Engineering protein self-assembling in protein-based nanomedicines for drug delivery and gene therapy

Neus Ferrer-Miralles $^{1,2,3}$, Escarlata Rodríguez-Carmona $^{1,2,3}$, José Luis Corchero $^{1,2,3}$, Elena García-Fruitós $^{1,2,3}$, Esther Vázquez $^{1,2,3}$, Antonio Villaverde $^{1,2,3}$

1 Instituto de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain
2 Department de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain
3 CIBER en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, 08193 Barcelona, Spain

§ Equally contributed.

* Corresponding author: Antonio Villaverde; antoni.villaverde@uab.cat; +34935813086

Keywords: Protein engineering; nanoparticles; drug delivery; non-viral gene therapy; nanomedicine.
Abstract
Lack of targeting and improper biodistribution are major flaws in current drug-based therapies that prevent reaching high local concentration of the therapeutic agent. Such weaknesses impose the administration of high drug doses, resulting in undesired side effects, limited efficacy and enhanced production costs. Currently missing nanosized containers, functionalized for specific cell targeting will be then highly convenient for the controlled delivery of both conventional and innovative drugs. In an attempt to fill this gap, health-focused nanotechnologies put under screening a growing spectrum of materials as potential components of nanocages, whose properties can be tuned during fabrication. However, most of these materials pose severe biocompatibility concerns. We will revise here how proteins, the most versatile functional macromolecules, can be conveniently exploited and adapted by conventional genetic engineering as efficient building blocks of fully compatible nanoparticles for drug delivery, and how selected biological activities can be recruited to mimic viral behavior during infection. Although engineering of protein self-assembling is still excluded from fully rational approaches, the exploitation of protein nano-assemblies occurring in nature and the direct manipulation of protein-protein contacts in bioinspired constructs open intriguing possibilities for further development. These methodologies empower the construction of new and potent vehicles that offer promise as true artificial viruses for efficient and safe nanomedical applications.
Introduction

Nanomedical approaches to drug delivery are aimed to the development of smart nanosized cages with high stability, appropriate pharmacokinetics and efficient cell penetrability. Ideally, any drug delivery system should perform a set of biological activities enabling them to reach an appropriate biodistribution and to deliver the cargo molecules to the appropriate compartment of a target cell type. Success in these activities would be only reached by vehicles overcoming several consecutive biological barriers, at both cellular and organic levels, in the absence of undesired side effects. In this regard, progresses in the design and fabrication of nanocontainers are in part restricted by the severe toxicity exhibited by some of the materials so far explored for nanofabrication and by the limited success in the targeting and delivery events. Different synthetic delivery carriers including cationic liposomes and micelles, polymeric nano-micro-particles, block copolymers, carbon nanotubes, dendrimers and inorganic nanoparticles have been actively developed and studied not only for facilitating targeted drug delivery but also for maximizing therapeutic efficacy (Kang et al., 2012). Being in general biologically inert, these materials, once nanostructured, must be functionalized to gain targeting abilities and other properties necessary for a proper biodistribution such as cell surface receptor binding, membrane crossing and nuclear penetration.

In contrast to engineering approaches involving synthetic materials, natural or modified gene delivery carriers are also offering potential promises and alternatives over synthetic vehicles because they could accomplish most of the main requirements needed for an ideal delivery including biocompatibility, solubility in water and high uptake efficiency (Ma, Nolte, & Cornelissen, 2012). These bioengineered biological entities can be living or non-living bacteria, viral gene vectors, virus-like-particles, virosomes, cell organelles, red blood cells, immune cells and stem cells, among others (Yoo, Irvine, Discher, & Mitragotri, 2011). Among all these vehicles, the majority of clinical trials for gene therapy still use virus-based nanocarriers (Giacca & Zacchigna, 2012). In fact, viruses, as pathogenic nanosized entities, have developed mechanisms and strategies successful in protecting their genomes and associated proteins and transport them through the extracellular environment, evading the host immune system, interacting with specific receptors of target cells and driving nucleic acids to the right cell compartment. These activities are based on the functional domains of viral proteins that have been continuously submitted to selective pressures during interaction with hosts. Therefore, as functional vehicles for targeted delivery, viruses offer superior advantages compared to synthetic constructs. Obviously, residual viral
pathogenic potential, even upon genetic inactivation, is a major concern when considering this category of vehicles and their generalized use.

Far from such biosafety concerns, proteins themselves fulfill the requirement of biocompatibility as they are main structural and functional components of living beings. In addition, proteins, protein domains or shorter polypeptide stretches, peptides, are responsible for most of the cross-molecular interactions, biochemical reactions, transformations and regulation events supporting life. Therefore, the ability of proteins to form supramolecular complexes but also to carry out functions relevant to targeted delivery make them appealing building blocks of nanocages for therapeutic purposes. In addition, they can be produced at industrial scale by recombinant DNA technologies (Ferrer-Miralles, Domingo-Espin, Corchero, Vazquez, & Villaverde, 2009). Many of these functions, especially those responsible for specific intermolecular contacts, have been mapped in surface-exposed regions of polypeptides, proteins, and they can be performed by linear peptides. Unlike complex biological protein production, the straightforward chemical synthesis of peptides benefits from the potential to incorporate unnatural amino acids and to introduce chemical modifications. Also, inter-batch variability is minimized and the quality concerns associated to recombinant protein production, such as conformational stress and midstream and downstream aggregation (Vazquez, Corchero, & Villaverde, 2011), are avoided. The lower functional plasticity, when compared to full-length natural or engineered proteins, might be balanced by the high control of the amino acid sequence and physicochemical properties of the products.

