 Oh, N. & Park, J. H., Endocytosis and exocytosis of nanoparticles in mammalian cells. *Int.J.Nanomedicine.* **2014,** *9 Suppl 1*, 51-63.

59 Murugan, K. *et al.*, Parameters and characteristics governing cellular internalisation and trans-barrier trafficking of nanostructures. *Int.J.Nanomedicine.* **2015,** *10*, 2191-2206.
60 Jiang, W., Kim, B. Y. S., Rutka, J. T. & Chan, W. C. W., Nanoparticle-mediated cellular response is size-dependent. *Nature Nanotechnology* **2008,** *3*, 145-150.
61 Sahay, G., Alakhova, D. Y. & Kabanov, A. V., Endocytosis of nanomedicines. *J.Control Release* **2010,** *145* (3), 182-195.
62 Lee, H., Fonge, H., Hoang, B., Reilly, R. M. & Allen, C., The effects of particle size and molecular targeting on the intratumoral and subcellular distribution of polymeric nanoparticles. *Mol Pharm* **2010,** *7* (4), 1195-1208.
63 Blanco, E., Shen, H. & Ferrari, M., Principles of nanoparticle design for overcoming biological barriers to drug delivery. *Nat.Biotechnol.* **2015,** *33* (9), 941-951.
64 Qiu, Y. *et al.*, Surface chemistry and aspect ratio mediated cellular uptake of Au nanorods. *Biomaterials* **2010,** *31* (30), 7606-7619.
65 Li, Y., Kroger, M. & Liu, W. K., Shape effect in cellular uptake of PEGylated nanoparticles: comparison between sphere, rod, cube and disk. *Nanoscale.* **2015,** *7* (40), 16631-16646.
66 Gratton, S. E. *et al.*, The effect of particle design on cellular internalisation pathways. *Proc.Natl.Acad.Sci.U.S.A* **2008,** *105* (33), 11613-11618.
67 Geng, Y. *et al.*, Shape effects of filaments versus spherical particles in flow and drug delivery. *Nat.Nanotechnol.* **2007,** *2* (4), 249-255.
68 Devarajan, P. V. *et al.*, Particle shape: a new design parameter for passive targeting in splenotropic drug delivery. *J.Pharm.Sci.* **2010,** *99* (6), 2576-2581.
69 Cano-Garrido, O. *et al.*, CXCR4(+)-targeted protein nanoparticles produced in the food-grade bacterium Lactococcus lactis. *Nanomedicine (London, England)* **2016,** *11* (18), 2387-2398.
70 Lynch, I. & Dawson, K., Protein-nanoparticle interactions. *Nano Today* **2008,** *3*, 40-47.
71 Jallouli, Y., Paillard, A., Chang, J., Sevin, E. & Betbeder, D., Influence of surface charge and inner composition of porous nanoparticles to cross blood-brain barrier in vitro. *Int.J.Pharm.* **2007,** *344* (1-2), 103-109.
72 Lin, I. C., Liang, M., Liu, T. Y., Monteiro, M. J. & Toth, I., Cellular transport pathways of polymer coated gold nanoparticles. *Nanomedicine.* **2012,** *8* (1), 8-11.
73 Bruckman, M. A., Czapar, A. E., VanMeter, A., Randolph, L. N. & Steinmetz, N. F., Tobacco mosaic virus-based protein nanoparticles and nanorods for chemotherapy delivery targeting breast cancer. *J Control Release* **2016,** *231*, 103-113.
74 Rajagopal, K. & Schneider, J. P., Self-assembling peptides and proteins for nanotechnological applications. *Curr.Opin.Struct.Biol.* **2004,** *14* (4), 480-486.
75 Drexler, K. E., Molecular engineering: An approach to the development of general capabilities for molecular manipulation. *Proc.Natl.Acad.Sci.U.S.A* **1981,** *78* (9), 5275-5278.
76 Janin, J., Bahadur, R. P. & Chakrabarti, P., Protein-protein interaction and quaternary structure. *Q.Rev.Biophys.* **2008,** *41* (2), 133-180.
77 King, N. P. *et al.*, Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **2012,** *336* (6085), 1171-1174.
78 Radomska, A., Leszczyszyn, J. & Radomski, M. W., The Nanopharmacology and Nanotoxicology of Nanomaterials: New Opportunities and Challenges. *Adv.Clin.Exp.Med.* **2016,** *25* (1), 151-162.

