nano-shells as artificial viruses for gene delivery" in Nanomedicine (Elsevier), vol. 10 (2014), p.535-541. The final version is available at DOI 10.1016/j.nano.2013.11.006

Original article

Nanomedicine: Nanotechnology, Biology and Medicine

Sheltering DNA in self-organizing, protein-only nano-shells as artificial viruses for gene delivery.

Ugutz Unzueta [1, 2, 3 §], Paolo Saccardo [1, 2, 3 §], Joan Domingo-Espín [1, 2, 3], Juan Cedano [4], Oscar Conchillo-Solé [1], Elena García-Fruitós [3, 1, 2], María Virtudes Céspedes [3, 5], José Luis Corchero [3, 1, 2], Xavier Daura [1, 6], Ramón Mangues [5, 3], Neus Ferrer-Miralles [1, 2, 3], Antonio Villaverde [1, 2, 3 *], and Esther Vázquez [1, 2, 3 *]

[1] Institut 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 de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, 08193 Barcelona, Spain
[4] Laboratory of Immunology, Regional Norte, Universidad de la República, Gral. Rivera 1350; Salto, 50.000, Uruguay
[5] Grup d'Oncogènesi i Antitumorals, Institut de Recerca, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain
[6] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

[§] Equally contributed

***** Corresponding authors: A. Villaverde; antoni.villaverde@uab.cat
E. Vazquez; esther.vazquez@uab.cat

Keywords: Nanoparticles; protein building blocks; self-assembling; artificial viruses; gene therapy

The authors declare no competing interests.

Text word count: 3412; Abstract word count: 133; Number of references: 43; Number of figures/tables: 4

The authors acknowledge the financial support granted to E.V. from FIS (PI12/00327) and to A.V. from Agència de Gestió d'Ajuts Universitaris i de Recerca (2009SGR-108), and from the Centro de Investigación Biomédica en Red (CIBER) de Bioingeniería, Biomateriales y Nanomedicina (NANOPROVIR project), financed by the Instituto de Salud Carlos III with assistance from the European Regional Development Fund. U.U. and P.S. received PhD fellowships from ISCIII and J.D.E. from MICINN. A.V. has been distinguished with an ICREA ACADEMIA Award.

Abstract

By recruiting functional domains supporting DNA condensation, cell binding, internalization, endosomal escape and nuclear transport, modular single-chain polypeptides can be tailored to associate with cargo DNA for cell-targeted gene therapy. Recently, an emerging architectonic principle at the nanoscale has permitted tagging protein monomers for self-organization as protein-only nanoparticles. We have studied here the accommodation of plasmid DNA into protein nanoparticles assembled with the synergistic assistance of end terminal poly-arginines (R9) and poly-histidines (H6). Data indicate a virus-like organization of the complexes, in which a DNA core is surrounded by a solvent-exposed protein layer. This finding validates end-terminal cationic peptides as pleiotropic tags in protein building blocks for the mimicry of viral architecture in artificial viruses, representing a promising alternative to the conventional use of viruses and virus-like particles for nanomedicine and gene therapy.

Background

Non-viral gene therapy and in general emerging nanomedicines aim to mimic viral activities in tuneable nanoparticles, for the cell-targeted delivery of cargo nucleic acids and other drugs [1;2]. Among a diversity of tested materials (including lipids, natural polymers, quantum dots, carbon nanotubes and dendrimers), proteins offer full biocompatibility, biodegradability, and a wide spectrum of functionalities that can be further adjusted by genetic engineering. Such a functional versatility is in contrast with the null control so far exercised over the supramolecular organization of *de novo* designed building blocks for protein-based complexes [3]. While protein nanoparticles based on natural cages, mainly infectious viruses [4], virus-like particles (VLPs) [5], eukaryotic vaults [6] and bacterial microcompartments (BMCs) [7] take advantage of the evolutionarily optimized self-assembling activities of their building blocks, fully the novo multifunctional protein monomers fail to reach predefined nanoscale organization. Only a very limited number of approaches, based on the engineering of oligomerization domains present in nature have resulted in the successful construction of efficient building blocks for protein shell generation [8]. Complexes of DNA and cationic proteins often result in polydisperse soluble aggregates probably derived from intrinsically disordered protein-protein interactions [9;10], or in which the DNA itself plays a leading architectonic role, stabilizing aggregation-prone protein monomers in form of monodisperse nanoparticles [11]. Self-assembling peptides, that organize as different types of nanostructured materials [12], promote unspecific aggregation when fused to larger proteins [13;14], making them useless as fine architectonic tags. In summary, the rational de novo design of protein monomers with self-assembling activities has remained so far unreachable. Very recently [15], we have described that pairs of ‘architectonic’ peptides consisting of an N-terminal cationic stretch plus a C-terminal polyhistidine, when combined in structurally diverse scaffold proteins (GFP, p53 and others), generate strongly dipolar charged monomers that spontaneously self-assemble. The resulting protein oligomers, ranging from 10 to 50 nm, show fast nuclear