However, protein functions can be also adjusted, tailored in natural polypeptides or in fully de novo designed proteins through conventional genetic engineering. Since late 70’s, the progressive development of cell factories and genetic tools for protein modification and production allowed the small and large scale fabrication of modified proteins as drugs, what has expanded the bio-pharma industries, with almost 200 protein drugs currently in use for human therapies. The enormous functional diversity of proteins in nature includes activities that would be highly convenient in targeted drug delivery, such as binding and condensation of nucleic acids, specific interactions with cell surface receptors, receptor-mediated endocytosis, endosomal escape, cytosolic trafficking and nuclear transport (Vazquez, Ferrer-Miralles, & Villaverde, 2008; Aris & Villaverde, 2004; Ferrer-Miralles, Vazquez, & Villaverde, 2008). In fact, the combination
of such abilities in viral capsids characterizes the virus life cycle and permits the horizontal transmission of both proteins and nucleic acids during infection. Being viruses the main referent in the design of functional protein-based nanoparticles for gene transfer, the artificial virus concept (Douglas, 2008; Mastrobattista, van der Aa, Hennink, & Crommelin, 2006; Wagner, 2004) orbits around the possibility to mimic these abilities in manufactured or recombinant constructs of viral size, devoid of any infectious material and being then chemically and biologically safe. While recruiting desired functions by the combination of active peptides, protein domains or modular full proteins (for instance, as multifunctional proteins) seems to be an affordable task. However engineering self-assembling of the forming building blocks is still excluded from a rational approach (Vazquez & Villaverde, 2010). Indeed, lack of controlled self-assembling as nanoparticles of pre-defined properties is probably the main obstacle to the construction of highly desired protein-only nanoscale shells as biomimetic of natural viruses. The potential therapeutic value that protein-coding nucleic acids as well as non-coding nucleic acids (such as small regulatory RNAs, DNA oligonucleotides, small catalytic RNAs and DNAs, long antisense RNAs, decoy RNAs and aptamers) is gaining, points out nucleic acids as main therapeutic drugs in a next future for a wide spectrum of potential applications (Giacca & Zacchigna, 2012). This fact makes the artificial virus concept (namely, the full adaptation of nanocages to the particular delivery of nucleic acids) highly promising in nanomedicine. Protein self-assembling into nucleic acid-embracing shells is here mandatory, as nucleic acids are highly labile molecules.

Naturally occurring, self-assembling protein nanoparticles

Viruses

Viral life cycle is supported by self-assembling protein nanoparticles that efficiently interact with host cells and deliver their cargo genome to the correct cell compartment. In the gene therapy arena, viruses have been exploited as nucleic acid-containing delivery agents, due to its specificity in cell-surface receptor binding, cell penetrability and efficient nucleic acid delivery (Limberis, 2012; Giacca & Zacchigna, 2012). More than 1800 ongoing or completed gene therapy clinical trials have been recorded, most of them using species from the families Adenoviridae, Parvoviridae, and Retroviridae (Table 1), upon engineering to include the desired transgene and to eliminate their replicative potential (http://www.wiley.com/legacy/wileychi/genmed/clinical/). Adenovirus are linear dsDNA icosahedral non-enveloped virus with a capsid composed by three major capsid proteins and four minor capsid proteins forming viral particles of approximately 90 nm (Martin, 2012; Vellinga, Van der Heijdt, & Hoeben, 2005).
Adenovirus-based vectors have a large packaging capacity and can transduce several cell cultures and tissues in vivo. However, transient transgene expression level depends on the immune response of the host that might be highly activated. Actually, a fatal systemic inflammatory response syndrome was reported in a patient enrolled in a safety gene therapy trial of second generation adenovirus vector, highlighting the limitation of the animal models used in preclinical trials and the need of determining the patient-to-patient immunogenic variability (Raper et al., 2003; Smaglik, 1999). On the other hand, adeno-associated vectors (Parvoviridae) have reduced packaging capacity but can transduce dividing and non-dividing cells, which is quite convenient when treating diseases affecting non-growing tissues. It has been demonstrated that the double strand DNA of adeno-associated vectors can integrate randomly in the genome of the transduced cell (Giacca & Zacchigna, 2012). However, in preclinical and clinical trials performed so far, no integration events have been documented. Finally, Gammaretrovirus and Lentivirus, both genus of the Retroviridae family, are also under exploration for gene therapy purposes. Gammaretrovirus have a large packaging capacity and a high level of stable transgene expression. Lentiviruses show the same properties than their Gammaretrovirus counterparts but can also transduce non-dividing cells. Nevertheless, in both cases, the encapsulated RNA, once reverse transcribed, is integrated in the host genome, which can result in undesired insertional mutagenesis. This has been already disclosed in clinical trials for patients suffering from X-linked severe combined immunodeficiency (SCID-X1). These patients were reinfused with CD34+ cells treated ex vivo with retroviruses containing the cytokine receptor gamma gene. Some of the patients developed leukemia linked to the random insertion of the retroviral DNA (Hacein-Bey-Abina et al., 2003a; Hacein-Bey-Abina et al., 2003b; Howe et al., 2008).