79 Lee, E. J., Lee, N. K. & Kim, I. S., Bioengineered protein-based nanocage for drug delivery. *Adv.Drug Deliv.Rev.* **2016**.

80 Molino, N. M. & Wang, S. W., Caged protein nanoparticles for drug delivery. *Curr.Opin.Biotechnol.* **2014,** *28*, 75-82.

81 Giacca, M. & Zacchigna, S., Virus-mediated gene delivery for human gene therapy. *J.Control Release* **2012,** *161* (2), 377-388.

82 Lee, L. A. & Wang, Q., Adaptations of nanoscale viruses and other protein cages for medical applications. *Nanomedicine.* **2006,** *2* (3), 137-149.

83 Mellado, M. C. *et al.*, Impact of physicochemical parameters on in vitro assembly and disassembly kinetics of recombinant triple-layered rotavirus-like particles. *Biotechnol.Bioeng.* **2009,** *104* (4), 674-686.

84 Padilla, J. E., Colovos, C. & Yeates, T. O., Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc.Natl.Acad.Sci.U.S.A* **2001,** *98* (5), 2217-2221.

85 Gradisar, H. *et al.*, Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. *Nat.Chem.Biol.* **2013,** *9* (6), 362-366.

86 Peternel, S. & Komel, R., Active protein aggregates produced in Escherichia coli. *Int J Mol Sci* **2011,** *12* (11), 8275-8287.

87 Gaynon, P. S., Primary treatment of childhood acute lymphoblastic leukemia of non-T cell lineage (including infants). *Hematol Oncol Clin North Am* **1990,** *4* (5), 915-936.

88 Hamley, I. W., Peptide fibrillisation. *Angew.Chem.Int.Ed Engl.* **2007,** *46* (43), 8128-8147.

89 Sutter, M. *et al.*, Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nat Struct Mol Biol* **2008,** *15* (9), 939-947.

90 Yeates, T. O., Crowley, C. S. & Tanaka, S., Bacterial microcompartment organelles: protein shell structure and evolution. *Annu Rev Biophys* **2010,** *39*, 185-205.

91 Zhen, Z. *et al.*, RGD-modified apoferritin nanoparticles for efficient drug delivery to tumors. *ACS Nano* **2013,** *7* (6), 4830-4837.

92 Miele, E., Spinelli, G. P., Miele, E., Tomao, F. & Tomao, S., Albumin-bound formulation of paclitaxel (Abraxane ABI-007) in the treatment of breast cancer. *Int.J.Nanomedicine.* **2009,** *4*, 99-105.

93 Ferrer-Miralles, N. *et al.*, Engineering protein self-assembling in protein-based nanomedicines for drug delivery and gene therapy. *Crit Rev.Biotechnol.* **2015,** *35* (2), 209-221.

94 Lamm, M. S., Rajagopal, K., Schneider, J. P. & Pochan, D. J., Laminated morphology of nontwisting beta-sheet fibrils constructed via peptide self-assembly. *J.Am.Chem.Soc.* **2005,** *127* (47), 16692-16700.

95 Lopez, D. L. P. *et al.*, De novo designed peptide-based amyloid fibrils. *Proc.Natl.Acad.Sci.U.S.A* **2002,** *99* (25), 16052-16057.

96 Riley, J. M., Aggeli, A., Koopmans, R. J. & McPherson, M. J., Bioproduction and characterisation of a pH responsive self-assembling peptide. *Biotechnol Bioeng* **2009,** *103* (2), 241-251.

97 Robson, M. H. & Kros, A., Self-assembly of coiled coils in synthetic biology: inspiration and progress. *Angew.Chem.Int.Ed Engl.* **2010,** *49* (17), 2988-3005.