migration (compatible with cytoskeleton-linked active transport) and penetrability [16], high stability and proper biodistribution upon systemic administration [17]. Important levels of gene expression where also achieved when the protein was associated to plasmid DNA [18]. Yet these protein particles efficiently bind plasmid DNA for transgene expression and are very promising tools in nanomedicine [18], their supramolecular organization remains so far unexplored. The purpose of this study is to investigate the architectonic properties of the polyplexes formed by expressible DNA and the paradigm protein R9-GFP-H6, to better understand the basis of the high cell penetrability and at which extent the resulting complexes adopt virus-like organization. A solid comprehension of how multifunctional proteins interact with exogenous DNA should enable the design and efficient biofabrication of true artificial viruses.

Methods

Protein production and DNA binding

The modular organization of R9-GFP-H6 [18], T22-GFP-H6 [17] and HNRK [11] has been described elsewhere. GFP-H6 is a parental version of R9-GFP-H6 and T22-GFP-H6 that does not self-assemble under physiological conditions [15;18]. Apart from their architectonic capability, R9 (RRRRRRRRR) acts as a cell penetrating peptide and nuclear localization signal [18] and T22 (RRWCYRKCYKGYCYRKCR) as a powerful ligand of the cell surface receptor CXCR4 [17]. Both stretches, being cationic, are potentially able to bind DNA. H6 (HHHHHH) is at the same time a useful tag for one-step chromatographic protein purification and a potent endosomolytic agent [19]. Precise amino acid sequences at the links between GFP and the fused peptides can be found elsewhere [17]. The protein constructs indicated above were produced in bacteria following conventional procedures and purified in a single step by His-based affinity chromatography [15], through activities assisted by the Protein Production Platform (CIBER-BBN) (http://www.bbn.ciber-bbn.es/programas/plataformas/equipamiento). Protein-DNA complexes were

generated by incubation at appropriate ratios in HBS buffer (pH 5.8) for 60 min at room temperature.

Cell culture, confocal microscopy and transmission electron microscopy (TEM)

HeLa (ATCC-CCL-2) cell line was cultured as previously described [16] and always monitored in absence of fixation to prevent internalization artefacts. Nuclei were labelled with 200 ng/ml Hoechst 33342 (Molecular Probes, Eugene, Oregon, USA) and plasma membranes with 2.5 µg/ml CellMask^{TM} Deep Red (Molecular Probes, Invitrogen, Carlsbad, CA, USA) for 5 min. Cells exposed to nanoparticles were recorded with a TCS-SP5 confocal laser scanning microscope (Leica Microsystems, Heidelberg, Germany) with a Plan Apo 63x / 1.4 (oil HC x PL APO lambda blue) objective. Three-dimensional cell models were generated with the Imaris v. 6.1.0 software (Bitplane; Zürich, Switzerland). For TEM, protein/DNA complexes were contrasted by evaporation of 1 nm platinum layer in carbon-coated grids and then visualized in a Hitachi H-7000 transmission electron microscope.