In China, the State Food and Drug administration (SFDA) approved, a few years ago, two viral gene therapy drugs, namely Gendicine (Gao, 2011) and Oncorine (Liang, 2012). Gendicine is an adenovirus vector containing p53 gene that has been already tested in more than 50 types of solid tumours. Oncorine is a conditional replicative oncolytic adenovirus–based vector that only replicates in cancer target cells that are then lysed. Although the results obtained by Chinese gene therapy products based on viral vectors have been met with skepticism by the scientific community (Guo & Xin, 2006), the high number of ongoing clinical trials in this field advancing towards improved methodologies to overcome secondary effects might initiate a new era in the treatment of genetic diseases. The reluctance of USA and European regulatory agencies (FDA and EMA) to approve viral-based drugs for gene therapy (linked to the
adverse effects and to the requirement to demonstrate a significant increase in overall survival rate) was finally broken by the very recent (November, 2012) approval of Glybera (alipogene tiparvovec), that is based on a modified adeno-associated virus (AAV) of serotype 1. This approval derives from a tortuous administrative process involving four revisions by the EMA and by the Committee on Human Medicinal Products (CHMP). Such bottlenecks resulted in the withdrawal of the company that initially presented the product (Amsterdam Molecular Therapeutics), which was finally replaced by UniQure in achieving the approval (Yla-Herttuala, 2012). Glybera is intended to treat a rare disease called lipoprotein lipase deficiency (LPLD), also known as familial hyperchylomicronemia, whose patients suffer recurring acute pancreatitis (http://www.uniQure.com/products/glybera/). Glybera delivers a normal, healthy LPL gene packaged in the AAV-based vector with tropism to muscle cells, the natural LPL producers, and it was tested in three clinical interventional studies conducted in the Netherlands and in Canada, in only 27 LPLD patients. The drug is administered via a one-time series of small intramuscular injections in the legs, and the approval for this prescription is under exceptional circumstances, and only allowed in selected centers of excellence with expertise in treating LPLD. Furthermore, Advexin, an adenovirus-based vector for p53 tumour suppressor therapy, is expected to be approved by the FDA after successfully ended phase III clinical trials.

In the line of safest approaches, bacteriophages have been extensively explored as vehicles as they can display foreign peptides or proteins to target specific cell types (Clark & March, 2006). In fact, phage display technology allows the identification of tissue-homing peptides to target tumors in potentially personalized treatments (Krag et al., 2006), that can be then used to functionalize nanovehicles of different chemical nature or just exposed to phages' surface itself (Khalaj-Kondori, Sadeghizadeh, Behmanesh, Saggio, & Monaci, 2011; Pan et al., 2012b). In fact, conventional genetic engineering can be used to modify the tropism of not only bacteriophages but also animal viruses, opening a door for flexibility and potential design of targets. This should enable the interaction with receptors overexpressed in target tissues that might increase the specificity of the viral vector and reduce its generic toxicity (Kaufmann & Nettelbeck, 2012).

**Virus-like particles**

Virus-like-particles (VLPs) are tuneable nanometric virus coat-protein cages produced in recombinant cells (usually insect cell lines but also microorganisms) through the
controlled expression of one or more cloned structural genes from a given viral species (Palomares, Mena, & Ramirez, 2012). VLPs mimic supramolecular (hierarchical) self-assembly behavior of viruses and are non-infectious, non-replicative and then safer than viral vectors. One or few major viral structural capsid proteins (not necessarily all those forming natural capsids) can spontaneously self-assemble into highly organized, homogeneous and morphological defined nanoparticles with uniform size (ranging from 10 to 1000 nm) and shape distributions (icosahedrons, spheres or tubes) with high stability (Ma, Nolte, & Cornelissen, 2012; Lee, Nguyen, & Wang, 2011). VLPs of diverse viruses such as papillomaviruses, hepatitis B, C and E viruses, polyoma viruses, lentiviruses, rotaviruses, paroviruses, noroviruses, and the particular species CCMV, CPMV, MS2, M13, or Qβ have been so far generated (Lee & Wang, 2006; Ma, Nolte, & Cornelissen, 2012).

Moreover, disassembly and reassembly of VLPs can be controlled in vitro (Mellado et al., 2009) resulting into natural empty “shells”, “hollow scaffolds” or “nano-containers” that are also suitable for ex vivo packaging and loading of therapeutic cargo molecules such as dyes, quantum dots, magnetic nanoparticles, chemicals, foreign proteins and exogenous nucleic acids. In addition, VLPs conserve some unique properties of the natural infective particles such as natural tropism, cellular uptake and intracellular trafficking. A broad range of organs can be targeted with different VLPs, including liver (hepatitis B VLPs), spleen (some papilloma and polyoma VLPs), antigen presenting cells (certain papilloma VLPs), and glial cells (JC virus VLPs), among others (Seow & Wood, 2009).