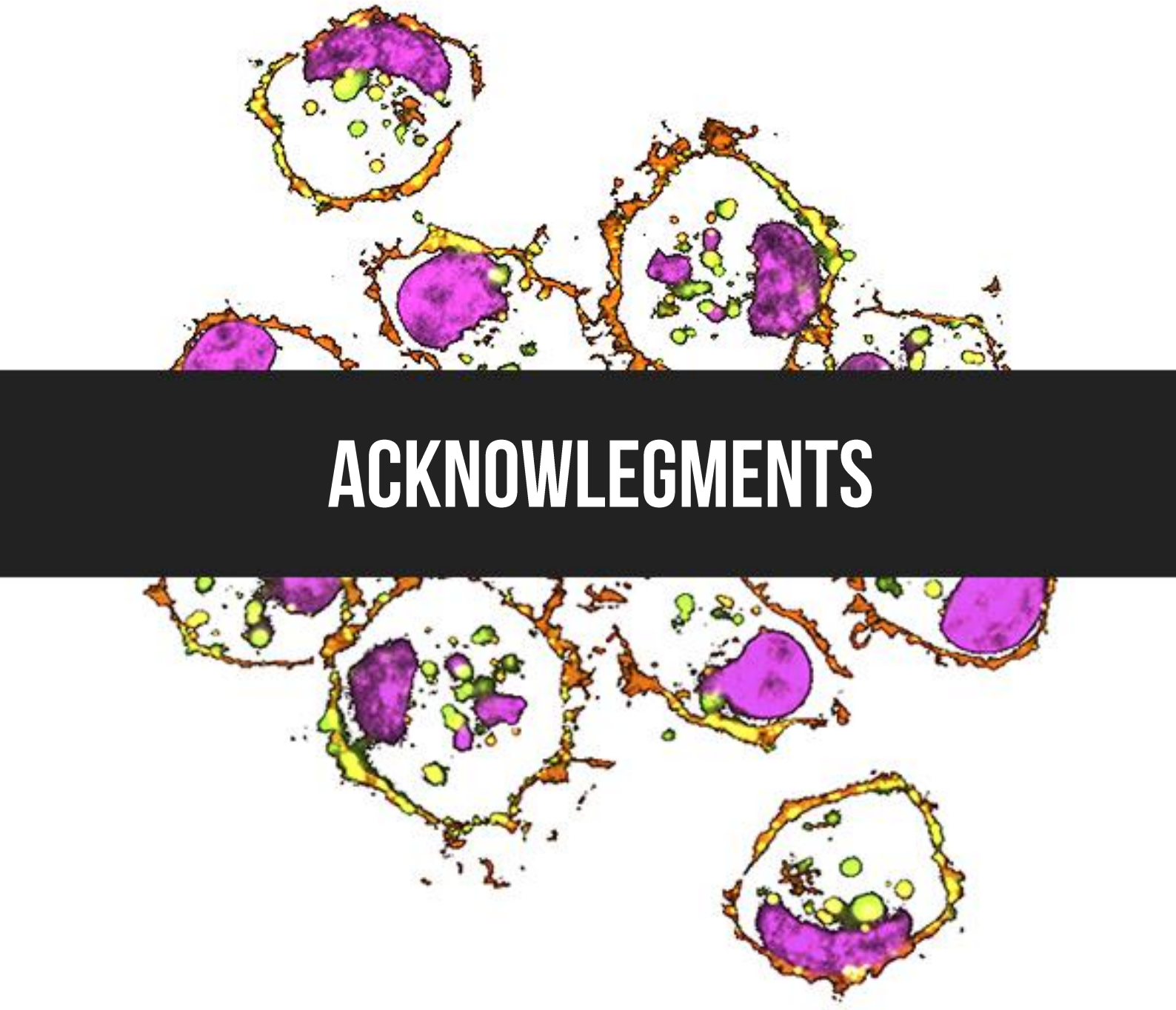
98 Raman, S., Machaidze, G., Lustig, A., Aebi, U. & Burkhard, P., Structure-based design of peptides that self-assemble into regular polyhedral nanoparticles. *Nanomedicine.* **2006,** *2* (2), 95-102.