DNA protection assay

In the buffers optimal for their respective stability [11;15], R9-GFP-H6 and GFP-H6 (HBS pH 5.8), T22-GFP-H6 (carbonate buffer, pH 5.8) and HNRK (HBS + dextrosa pH 5.8) were mixed with 1 µg of plasmid DNA (pTurboFP635, [18]) at 1 and 2 retardation units. Mixtures were incubated at room temperature for 1 h and then threated with 0.5 µg/ml DNAse I (Roche) at 37º C, in presence of 2.5 mM $MgCl_2$ and 0.5 mM $CaCl_2$. Samples were collected just before DNAse I addition and at 5, 20 and 60 min of the digestion reaction. DNAse I was inactivated by adding EDTA 2.3 µM final concentration and by heating the samples for 20 min at 70º C. The remaining DNA was released from protein complexes by adding 10 U of Heparin followed by 2 hours incubation at 25º C. Subsequently, samples were analyzed in 1% agarose gels. DNA

signals in agarose gel were interpreted and analyzed with Quatity One software (Bio-Rad). Experiments were performed by triplicate.

Determination of particle size and Z potential

Volume size distributions of self-assembled protein nanoparticles and protein-DNA complexes were determined by triplicate using a dynamic light scattering (DLS) analyser at the wavelength of 633 nm, combined with non-invasive backscatter technology (NIBS) (Zetasizer Nano ZS, Malvern Instruments Limited, Malvern, U.K.). Z Potencial of these materials was determined in the same device in HBS buffer (pH 5.8, 10 µg/mL final protein concentration). Measurements were carried out at 25 °C using a disposable plastic cuvette. Each sample was analysed by triplicate.

Molecular modelling

To build R9-GFP-H6-based particles, a model of the monomer was first generated using Modeller 9v2 [20] and the pdb structure "1qyo" as template. The arginine and histidine tails were modeled using the loopmodel function of this package. The structural models of the assembled monomers at pH 7 and pH 5.8 were then created using HADDOCK 2.0 [21], with the protonation states chosen according to pH and residue pKas, defining the 9 arginines at the N-terminus as active residues and the 6 histidines at the C-terminus as passive residues and enforcing C5 symmetry led to star-shaped conformations. Alternative conformations were obtained using the tail arginines as active residues and no passive ones. All these models where analysed with FoldX using the function "AnalyseComplex" [22]. Defaults were taken for any other simulation parameters. This protocol has been already used in a previous study [18]. DNA was modeled for a 26 bp random sequence with the 3DDART server [23] using default parameters. The structural model of the (1:1) DNA-protein complex was created with HADDOCK2.0 using N-terminal-tail arginines and C-terminal-tail histidines as active residues and all DNA bases as passive ones. Superposition of all

resulting solutions was performed with PROFIT [24] (an implementation of the McLachlan algorithm, [25]) , using only the DNA molecule as subject of the structural fit. The structural comparison of disks made of TMV coat protein and R9-GFP-H6 was performed with SwissPdbViewer* [26] to superimpose the 2om3 PDB structure and the modelled building block [27]. To facilitate the visualization of the resulting models, images were generated using Chimera [28] as rendering tool.

Results

Hexahistidine tails, when combined in single chain polypeptides with N-terminal cationic peptides, such as R9 or T22, promote assembling of these building blocks as regular particles at neutral or slightly acidic pH values [15], at which the imidazol group gets protonated and the tag moderately cationic [19]. When nanoparticles formed by R9-GFP-H6 at pH 7 and 8 (Figure 1a) were incubated with DNA, particle size remained close to 20 nm (Figure 1 b), the size previously observed in absence of DNA [15]. At pH 4 and 10, protein-DNA complexes peaked at 0.8 and 2 µm respectively (Figure 1 b), which is in agreement with the tendency of the protein alone to form amorphous aggregates under denaturing conditions Figure 1 a). Interestingly, at slightly acidic pH (5.8), where the transfection mediated by R9-GFP-H6 had resulted more efficient [15], the population of polyplexes split in two fractions, peaking at 38 and 700-800 nm respectively, with no symptoms of protein instability or aggregation (protein-only nanoparticles peaked between 20 and 30 nm). The ability of these protein constructs to bind DNA was generically confirmed by retardation mobility assays (Figure 1 c).