While VLPs were initially developed for vaccination purposes, they have recently emerged as biocompatible self-assembling nanomaterials in form of bioimaging scaffolds, nanowires and nanocomposites (Manchester & Singh, 2006). As nanocages, VLPs are superior to conventional viral vectors as their production is fully scalable and result in large yields from cost-effective production processes (Pattenden, Middelberg, Niebert, & Lipin, 2005), even in microbial cells (Lunsdorf, Gurramkonda, Adnan, Khanna, & Rinas, 2011; Rodríguez-Limas, Tyo, Nielsen, Ramirez, & Palomares, 2011). Moreover, VLPs resist purification processes better than viral counterparts, making easier the further functionalization and modification of their surfaces. In this context, VLPs exhibit a tunable architecture that enables their loading with diverse cargoes and the improvement of cell-specific targeting.
VLPs of different viruses including cowpea chlorotic mottle virus (CCMV), the brome mosaic virus (BMV), polyomavirus (such as the Simian virus (SV40) and the human polyoma JC virus) and papillomavirus (Lee & Wang, 2006; Ma, Nolte, & Cornelissen, 2012) have been used, through in vitro assembly and disassembly hierarchical processes as well as through osmotic shock, to package and eventually deliver exogenous DNAs (Fang et al., 2012), oligonucleotides (Mateu, 2011), peptides and proteins (Henke, Rohmann, Bertling, Dingermann, & Zimmer, 2000; Kaczmarczyk, Sitaraman, Young, Hughes, & Chatterjee, 2011), small interfering RNAs and plasmids expressing short hairpin RNAs (Chou et al., 2010), small chemicals (Zhao et al., 2011), quantum dots (Li et al., 2009) and magnetic nanoparticles (Goicochea, De, Rotello, Mukhopadhyay, & Dragnea, 2007).

Further, VLP inner surfaces can be genetically engineered and/or chemically modified to incorporate directed functionalities allowing the attachment of different small molecules, smaller nanoparticles, drugs or genes. The most common approach of chemical bioconjugation is based on the identification of endogenously exposed amino acid residues (such as lysines, cysteines, tyrosines, aspartic acid or glutamic acid) with highly reactive sites (amino, carboxylic acid, thiol and phenol groups) that permit targeted molecular anchoring (Qu et al., 2004) without compromising the structure and functionality of VLPs (Ma, Nolte, & Cornelissen, 2012; Patel & Swartz, 2011; Lee, Nguyen, & Wang, 2011). For example, Pan and coauthors loaded human pre-miR146a RNA into bacteriophage MS2 VLPs by chemical conjugation with the trans-activator of transcription (Tat) 47–57 peptide. This strategy offers a novel miRNA delivery system to promote delivery and subcellular localization of miR-146a with the effective suppression of the targeting gene (Pan et al., 2012a). Recently, it has been described an efficient, easy and convenient production of multifunctional VLPs composed by the major structural protein VP6 of RV (rotavirus). VP6, refolded from Inclusion Bodies (IBs) was covalently loaded with the anticancer drug DOX and subsequent self-assembled into VLPs (Zhao et al., 2011). Similarly, it has been reported the delivery of active Gag-Cre recombinase, Gag-Fcy::Fur, and Gag-human caspase-8 proteins using avian retrovirus VLPs (Kaczmarczyk, Sitaraman, Young, Hughes, & Chatterjee, 2011).

On the one hand, outer-surfaces confer the natural tropism of the virus for the cell targeting of VLPs and their cargoes. Also, outer VLPs surfaces can be chemically modified via bioconjugation through the attachment of appropriate cell-surface receptor ligands (such as folic acid or lactobionic acid), or alternatively, genetically engineered through the insertion into solvent-exposed loops of specific binding domains.
(Stubenrauch, Gleiter, Brinkmann, Rudolph, & Lilie, 2001; Gleiter & Lilie, 2001) such as short peptides, antibodies, transferrin or cell penetrating peptides among others. These strategies for tropism modification, successfully applied in CPMV, CCMV, TMV and the bacteriophages Qβ, MS2, and M13 have been reviewed elsewhere (Lee & Wang, 2006; Strable & Finn, 2009; Mateu, 2011; Ma, Nolte, & Cornelissen, 2012).

In addition, it has been demonstrated (Zhao et al., 2011) that Doxorubicin-loaded VLPs decorated with lactobionic acid (LA chemically bound to the reactive amino groups of the VP6 capsid proteins) gained specific targeting to the hepatoma cell line HepG2. More recently, it has been developed a novel microRNA delivery system based on bacteriophage MS2 virus-like particles, chemically conjugated with the HIV-1 Tat47–57 peptide (Pan et al., 2012a). Moreover, VLPs chimeras have been generated by conventional genetic engineering resulting in the incorporation of cell-specific short peptides to phage M13 for tumor targeting (Chen et al., 2004), and of a liver cell-binding ligand (preS1) to hepatitis B core antigen (HBcAg) VLPs (Lee, Tey, Ho, & Tan, 2012). In addition, VLPs can be modified with polymers in order to improve their half-life, reduce immunogenicity and to stabilize cargo nucleic acids (Ma, Nolte, & Cornelissen, 2012). In this context, a novel polyethyleneimine-coated adeno-associated VLP formulation (PEI-AAV2-VLP) has been generated that show high siRNA transfer efficiency in MCF-7 breast cancer cells (Shao et al., 2012).

**Bacterial organelles**

Bacteria are microorganisms with a complex subcellular architecture (Bobik, 2006; Cheng, Liu, Crowley, Yeates, & Bobik, 2008; Gitai, 2005; Rudner & Losick, 2010). Many bacterial species contain self-organizing nano- and micro-compartments (bacterial microcompartments, or BMCs) consisting of virus-like protein-only shells with 60 to 10,000–20,000 self-assembled copies of one or a few protein species. These organelles package enzymes involved in specific metabolic pathways, confining such reactions and their putative unstable or toxic intermediate metabolites. Carboxysomes, the firstly identified bacterial organelles, were described in 1973 (Shively, Decker, & Greenawalt, 1970; Shively, Ball, Brown, & Saunders, 1973) as polyhedral inclusions of 100–150 nm in cross section, with a 3–4 nm protein shell composed of six to ten different proteins. Carboxysomes contains the RuBisCO enzyme (Shively, Ball, Brown, & Saunders, 1973) and the carbonic anhydrase enzyme, responsible of the conversion of HCO$_3^-$ to CO$_2$, the substrate for RuBisCO (Price, Coleman, & Badger, 1992; So et al., 2004; Yu, Price, Song, & Badger, 1992). The function of carboxysomes is to enhance autotrophic CO$_2$ fixation at low CO$_2$ levels. This role is supported by the
findings that carboxysome formation is induced by CO$_2$ limitation (McKay, Gibbs, & Espie, 1993) and that mutant strains unable to properly form carboxysomes require high CO$_2$ levels for autotrophic growth (English, Jin, & Shively, 1995; Price & Badger, 1989).