99 Joh, N. H. *et al.*, De novo design of a transmembrane Zn(2)(+)-transporting four-helix bundle. *Science* **2014,** *346* (6216), 1520-1524.
100 Boyle, A. L. *et al.*, Squaring the circle in peptide assembly: from fibers to discrete nanostructures by de novo design. *J.Am.Chem.Soc.* **2012,** *134* (37), 15457-15467.
101 Lai, Y. T. *et al.*, Structure of a designed protein cage that self-assembles into a highly porous cube. *Nat.Chem.* **2014,** *6* (12), 1065-1071.
102 Lai, Y.-T., Tsai, K.-L., Sawaya, M. R., Asturias, F. J. & Yeates, T. O., Structure and flexibility of nanoscale protein cages designed by symmetric self-assembly. *J Am Chem Soc* **2013,** *135* (20), 7738-7743.
103 Smits, F. C., Buddingh, B. C., van Eldijk, M. B. & van Hest, J. C., Elastin-like polypeptide based nanoparticles: design rationale toward nanomedicine. *Macromol.Biosci.* **2015,** *15* (1), 36-51.
104 Xia, X. X., Wang, M., Lin, Y., Xu, Q. & Kaplan, D. L., Hydrophobic drug-triggered self-assembly of nanoparticles from silk-elastin-like protein polymers for drug delivery. *Biomacromolecules.* **2014,** *15* (3), 908-914.
105 Vazquez, E. *et al.*, Modular protein engineering in emerging cancer therapies. *Curr.Pharm.Des* **2009,** *15* (8), 893-916.
106 Aris, A. & Villaverde, A., Engineering nuclear localisation signals in modular protein vehicles for gene therapy. *Biochem.Biophys.Res.Commun.* **2003,** *304* (4), 625-631.
107 Fominaya, J. & Wels, W., Target cell-specific DNA transfer mediated by a chimeric multidomain protein. Novel non-viral gene delivery system. *J.Biol.Chem.* **1996,** *271* (18), 10560-10568.
108 Liu, T. F. *et al.*, Diphtheria toxin-epidermal growth factor fusion protein and Pseudomonas exotoxin-interleukin 13 fusion protein exert synergistic toxicity against human glioblastoma multiforme cells. *Bioconjug.Chem.* **2003,** *14* (6), 1107-1114.
109 Woodworth, T. G. & Nichols, J. C., Recombinant fusion toxins--a new class of targeted biologic therapeutics. *Cancer Treat.Res.* **1993,** *68*, 145-160.
110 Vazquez, E. *et al.*, Protein nanodisk assembling and intracellular trafficking powered by an arginine-rich (R9) peptide. *Nanomedicine.(Lond)* **2010,** *5* (2), 259-268.
111 Unzueta, U. *et al.*, Improved performance of protein-based recombinant gene therapy vehicles by tuning downstream procedures. *Biotechnol.Prog.* **2013,** *29* (6), 1458-1463.
112 Sanchez-Garcia, L. *et al.*, Recombinant pharmaceuticals from microbial cells: a 2015 update. *Microbial cell factories* **2016,** *15*, 33.
113 Kobayashi, N. *et al.*, Self-Assembling Nano-Architectures Created from a Protein Nano-Building Block Using an Intermolecularly Folded Dimeric de Novo Protein. *J Am Chem Soc* **2015,** *137* (35), 11285-11293.
114 Hsia, Y. *et al.*, Design of a hyperstable 60-subunit protein icosahedron. *Nature* **2016,** *535* (7610), 136-139.
115 Rosano, G. L. & Ceccarelli, E. A., Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol* **2014,** *5*, 172.
116 Baneyx, F. & Mujacic, M., Recombinant protein folding and misfolding in Escherichia coli. *Nat.Biotechnol.* **2004,** *22* (11), 1399-1408.
117 Dobson, C. M., Protein folding and misfolding. *Nature* **2003,** *426* (6968), 884-890.
118 Hartl, F. U., Bracher, A. & Hayer-Hartl, M., Molecular chaperones in protein folding and proteostasis. *Nature* **2011,** *475* (7356), 324-332.
119 Langer, T. *et al.*, Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature* **1992,** *356* (6371), 683-689.