These polyplexes were examined by confocal microscopy during exposure to cultured cells, taking advantage of the natural green fluorescence of the protein partner and upon staining the DNA with the blue fluorescent dye Hoechst 33342. Small spherical particles (Figure 2 a) and larger rod-shaped versions, some slightly twisted or ramified (Figure 2 b) were observed, whose size fitted respectively to the two main peaks

194 determined by DLS (Figure 2 b). The blue DNA signal appeared coincident with the
195 green label, but its slightly smaller size suggested that DNA occurred in inner cavities
196 of protein entities. Qualitatively, rod-shaped nanoparticles seemed more efficient in
197 embedding DNA than the regular versions, as an important fraction of spheres, but not
198 rods, appeared to be empty (Figure 2 a, b). Fine confocal sections and 3D isosurface
199 reconstructions strongly suggested that a core DNA was shielded by a solvent-exposed
200 protein layer (Figure 2 c), in a virus-like architectonic scheme.
201
202 In this regard, rod-shaped forms shown in Figure 2 a and c strongly evoked the
203 morphologies of capsid proteins observed in plant viruses. Furthermore, a
204 superimposition of the RNA-containing, rod-shaped tobacco mosaic virus (TMV) disk (a
205 structural intermediate in the construction of helical capsids) and an energetically
206 stable, planar, star-shaped molecular model of the self-assembled R9-GFP-H6 at pH
207 5.8 are presented (Figure 2 d), showing coincidence in diameter and in monomer
208 organization. Interestingly, a similar spatial distribution of arginines around the central
209 cavities was found in both viral and non-viral complexes (Figure 2 d, inset). TEM
210 images of material deposited on the gird in absence of cells indicated again a
211 prevalence of tubular structures (Figure 2 e), with a diameter compatible with the
212 particles observed by confocal analyses (between 20 and 30 nm) and with R9-GFP-H6
213 disks obtained by molecular modelling (Figure 2 d). Importantly, no DNA was found
214 associated to internalized R9-GFP-H6 protein-only nanoparticles (Figure 2 f). This
215 indicates that cellular nucleic acids that the protein complexes might eventually find
216 during the intracellular trafficking would result not available for binding, and that the
217 only cargo suitable to form artificial viruses is the nucleic acid loaded *in vitro*.
218
219 Furthermore, DNA embedded in R9-GFP-H6 shells resulted highly protected from
220 DNAse I attack (Figure 3 a). This effect was similar to that promoted by the closely
221 related, self-assembling construct T22-GFP-H6. Contrarily, the short modular peptide

HNRK [18;29], that although being positively charged does not exhibit architectonic properties, failed in protecting DNA from digestion (Figure 3 a). In the HNRK-DNA polyplexes, from which DNA overhangs, the nucleic acid is the main architectonic regulator of the resulting particles (of around 80 nm), the protein fraction being clustered by DNA instead of entrapping it in shell-like structures [11]. The high protection of R9-GFP-H6-linked DNA also indicates that whether DNA molecules are externally associated to some protein particles as suggested by confocal analysis (Figure 2), the fraction of such material is statistically low.

Why at slightly acidic pH and in presence of DNA, R9-GFP-H6 ~20 nm-nanoparticles rearrange as alternative spherical or cylindrical shells remains to be solved, but it might be speculated that the dipolar nature of the modular protein would permit a reorganization of the building blocks, to orient the positive protein patches at the inner surface of the shell, in contact with DNA. For that, spheres and cylinders would permit appropriate protein-protein interactions. In agreement with this hypothesis, the superficial charge of protein-only particles was -16.2±1.8 mV, while in presence of plasmid DNA (2 RU) it shifted to a more negative value (-24.5±2.0 mV) (Figure 3 b). Interestingly, by applying the same amount of protein, the number of nanoparticles was reduced by more than 50 % in the presence of DNA, consistent with a higher protein demand to form nanoparticles up to 800 nm than to form protein-only nanoparticles of ~20 nm. On the other hand, the organization of protein shells as spheres or alternatively as rods would require a certain degree of flexibility in monomer-monomer contacts, allowing alternative arrangements of the oligomers. The in-equilibrium protonation and charge profile of the histidine tail population (pK~6) [19], would confer enough structural versatility of these interactions supportive of spherical and disk-based cylindrical organization. In agreement, alternative stable versions of R9-GFP-H6 oligomers (pentamers) resulted from the docking process, sustained by slightly divergent styles of inter-molecular interactions (Figure 4 a). Such pentamers, similarly