Other BMC proteins were later found encoded by the propanediol utilization operon (pdu operon) of Salmonella (Chen, Andersson, & Roth, 1994) and by an operon for metabolizing ethanolamine (eut operon) in enteric bacterial species, including Salmonella and Escherichia (Kofoid, Rappleye, Stojilkovic, & Roth, 1999). Salmonella enterica forms a polyhedral organelle when growing on 1,2-propanediol (1,2-PD) as a unique carbon and energy source (Bobik, Havemann, Busch, Williams, & Aldrich, 1999). S. enterica forms structures similar in size and shape to carboxysomes during growth on 1,2-PD but not during growth on other carbon sources (Bobik, Havemann, Busch, Williams, & Aldrich, 1999; Havemann, Sampson, & Bobik, 2002). More recently, Sutter and colleagues (Sutter et al., 2008) described the smallest (20–24 nm in diameter) known protein-based organelle in the hyperthermophilic bacterium Thermotoga maritima. This protein family was initially named “linocins” (Valdes-Stauber & Scherer, 1994), but it was renamed by the authors as “encapsulins”.

Carboxysomes are approximately icosahedral in shape, as revealed by electron microscopy studies. As in isometric viral capsids, the construction of such structures typically requires a combination of hexameric and pentameric units. The typical BMC domain consists of 90 amino acids in length with an alpha/beta fold pattern (Kerfeld et al., 2005; Yeates, Crowley, & Tanaka, 2010). BMC proteins self-assemble to form disc-shaped hexamers, the basic building blocks of the shell (Figure 1, green proteins). Such hexamers further assemble side-by-side, forming a flat molecular layer (Kerfeld et al., 2005; Tanaka et al., 2008; Tanaka, Sawaya, & Yeates, 2010). On the other hand, pentamers occupy the vertices of the icosahedral shell, generating curvature in an otherwise flat hexagonal sheet. Supporting this, homologous proteins CcmL and CsoS4A from two different types of pentamers (Figure 1, blue proteins), compatible with their placement at the vertices of 12 pentamer-icosahedral shells (Tanaka et al., 2008). Other BMCs do not follow this architectonic pattern. For example, the homologous protein of CcmL and CsoS4 in the Eut microcompartment is EutN, whose oligomeric status is hexameric, what is reflected in the structural differences (more irregular in shape) found between Eut and carboxysome microcompartments. In general, Eut (and also Pdu) microcompartments do not resemble regular icosahedrons as closely as the carboxysome does.
Mechanisms directing enzyme encapsulation within protein-based bacterial organelles have been recently elucidated. In some cases, a stretch of a few (~15-20) amino acids at the N-terminus of the inner cargo protein directs and binds it to specific sites on the inner surface of the shell protein. Then, in β carboxysomes, the protein CcmM is used as a scaffold to form interactions between shell proteins and enzymes (Cot, So, & Espie, 2008; Long, Badger, Whitney, & Price, 2007), through a CcmM C-terminal region with homology to the small subunit of RuBisCO (Price, Howitt, Harrison, & Badger, 1993). In Pdu microcompartments, some of the cargo enzymes are also encapsulated by N-terminal targeting sequences (Fan et al., 2010; Parsons et al., 2010). Interestingly, carboxysomes can self-assemble in vivo when RuBisCO has been deleted (Menon, Dou, Heinhorst, Shively, & Cannon, 2008), offering appealing opportunities to fill such nanocages with therapeutic molecules. A second example of this encapsulation strategy was provided by Sutter and colleagues (Sutter et al., 2008), regarding packaging of different proteins within encapsulins. In many bacteria, the encapsulin gene is being preceded, within an apparent two-gene operon, by the gene encoding either an iron-dependent peroxidase (DyP) or a protein closely related to the iron transporter ferritin (Flp). Sequence alignment of DyP and Flp genes revealed that only those followed by the encapsulin gene carry a C-terminal extension with a conserved amino acid sequence, responsible for the protein’s physical interaction with the encapsulin protein, through the binding to the inner shell surface (Sutter et al., 2008).

In a second route, there is not a directing peptide present in the cargo protein, but it is synthesized together with the shell-forming domain from one unique gene. In the hyperthermophilic archaeon Pyrococcus furiosus, a Flp coding sequence (without any targeting sequence directing its encapsulation by physical interaction with BMC proteins) is found fused in frame with an encapsulin gene (Sutter et al., 2008). Thus, both cargo and encapsulin proteins are synthesized as a fusion that further self-assembles to form a nano-cage containing the cargo protein in its inner space.