120 Kuczynska-Wisnik, D. *et al.*, The Escherichia coli small heat-shock proteins IbpA and IbpB prevent the aggregation of endogenous proteins denatured in vivo during extreme heat shock. *Microbiology* **2002,** *148* (Pt 6), 1757-1765.
121 Lee, S. *et al.*, The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state. *Cell* **2003,** *115* (2), 229-240.
122 Hoskins, J. R., Pak, M., Maurizi, M. R. & Wickner, S., The role of the ClpA chaperone in proteolysis by ClpAP. *Proc Natl Acad Sci U S A* **1998,** *95* (21), 12135-12140.
123 Porankiewicz, J., Wang, J. & Clarke, A. K., New insights into the ATP-dependent Clp protease: Escherichia coli and beyond. *Mol Microbiol* **1999,** *32* (3), 449-458.
124 Goldberg, A. L., Moerschell, R. P., Chung, C. H. & Maurizi, M. R., ATP-dependent protease La (lon) from Escherichia coli. *Methods Enzymol* **1994,** *244*, 350-375.
125 Garcia-Fruitos, E. *et al.*, Divergent genetic control of protein solubility and conformational quality in Escherichia coli. *Journal of molecular biology* **2007,** *374* (1), 195-205.
126 Sorensen, H. P. & Mortensen, K. K., Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microbial cell factories* **2005,** *4*.
127 Gonzalez-Montalban, N., Natalello, A., Garcia-Fruitos, E., Villaverde, A. & Doglia, S. M., In situ protein folding and activation in bacterial inclusion bodies. *Biotechnol Bioeng* **2008,** *100* (4), 797-802.
128 Gonzalez-Montalban, N., Garcia-Fruitos, E. & Villaverde, A., Recombinant protein solubility - does more mean better? *Nat Biotechnology* **2007,** *25*, 718-720.
129 Martínez-Alonso, M., Garcia-Fruitos, E. & Villaverde, A., Yield, solubility and conformational quality of soluble proteins are not simultaneously favored in recombinant Escherichia coli. *Biotechnol.Bioeng.* **2008,** *101*, 1353-1358.
130 Cano-Garrido, O. *et al.*, Supramolecular organisation of protein-releasing functional amyloids solved in bacterial inclusion bodies. *Acta Biomater* **2013,** *9* (4), 6134-6142.
131 Ramón, A., Señorale-Pose, M. & Marín, M., Inclusion bodies: not that bad…. *Front Microbiol* **2014,** *5*, 56.
132 Ventura, S. & Villaverde, A., Protein quality in bacterial inclusion bodies. *Trends Biotechnol* **2006,** *24* (4), 179-185.
133 Jurgen, B. *et al.*, Quality control of inclusion bodies in Escherichia coli. *Microbial cell factories* **2010,** *9*, 41.
134 Singh, A., Upadhyay, V., Upadhyay, A. K., Singh, S. M. & Panda, A. K., Protein recovery from inclusion bodies of Escherichia coli using mild solubilisation process. *Microbial cell factories* **2015,** *14*, 41.
135 Tsumoto, K., Ejima, D., Kumagai, I. & Arakawa, T., Practical considerations in refolding proteins from inclusion bodies. *Protein expression and purification* **2003,** *28* (1), 1-8.
136 Jevsevar, S. *et al.*, Production of nonclassical inclusion bodies from which correctly folded protein can be extracted. *Biotechnol Prog* **2005,** *21* (2), 632-639.
137 Yang, Y., Ringler, P., Muller, S. A. & Burkhard, P., Optimizing the refolding conditions of self-assembling polypeptide nanoparticles that serve as repetitive antigen display systems. *J Struct Biol* **2012,** *177* (1), 168-176.
138 Mamat, U. *et al.*, Detoxifying Escherichia coli for endotoxin-free production of recombinant proteins. *Microbial cell factories* **2015,** *14*, 57.
139 Gao, B. & Tsan, M.-F., Endotoxin contamination in recombinant human heat shock protein 70 (Hsp70) preparation is responsible for the induction of tumor necrosis