distributed oligomers (eg hexamers) orf their combination, could support both spherical and rod-shaped architectures as in the case of virus shells. After careful analysis of these models, we have identified, apart from electrostatic interactions (-7.33 Kcal/mol), van der Wals forces as the main components keeping the monomers together (-42.38 Kcal/mol), in some cases with hydrogen bonds (-29.13 Kcal/mol) contributing significantly to the stability of the oligomers (data taken from the model disk represented in Figure 1 d and in Figure 4 a, left).

Figure 4 b shows a potential mode of interaction between DNA and R9-GFP-H6, based on unspecific charge-charge interactions between DNA and the GFP-overhanging tails. This architecture would enable the organization of several GFP molecules around a single DNA helix in a form similar to those shown in Figure 2 d for RNA, as suggested by the superposition of the best 50 solutions of a (1:1) DNA-protein docking simulation, which shows a uniform distribution of GFP-based building blocks around the DNA.

Discussion

The severe biological risks and negative media perception associated to the administration of natural viruses [30] have dramatically compromised the development of viral gene therapy [31;32] and prompted researchers to explore manmade alternatives as vehicles for the delivery of therapeutic genes. The artificial virus concept [2] claims the use of nanoparticles, that upon convenient upstream design, biological fabrication and engineering can successfully mimic properties of the viral infectious cycle that are relevant to transgene delivery and expression [33]. Nanotechnologies and material sciences offer interesting approaches to generate functional nanostructured carriers, and a spectrum of materials are being explored in this regard [34], even under suspicion of potential toxicity [35]. Among them, proteins are the most versatile regarding structure and function, being fully biocompatible, suitable of biological fabrication and not posing safety of toxicity concerns. In fact, vaults and

BMCs, or the recombinant version of viruses, namely VLPs, can be conveniently adapted to embed cargo molecules for targeted delivery [36]. In a more versatile approach, modular proteins containing cationic stretches for nucleic acid binding and condensation, as well as other functional segments such as cell penetrating peptides, ligands or nuclear localization signals, have been under continuous design to recruit virus-like functions in single chain molecules [37-40]. However, despite the functional versatility of these constructs they fail to reach ordered nanoscale structures, in most cases being the DNA the main driving force of the polyplexe architecture [11]. In fact, the assembly of viral capsids results from a complex combination of intermolecular interactions including hydrophobic, electrostatic, van der Waals, and hydrogen bonds [41] that are excluded from a rational design in the novo designed recombinant proteins. Recently, we have determined that a combination of a cationic peptide plus a hexahistidine, placed at the amino and caboxy termini respectively of modular proteins grant them with the ability to self-organize as regular protein-only nanoparticles, able to penetrate target cells and to reach the nucleus in a very efficient way [15-17]. We have here shown how at a slightly acidic pH and in presence of DNA, the contacts promoted by the hexahistidine tail are able to accommodate structural rearrangements, among others those promoting a re-orientation of cationic segments in the inner surface, that convert plain oligomers into more complex supramolecular structures, namely closed protein shells, in a virus-like fashion (Figures 1, 2). Both conventional isometric and rod-shaped architectonic models occurring in natural viruses are spontaneously reached by the self-assembling of tagged GFP-H6, efficiently embedding the foreign DNA in the inner cavity of a protein-only shell (Figure 2). Such a dual construction scheme at the nanoscale reminds the organization of viral proteins. The rotavirus VP6 capsid protein, whose essential organization is a trimer, assembles into either nanotubes or nanospheres when produced as a recombinant version [42]. Cationic peptides R9 and H6 promotes the oligomerization of a monomeric GFP into particles whose size measured by DLS (Figure 1 a) is compatible with that of pentamers (or