Specific targeting sequences could be of use in nanomedical applications to package proteins inside the stable self-assembled icosahedral shell of encapsulin. As an example, an icosahedral enzyme complex, lumazine synthase, was engineered to encapsulate target molecules by means of charge complementarity. The lumazine synthase from Aquifex aeolicus (AaLS) forms icosahedral capsids large enough to encapsulate proteins. The electrostatic charge within the capsid was modified by
replacing in each monomer four residues projecting into the lumen to glutamates, thus providing extra negative charges to the inner surface of the cage. On the other hand, the addition of a short stretch of positively charged amino acids to the cargo protein promoted its specific encapsulation by means of charge complementarity, as was demonstrated with a modified GFP containing ten arginine residues (Seebeck, Woycechowsky, Zhuang, Rabe, & Hilvert, 2006). Another packaging strategy would be the use of peptides naturally directing enzyme packaging into BMCs. In this regard, a short N-terminal sequence of propionaldehyde dehydrogenase (PduP) is required for its efficient packaging into the Pdu BMC. Fusion of this short (18 amino acids) N-terminal stretch from PduP to GFP, GST, or maltose-binding (Fan et al., 2010), and to green fluorescent protein or glutathione S-transferase proteins (Fan & Bobik, 2011) resulted in their successful encapsulation within Pdu BMC. As many peptides targeting to the vasculature of a variety of tissues, organs, and tumors have been identified (Pasqualini & Ruoslahti, 1996b; Pasqualini & Ruoslahti, 1996a), they can be used to functionalize BMC, as demonstrated for a small Hsp cage structure modified with the peptide RGD-4C or conjugation with an anti-CD4 monoclonal antibody (Flenniken et al., 2006). Release of the cargo in target cells can be enhanced by using polyhistidine tags, a powerful membrane-disrupting agent (Ferrer-Miralles et al., 2011), to promote pH-dependent disassembly in the endosomes (Dalmau, Lim, & Wang, 2009a; Dalmau, Lim, & Wang, 2009b). As far as we know, no nucleic acids have been encapsulated into BMCs to date.

Finally, flagella are also self-assembling protein-based bacterial nano-structures, organizing as nanotubes. Flagella are protein filaments involved in motility, up to 10-15 µm in length with a typical diameter of 12-25 nm, that consist of a helical assembly of the flagellin protein, with 11 protein subunits per helix turn (Macnab, 2003; Yonekura, Maki-Yonekura, & Namba, 2003). Engineered flagella can function as bionanotube scaffolds, displaying functional groups on their surfaces. So far, several peptides and proteins have been successfully displayed into flagella for use as an analytical tool (Tripp, Lu, Bourque, Sookdeo, & McCoy, 2001; Westerlund-Wikstrom, 2000). However, the investigation exploration of flagella as potential vehicles to deliver nucleic acids or drugs – artificial viruses or generic delivery agents – has been so far still very limited/narrow.

Artificial, self-organizing protein constructs
Inclusion bodies
Bacterial IBs are self-organizing protein-based amyloid granules, ranging between ~50 and 1000 nm, often occurring in the cytoplasm of the bacterium *Escherichia coli* when producing recombinant proteins. IBs are formed by a stereospecific protein-protein contacts (Speed, Wang, & King, 1996) based on cross-beta sheet interactions (Gonzalez-Montalban, Garcia-Fruitos, Ventura, Aris, & Villaverde, 2006), leading to a selective protein deposition with a high degree of purity (Morell *et al.*, 2008). Interestingly, although IBs have been traditionally considered by-products of recombinant protein production processes, these nanoparticles show potential in different biomedical applications (Garcia-Fruitos *et al.*, 2012) apart from as biocatalysts (Garcia-Fruitos & Villaverde, 2010) with tuneable biological activity (Garcia-Fruitos *et al.*, 2007). Additionally, these protein-based nanoparticles are mechanically stable (Garcia-Fruitos *et al.*, 2009), showing regulatable size (Garcia-Fruitos *et al.*, 2009), geometry (Garcia-Fruitos, Seras-Franzoso, Vazquez, & Villaverde, 2010), density (Peternel, Jevsevar, Bele, Gaberc-Porekar, & Menart, 2008), Z-potential, stiffness and wettability (Diez-Gil *et al.*, 2010). Importantly, it is feasible to obtain IBs formed by almost any protein of interest, eventually assisted by pull-down peptides that act as fuzzy architectonic tags (Nahalka & Nidetzky, 2007; Zhou, Xing, Wu, Zhang, & Lin, 2012; Wu, Xing, Zhou, & Lin, 2011; Garcia-Fruitos *et al.*, 2005). As IBs show high cell membrane penetrability (Figure 2) in absence of cytotoxicity and side effects at organic level, and the IB protein is made available to the cell upon uptake, IBs fabricated with therapeutic proteins act as unexpected vehicles for protein-based cell therapy (Villaverde *et al.*, 2012; Vazquez *et al.*, 2012; Liovic *et al.*, 2012; Villaverde, 2012), in form of Nanopills (administered in suspension) or Bioscaffolds (used as substrates for mammalian cell growth) (Table 2). In both presentations, IBs act as biomimetics of secretory granules from the endocrine system, and like them, they can be considered non-toxic functional amyloids (Maji *et al.*, 2008; Mankar, Anoop, Sen, & Maji, 2011) for the slow release of their forming proteins from a porous, fibrous protein scaffold (Cano-Garrido *et al.*, 2012). IBs, in addition, are excellent models to study in vivo amyloid assembling and their inhibitors (Espargaro, Villar-Pique, Sabate, & Ventura, 2012; Villar-Pique, Espargaro, Sabate, de Groot, & Ventura, 2012; Villar-Pique & Ventura, 2012; Garcia-Fruitos, Sabate, de Groot, Villaverde, & Ventura, 2011).