factor alpha release by murine macrophages. *The Journal of biological chemistry* **2003,** *278* (1), 174-179.
140 Szalay, L. *et al.*, Estradiol improves cardiac and hepatic function after trauma-hemorrhage: role of enhanced heat shock protein expression. *Am J Physiol Regul Integr Comp Physiol* **2006,** *290* (3), R812-818.
141 Taguchi, S., Ooi, T., Mizuno, K. & Matsusaki, H., Advances and needs for endotoxin-free production strains. *Applied microbiology and biotechnology* **2015,** *99* (22), 9349-9360.
142 Brune, K. D. *et al.*, Plug-and-Display: decoration of Virus-Like Particles via isopeptide bonds for modular immunisation. *Sci Rep* **2016,** *6*, 19234.
143 Feng, B., LaPerle, J. L., Chang, G. & Varma, M. V. S., Renal clearance in drug discovery and development: molecular descriptors, drug transporters and disease state. *Expert Opin Drug Metab Toxicol* **2010,** *6* (8), 939-952.
144 Mastrobattista, E., van der Aa, M., Hennink, W. & Crommelin, D., Artificial viruses: a nanotechnological approach to gene delivery. *Nat Rev Drug Discov* **2006,** *5*, 115-121.
145 Gonzalez-Angulo, A. M., Hennessy, B. T. J. & Mills, G. B., Future of personalized medicine in oncology: a systems biology approach. *J Clin Oncol* **2010,** *28* (16), 2777-2783.
146 Duncan, R. & Gaspar, R., Nanomedicine(s) under the microscope. *Mol Pharm* **2011,** *8* (6), 2101-2141.
147 Daher, N. *et al.*, [Intra-diverticular tumors of the bladder. Natural history and management perspectives]. Tumeurs intra-diverticulaires de la vessie. Histoire naturelle et perspectives d'approche. *Ann Urol (Paris)* **1989,** *23* (4), 275-280.
148 Thapa, R. & Wilson, G. D., The Importance of CD44 as a Stem Cell Biomarker and Therapeutic Target in Cancer. *Stem Cells Int* **2016,** *2016*, 2087204.
149 Poglazova, M. N. & Mashkovtseva, A. V., [A fluorescent-cytochemical method for selective detection of yeast mitochondrial DNA]. Fluorestsentno-tsitokhimicheskii metod izbiratel'nogo vyiavleniia mitokhondrial'noi DNK drozhzhei. *Mikrobiologiia* **1990,** *59* (6), 1024-1031.
150 Villaverde, A. *et al.*, A cell adhesion peptide from foot-and-mouth disease virus can direct cell targeted delivery of a functional enzyme. *Biotechnol Bioeng* **1998,** *59* (3), 294-301.
151 Serna, N. *et al.*, Rational engineering of single-chain polypeptides into protein-only, BBB-targeted nanoparticles. *Nanomedicine.* **2016,** *12* (5), 1241-1251.
152 Elzoghby, A. O., Samy, W. M. & Elgindy, N. A., Albumin-based nanoparticles as potential controlled release drug delivery systems. *J Control Release* **2012,** *157* (2), 168-182.
153 Vijayendra, S. V. N. & Shamala, T. R., Film forming microbial biopolymers for commercial applications--a review. *Crit Rev Biotechnol* **2014,** *34* (4), 338-357.
154 Park, T. J., Lee, K. G. & Lee, S. Y., Advances in microbial biosynthesis of metal nanoparticles. *Applied microbiology and biotechnology* **2016,** *100* (2), 521-534.
155 Villaverde, A., Nanotechnology, bionanotechnology and microbial cell factories. *Microbial cell factories* **2010,** *9*, 53.
156 Seras-Franzoso, J. *et al.*, A nanostructured bacterial bioscaffold for the sustained bottom-up delivery of protein drugs. *Nanomedicine (London, England)* **2013,** *8* (10), 1587-1599.

157 Carbonell, X. & Villaverde, A., Protein aggregated into bacterial inclusion bodies does not result in protection from proteolytic digestion. *Biotechnology Letters* **2002,** *24* (23), 1939-1944.
158 Peternel, S. & Komel, R., Isolation of biologically active nanomaterial (inclusion bodies) from bacterial cells. *Microbial cell factories* **2010,** *9*, 66.
159 Peternel, S., Jevsevar, S., Bele, M., Gaberc-Porekar, V. & Menart, V., New properties of inclusion bodies with implications for biotechnology. *Biotechnol Appl Biochem* **2008,** *49* (Pt 4), 239-246.
160 Korc, M. & Friesel, R. E., The role of fibroblast growth factors in tumor growth. *Curr Cancer Drug Targets* **2009,** *9* (5), 639-651.
161 Xu, Z. K. *et al.*, Formulating tumor-homing peptides as regular nanoparticles enhances receptor-mediated cell penetrability. *Mater Lett* **2015,** *154*, 140-143.
162 Decuzzi, P. *et al.*, Size and shape effects in the biodistribution of intravascularly injected particles. *Journal of Controlled Release* **2010,** *141* (3), 320-327.
163 Hirn, S. *et al.*, Particle size-dependent and surface charge-dependent biodistribution of gold nanoparticles after intravenous administration. *Eur J Pharm Biopharm* **2011,** *77* (3), 407-416.
164 Osborne, O. J. *et al.*, Organ-Specific and Size-Dependent Ag Nanoparticle Toxicity in Gills and Intestines of Adult Zebrafish. *Acs Nano* **2015,** *9* (10), 9573-9584.
165 Tan, J. F., Shah, S., Thomas, A., Ou-Yang, H. D. & Liu, Y. L., The influence of size, shape and vessel geometry on nanoparticle distribution. *Microfluid Nanofluid* **2013,** *14* (1-2), 77-87.
166 Toy, R., Peiris, P. M., Ghaghada, K. B. & Karathanasis, E., Shaping cancer nanomedicine: the effect of particle shape on the in vivo journey of nanoparticles. *Nanomedicine* **2014,** *9* (1), 121-134.
167 Powers, K. W., Carpinone, P. L. & Siebein, K. N., Characterisation of nanomaterials for toxicological studies. *Methods Mol Biol* **2012,** *926*, 13-32.
168 Chithrani, B. D., Ghazani, A. A. & Chan, W. C. W., Determining the size and shape dependence of gold nanoparticle uptake into mammalian cells. *Nano Letters* **2006,** *6* (4), 662-668.
169 Herd, H. *et al.*, Nanoparticle Geometry and Surface Orientation Influence Mode of Cellular Uptake. *Acs Nano* **2013,** *7* (3), 1961-1973.