eventually hexamers, Figure 4 a). The presence of exogenous DNA upon in vitro incubation stimulates the arrangement of these building blocks in higher order, larger complexes (Figure 1 b) with flexibility to form nanospheres and nanotubes (Figure 2). The organizing ability of DNA over cationic proteins to rend ordered protein-DNA complexes has been reported previously ([11] and references therein), and cationic interactions seem to be the driving force for the primary DNA-protein interaction (Figure 1 c), that result in nuclease attack protection (Figure 3). The ability of R9-GFP-H6 oligomers to bind and combine with nucleic acids is restricted to exogenous DNA, as not protein-DNA complexes were observed when mammalian cells were exposed to protein alone, which efficiently internalizes cultured cells ([16] and Figure 2 f). In addition, the carrier DNA promotes important levels of gene expression, the whole R9-GFP-H6-DNA complexes acting structurally and functionally like artificial viruses.

Importantly, the ability of the end-terminal tags of cationic nature to promote protein self-assembling seems to be irrespective of the polypeptide chosen as the core of the assembly, or at least not limited to a particular protein species [15]. This opens a door to select non-immunogenic homologous protein candidates as building blocks of nanoparticles in order to avoid any immune response upon systemic administration, what could be a critical bottleneck to the therapeutic use of artificial viruses based on *de novo* designed self-assembling proteins.

In summary, we have demonstrated for the first time how protein-based artificial viruses, namely functional nanoparticles formed by self-assembling protein shells shielding a core DNA, can be generated by the fully de novo design of building blocks. This fact not only validates R9 and H6 as pleiotropic peptides in vehicles for non-viral gene therapy, but it also reveals an unexpected architectonic potential of these tags in the generation of tuneable protein shells, whose properties can be further polished by conventional protein engineering. These versatile agents are promising alternatives to

natural protein constructs, including viruses, VLPs, vaults and BMCs, which because of several limitations including rigid architecture but also biosafety concerns, are less suitable for engineering and adaptation to nanomedical purposes.

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), and of Amable Bernabé from Soft Materials Service (ICMAB-CSIC/CIBER-BBN).

References

1. Wagner E. Strategies to improve DNA polyplexes for in vivo gene transfer: will "artificial viruses" be the answer? *Pharm.Res.* 2004; **21:** 8-14.

2. Mastrobattista E, van der Aa MA, Hennink WE, Crommelin DJ. Artificial viruses: a nanotechnological approach to gene delivery. *Nat.Rev.Drug Discov.* 2006; **5:** 115-21.

3. Tu RS, Tirrell M. Bottom-up design of biomimetic assemblies. *Adv.Drug Deliv.Rev.* 2004; **56:** 1537-63.

4. Giacca M, Zacchigna S. Virus-mediated gene delivery for human gene therapy. *J.Control Release* 2012; **161:** 377-88.

5. Ma Y, Nolte RJ, Cornelissen JJ. Virus-based nanocarriers for drug delivery. *Adv.Drug Deliv.Rev.* 2012; **64:** 811-25.

6. Han M, Kickhoefer VA, Nemerow GR, Rome LH. Targeted vault nanoparticles engineered with an endosomolytic peptide deliver biomolecules to the cytoplasm. *ACS Nano.* 2011; **5:** 6128-37.

7. Corchero JL, Cedano J. Self-assembling, protein-based intracellular bacterial organelles: emerging vehicles for encapsulating, targeting and delivering therapeutical cargoes. *Microb Cell Fact.* 2011; **10:** 92.

8. Doll TA, Raman S, Dey R, Burkhard P. Nanoscale assemblies and their biomedical applications. *J.R.Soc.Interface* 2013; **10:** 20120740.

9. Aris A, Villaverde A. Engineering nuclear localization signals in modular protein vehicles for gene therapy. *Biochem.Biophys.Res.Commun.* 2003; **304:** 625-31.

10. Aris A, Villaverde A. Molecular organization of protein-DNA complexes for cell-targeted DNA delivery. *Biochem.Biophys.Res.Commun.* 2000; **278:** 455-61.

11. Domingo-Espin J, Vazquez E, Ganz J *et al.* Nanoparticulate architecture of protein-based artificial viruses is supported by protein-DNA interactions. *Nanomedicine (Lond)* 2011; **6:** 1047-61.