**Peptides**

Many natural or in-silico designed peptides (amyloidgenic or amphiphilic) spontaneously self-assemble in higher order structures. Such self-organizing processes can be controlled by extrinsic parameters such as temperature, ionic...
strength, pH, and the nature of the solvent, as well as by intrinsic physicochemical properties that can be adjusted by the modification of the amino acid sequence or the conjugated chemical groups. Self-assembling of short peptides occurs by non-covalent, weak cross-molecular local contacts between monomers including the formation of disulphide bridges and water-mediated hydrogen bonds, induction of turns, and hydrophobic, electrostatic, π-π staking and van der Waals interactions (Nisbet & Williams, 2012; Liu & Zhao, 2011). This result in the formation of a diversity of hierarchical 0 dimensional (0D), 1D, 2D and 3D structures within the nanoscale including fibrils, tubes, spheres, ribbons, vesicles, micelles, monolayers, bilayers, hydrogels and tapes. Many amyloidogenic peptides render fibers that can organize as hydrogels and other valuable biomaterials with biomedical potential (Lakshmanan, Zhang, & Hauser, 2012; Sadatmousavi, Soltani, Nazarian, Jafari, & Chen, 2011; Liu & Zhao, 2011; Liu et al., 2011; Kyle, Aggeli, Ingham, & McPherson, 2010; Gelain, Unsworth, & Zhang, 2010; Yang et al., 2009). Unfortunately, no regular cages have been so far constructed with short self-assembling peptides. Also, when fused to large proteins and produced in biological systems, these tags promote the aggregation of the fusion protein into IBs (Wu, Xing, Zhou, & Lin, 2011), proving that they can even drive the formation of IBs, but these agents are not valuable as fine architectonic tags in larger modular protein constructs, at least under a straightforward fusion approach.

However, cationic peptides combined with nucleic acids might form micro- (Plank, Tang, Wolfe, & Szoka, Jr., 1999) or nanosized particles (Domingo-Espin et al., 2011a) useful in gene therapy probably due to the condensation of plasmid DNA molecules. In fact, the DNA itself often stabilizes the protein incorporated into the complex through charge neutralization (Domingo-Espin et al., 2011a). Size and properties of these polyplexes are unpredictable and so far excluded from rational engineering. Despite this fact, bi-functional (Harbottle et al., 1998) or multifunctional (Domingo-Espin et al., 2011a) peptides have been extensively explored as valuable nucleic acid-condensing and delivery agents in non-viral gene therapy (Dietrich, Durr, & Koch, 2012; Said, Saleh, Abes, Gait, & Lebleu, 2010; Saccardo, Villaverde, & Gonzalez-Montalban, 2009).

**Engineered polypeptides**

Proteins are natural building blocks of viruses and intracellular microbial organelles as discussed above. Genetic engineering technologies make possible the rational design of new polypeptides with desired structure or activities, and the screening of randomly generated large peptide/protein libraries for specific functions. In non-viral gene
therapy, multifunctional modular proteins have been constructed as a strategy to recruit, in a single polypeptide chain, all functions required for in vitro condensation and further cell-targeted delivery of nucleic acids (Aris & Villaverde, 2004; Vazquez et al., 2009; Vazquez, Ferrer-Miralles, & Villaverde, 2008; Ferrer-Miralles, Vazquez, & Villaverde, 2008). This enables the biological production of such polypeptides in cell factories through cost-effective production and purification processes (Rodríguez-Carmona & Villaverde, 2010; Vazquez & Villaverde, 2010). Such a versatile approach has resulted into the generation of diverse categories of protein/DNA complexes as non-viral vehicles, able to deliver therapeutic nucleic acids both in cell culture and in vivo. As a representative example, engineered β-galactosidases displaying functional peptides in several solvent-exposed loops, promote the functional recovery of brain-damaged rats in ischemia models (Peluffo et al., 2003; Peluffo et al., 2006; Peluffo et al., 2011). The most common functional modules incorporated in these kind of constructs (cell receptor binding, DNA condensation, nuclear transport and endosomal escape) do not show any particular architectonic potential. Therefore, the size and geometry of the resulting polyplexes is unpredictable and it results from the tendency of the engineered protein to form supramolecular complexes, combined with the electrostatic interactions generated between protein and DNA. For instance, large molecular weight β-galactosidases (about 120 kDa per monomer, tetrameric in their original form) organize as amorphous, largely polydisperse entities (small soluble aggregates in the nanoscale) when empowered with cationic peptides displayed at the protein’s surface (Aris & Villaverde, 2000; Aris & Villaverde, 2003). The further addition of plasmid DNA has then a poor influence in this architecture, and the resulting polyplexes remain essentially unchanged. In contrast, shorter multifunctional peptides of about 20 kDa, based on the sequential joining of active amino acid segments, tend to precipitate as large size aggregates. The further addition of plasmid DNA stabilizes the protein building blocks, what results in the generation of polyplexes as monodisperse and fully soluble nanoparticles of 80 nm, extremely efficient in the delivery of the cargo DNA (Domingo-Espin et al., 2011b).