ACKNOWLEGMENTS

Aquesta tesis ha estat el producte de gairebé cinc anys de treball, intensos però molt gratificants. I és que, m'han servit, com diria un amic, per sortir del capullo i convertir-me en papallona en l'àmbit professional. N'he aprés que l'experiència s'obté a base d'errors i que això fa que el camí sigui excitant, enriquidor i que l'èxit sigui molt més dolç. Ha estat una experiència boníssima en molts aspectes i és per això que vull agrair a totes aquelles persones que ho han fet possible i que m'han ajudat des del primer a l'últim dia.

Als meus tutors, estic immensament agraïda per que m'hagueu obert les portes i hagueu apostat per mi en un inici sense conèixer-me. Orgullosa és com em sento quan em demanen "Com va per el laboratori?" doncs és un orgull poder treballar en un grup competent, productiu, que es preocupa per els integrants i que desprèn molt bon ambient. Tot això, en bona part, és mèrit vostre.

Toni, dins de la meva incertesa de quin serà el següent pas després de la tesis, si una cosa tinc clara és que et trobaré a faltar. No sempre un es troba amb un "jefe" que t'escolta, confia amb tu, et motiva i es preocupa. Per tot això, gràcies!

Esther, tú me has ayudado en el día a día tanto a nivel profesional como personal. ¿Qué hubiese sido de esta tesis sin nuestras TO DO LISTS? Eres pura energía y la que, aun teniendo suficiente jaleo en casa, nos guías y te preocupas de todos nosotros. Has sido esencial para esta tesis como también lo eres para el laboratorio. Gracias :)

Neus, no només m'has aconsellat quan no trobava la manera de purificar una puteina. Has sigut tu qui m'ha ajudat tots aquests anys amb una part molt important de la tesis, la docència. Per mi, ets, en molts aspectes, un exemple. T'estic eternament agraïda!

Ugutz: mi mentor (y el de muchos jejejej), te debo tanto! Me has enseñado todo lo aprendido a nivel experimental como también la actitud, incluso a coger las puntas de las pipetas por orden :) Gracias por ayudarme, darme consejos, idees....soy muy fan tuya!

Paolo: eres el más considerado y la alma de la fiesta! Gracias por ser siempre el primero que se levanta para echar una mano cuando se necesita. Espero poder bachatear y cantar "Moulin Rouge" siempre contigo :)

Olivia: la meva confident i una amiga! Gracies per aconsellar-me i donar-me suport dins i fora el laboratori, en els bons i ens els pitjors moments. M'ho he passat teta organitzant la fondue amb tu i a totes les escapades que hem fet amb el Paolo. Espero que n'hi hagin moltes més!

Rosa: se t'hauria de regalar la lluna per tot el que fas al laboratori. Moltes gràcies per facilitar-nos la feina!

Pepe: no sé si apostaría por las bolas magnéticas pero sí lo haría por darte un papel más influyente en los vídeos de la fondue, eres un actorazo!

Elena: Aquest últims anys te trobat a faltar doncs és un plaer treballar amb tu i compartir “despatx”. Sempre m’has ajudat i animat a tirar endavant. Moltíssimes gràcies.

Laura: qui haguera dit quan erets estudiant de grau que al cap d’uns anys continuaries el PENTRI? Ets un tresor, m’has ajudat moltíssim i sempre amb un somriure. Però no només a mi si no a tot el laboratori. Et desitjo lo millor.

Alex: Si una cosa troba a faltar de les pràctiques són les sessions a microscòpia. Saps dibuixar-li un somriure a tothom. Gràcies per dibuixar-me’l cada cop que et veig i per totes les coses que he après amb tu.

Quim: m’ha agradat molt treballar amb tu dins i fora del laboratori, m’has ensenyat maneres de fer i com treballar amb cossos. Gràcies per tot això i per les birres que hem compartit junts.

Fabian: mi amol! En 4 años uno obtiene el doctorado, tú en cambio, doctor, papa, y esposo. Has demostrado tener una energía y una fuerza enorme. Gracias por tu ayuda y tu cariño.