While true architectonic tags to be incorporated as an extra module of multifunctional proteins remain to be identified, a recent engineering approach shows promise to create and pre-define the self-assembling properties of building blocks in the construction of artificial viruses. The combination of an amino-terminal cationic peptide and a carboxy-terminal polyhistidine, when flanking a core protein (can be as diverse as GFP or p53), promote the generation of dipolar building blocks that tend to self-organize as rather planar nanoparticles of defined size (Unzueta et al., 2012b). Such
self-assembling is empowered at slightly acidic pHs, when the imidazole group of the histidines result protonated (Ferrer-Miralles et al., 2011). This in turn confers them cationic charge while it also activates the potential of this residue in acting as proton sponge for endosomal escape. Interestingly, the size of the resulting nanoparticles is directly proportional to the number of cationic residues in the amino-terminal tag (Unzueta et al., 2012b). Differently lengthened polyarginines as well as naturally occurring cationic protein segments (such as T22, R9 or A5), are valuable as architectonic regulators at the amino terminus (Unzueta et al., 2012b). Once formed, the resulting nanoparticles are highly stable at physiological pH, show a correct biodistribution once injected in model animals and efficiently reach the target cell compartment (Unzueta et al., 2012a). Loaded with DNA, R9-empowered members of this family of artificial viruses promote nuclear delivery of expressible DNA (Vazquez et al., 2010a; Vazquez et al., 2010b). Although this nanoscale architectonic principle is just in its infancy, it opens so far unexpected ways to engineer shape and function in recombinant protein-based artificial viruses by the incorporation of architectonic tags to building blocks.

**Conclusions and future prospects**

Different categories of self-assembling protein-based building blocks show promise as components of vehicles for emerging nanomedicines. In contrast to other materials explored for drug encapsulation and delivery, proteins show enormous functional flexibility that can be pre-designed, engineered, combined and adapted to specific situations. Functions such as cell surface receptor binding, nucleic acid condensation, membrane crossing and nuclear delivery have been identified in natural peptides or protein domains, and can be incorporated to fully de novo designed constructs in an attempt to mimic viral infectious activities. However, protein self-assembling has been largely reluctant to engineering. Therefore, natural constructs, initially viruses and VLPs and lately BMCs, have been adapted to deliver nucleic acids, therapeutic proteins and chemicals (Table 3). So far, the extent of manipulation of these natural nanocages is limited and restricted by the architectonic constraints imposed by fine protein-protein contacts derived from long-term evolution. The recent emerging of bioinspired, protein-based constructs with regulatable self-organizing properties, namely multifunctional modular proteins and IBs, opens a door for a more versatile, functional and architectonic control of drug vehicles (Table 3). Since the protein sequence and functions, including those governing self-assembling, can be easily manipulated by conventional genetic engineering, the progressive comprehension of the mechanics of protein-protein contacts and supramolecular organization in recombinant nanocages...
and the development of architectonic tags should permit the so far neglected de novo design and biofabrication of bioinspired, fully biocompatible, multifunctional constructs, with utility in nanomedicines and personalized medicine for the targeted delivery of chemicals, therapeutic proteins and innovative drugs.

**Acknowledgments:** We are indebted to MINECO (BFU2010-17450, ACI2009-0919 and IT2009-0021), AGAUR (2009SGR-0108), FISS (PS09-00165 and PI12-00327) and CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN, Spain) for funding our research on protein-based nanotherapeutics. CIBER-BBN is an initiative funded by the VI National R&D&i Plan 2008-2011, Iniciativa Ingenio 2010, Consolider Program, CIBER Actions and financed by the Instituto de Salud Carlos III with assistance from the European Regional Development Fund. EGF is supported by the Programa Personal de Técnico de Apoyo (Modalidad Infraestructuras científico-tecnológicas, MICINN). AV received an ICREA ACADEMIA award.

**Declaration of interest:** The authors report no declarations of interest.
Figure legends:

**Figure 1. Hierarchical assembly of natural, non-viral protein nano-cages.**

A) Schematic model showing the hierarchical assembly of a typical bacterial microcompartment. Facets of these icosahedral nano-cages (BMCs) are usually made of one type of protein (in green) that self-assembles to render hexamers. The shape of such hexamers is tailored for further side-by-side assembly into a flat molecular layer. On the other hand, vertices are formed by pentamers resulting of the self-assembly of a different protein (in blue). BMC sizes can range from 20-25 nm for the encapsulin shell to 100–150 nm for carboxysomes. B) Eukaryotic vaults (a hollow, barrel-like structure) are formed by the assembly of two identical cup-like halves joined at their open ends (right figure). Each half vault is in turn composed of a single eight-petaled, flower-like structure, which is folded into a "cup-shaped” half vault. Each petal is composed of six copies of the MVP protein (left figure). Adapted from Kedersha et al. (1991), The Journal of Cell Biology, Volume 112 (2): 225–235.

**Figure 2. Intracellular protein delivery mediated by bacterial IBs.** IBs formed by a variant of the green fluorescent protein (GFP) crossing cytosolic (red) and nuclear (blue) membranes of cultured HeLa cells, observed in a 3D confocal reconstruction. IBs purified from produced bacteria as described (Rodriguez-Carmona, Cano-Garrido, Seras-Franzoso, Villaverde, & Garcia-Fruitos, 2010) were suspended in cell culture media and exposed to HeLa cells for several hours. Precise methodological details can be found elsewhere (Vazquez et al., 2012). The image is taken from reference (Villaverde et al., 2012).


Reference List


37


of the Immunoreceptor CD300f Has a Neuroprotective Role in a Model of Acute Brain Injury. *Brain Pathol.*


(epsilon class) is a component of the carboxysome shell. *J.Bacteriol.*, 186:623-630.


nuclear migration of protein-only, arginine-rich nanoparticles. *Biomaterials,* **31** 9333-9339.