Naroa: Gracias por todo lo aprendido trabajando contigo y por las cervezas y la fiesta fuera del labo. Eres pura salsa!

Marianna: Qué arte tienes! Qué alegría! Que suerte tenerte en el laboratorio! Has sido mi compañera en la guerra contra la purificación de f...ing puteins con DNA y tendencia a precipitar. Seguiremos luchando!

Cada dia em desperto pensant en la sort que he tingut de poder treballar durant aquests anys amb un ambient tant acollidor i enriquidor. Fran i Olga, “the culture duo”, moltes gràcies tant per l’ajuda com per els moment fora de l’institut. Volia agrair a aquelles persones fora del grup que m’han guiat i aconsellat en moments que no sabia què fer: Gràcies a la Sandra Villegas, al Xavier Serra, a la Manu, a la Sílvia i a la Mònica Roldán.

Thank you Rosemary to give this thesis a new look! :)

Eva, tu em vas animar a enviar el proposal, m’has ajudar moltíssim durant tots aquest últims mesos i és per això que et dec pràcticament un terç d’aquesta tesis. Et desitjo lo millor en aquesta nova etapa. Ets un sol.

Vull agrair a tot el grup PENTRI que ha fet possible tirar això endavant, però en especial a la Petra i la Yolanda, ha estat un plaer treballar amb vosaltres.

Uschi and Zhaopeng, I am very grateful to you for inviting me in your group, for taking care of me and teaching me a lot of new things. I am very pleased to meet you and work with you. All the short stay in Hannover was a great experience. Hope we keep in contact

Gràcies MarieEve! Tu ets la que m'has ensenyat més coses sobre purificació de proteïnes i sobre comportament d'aquestes molècules. T'admiro a molts nivells, una persona intel·ligent, treballadora, forta, atenta, simpàtica, guapa....puc seguir però no acabaria.

Oh Toni! Estic molt contenta que haguem iniciat una col·laboració després de tants anys d'amistat, treballar amb tu ha sigut genial. Des del dia en que vam dissenyar el programa de matrius a informàtica, sé que junts podem fer coses increïbles jejeej. T'estimo molt!

Gràcies Anna per ser TU! Per els "breaks" acompanyats de somriure, per les sortides, l'energia que transmets. Gracies a la fondue del 2014 he conegut a una molt bona amiga.

Pablo, un amic, un confident, bé...no sé molt bé com definir-te. M'has escoltat i recolzat durant aquests últims anys. Gràcies per ser-hi!

Com a investigador, moltes vegades és difícil explicar què fas i amb què treballes a una persona d'un àmbit completament allunyat. I més difícil és encara que t'escolti, s'interessi i es preocupi per comptes d'avorrir-se. És una realitat per malament que soni. Però he tingut molta sort de trobar persones que m'han escoltat, inclús m'han ajudat sempre que ho he necessitat. Isabel, Mireia, Nguyet, Bärbel, Marta, Gerard, Dani, Marcel, i segur que em deixo algú, moltes gràcies!

Salva, la persona que més m'ha escoltat, m'ha animat i apostat per mi quan ni tant sols jo ho faria. Et dec moltíssim! Company d'aventures, de desacords, de passions, d'amor...i de tantes altres coses. Gràcies de tot cor!

Per últim, vull agrair immensament a la meva família. M'heu donat tot l'afecte, suport i consell que he necessitat en totes les decisions que he pres. Heu fet tots els sacrificis que heu pogut perquè pogués arribar on sóc i m'heu fet la persona que sóc. Sé que esteu orgullosos, sempre m'ho heu demostrat, però jo també ho estic de tenir uns pares com vosaltres. Papa, sent com ets, des de petita has sigut el meu mirall, m'has despertat la curiositat per aprendre, la lluita per idear i el caràcter de superació. Mama, m'has ensenyat a lluitar i no abaixar mai el cap, a mourem per aconseguir resoldre tots els problemes, ets exemplar. Germanet tu sempre m'has fet sentir important, especial. Buscaves els meus "amulets" de la sort pensant que t'ajudarien a aprovar examens perquè volies assemblar-te a la teva germana, que sàpigues que ara tú ets un exemple per mi. Aquesta tesi és per mi però va dedicada a vosaltres!

Maria, ets una noia collonuda. És una sort tenir-te entre nosaltres. Gràcies per donar-li cara a aquesta tesis :)

Image use from published material in this thesis has been asked for permission.