12. Lakshmanan A, Zhang S, Hauser CA. Short self-assembling peptides as building blocks for modern nanodevices. *Trends Biotechnol* 2012; **30:** 155-65.

13. Zhou B, Xing L, Wu W, Zhang XE, Lin Z. Small surfactant-like peptides can drive soluble proteins into active aggregates. *Microb.Cell Fact.* 2012; **11:** 10.

14. Wu W, Xing L, Zhou B, Lin Z. Active protein aggregates induced by terminally attached self-assembling peptide ELK16 in Escherichia coli. *Microb.Cell Fact.* 2011; **10:** 9.

15. Unzueta U, Ferrer-Miralles N, Cedano J *et al.* Non-amyloidogenic peptide tags for the regulatable self-assembling of protein-only nanoparticles. *Biomaterials* 2012; **33:** 8714-22.

16. Vazquez E, Cubarsi R, Unzueta U *et al.* Internalization and kinetics of nuclear migration of protein-only, arginine-rich nanoparticles. *Biomaterials* 2010; **31:** 9333-9.

17. Unzueta U, Cespedes MV, Ferrer-Miralles N *et al.* Intracellular $CXCR4^{+}$ cell targeting with T22-empowered protein-only nanoparticles. *Int.J.Nanomedicine* 2012; **7:** 4533-44.

18. Vazquez E, Roldan M, Diez-Gil C *et al.* Protein nanodisk assembling and intracellular trafficking powered by an arginine-rich (R9) peptide. *Nanomedicine.(Lond)* 2010; **5:** 259-68.

19. Ferrer-Miralles N, Corchero JL, Kumar P *et al.* Biological activities of histidine-rich peptides; merging Biotechnology and Nanomedicine. *Microb Cell Fact.* 2011; **10:** 101.

20. N.Eswar, M.A.Marti-Renom, B.Webb, M.S.Madhusudhan, D.Eramian, M.Shen, U.Pieper, and A.Sali. Comparative Protein Structure Modeling With MODELLER. Current Protocols in Bioinformatics, John Wiley & Sons, Inc. [Supplement 15], 5.6.1-5.6.30, 200. 2009.

Ref Type: Generic

21. de Vries SJ, van Dijk AD, Krzeminski M *et al.* HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins* 2007; **69:** 726-33.

22. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J.Mol.Biol.* 2002; **320:** 369-87.

23. van DM, Bonvin AM. 3D-DART: a DNA structure modelling server. *Nucleic Acids Res.* 2009; **37:** W235-W239.

24. Martin ACR. ProFit. *http://www.bioinf.org.uk/software/profit/* 2009.

25. McLachlan AD. Rapid Comparison of Protein Structres. *Acta Crystallogr.* 1982; **A38:** 871-3.

415 26. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an
416 environment for comparative protein modeling. *Electrophoresis* 1997; **18:** 2714-
417 23.

418 27. Guex N, Diemand A, Peitsch MC. Protein modelling for all. *Trends Biochem.Sci.*
419 1999; **24:** 364-7.

420 28. Pettersen EF, Goddard TD, Huang CC *et al.* UCSF Chimera--a visualization
421 system for exploratory research and analysis. *J.Comput.Chem.* 2004; **25:** 1605-
422 12.

423 29. Domingo-Espin J, Petegnief V, de VN *et al.* RGD-based cell ligands for cell-
424 targeted drug delivery act as potent trophic factors. *Nanomedicine* 2012; **8:**
425 1263-6.

426 30. Edelstein ML, Abedi MR, Wixon J. Gene therapy clinical trials worldwide to
427 2007--an update. *J.Gene Med.* 2007; **9:** 833-42.

428 31. Abbott A. Questions linger about unexplained gene-therapy trial death.
429 *Nat.Med.* 2006; **12:** 597.

430 32. Williams DA, Baum C. Medicine. Gene therapy--new challenges ahead.
431 *Science* 2003; **302:** 400-1.

432 33. Aris A, Villaverde A. Modular protein engineering for non-viral gene therapy.
433 *Trends Biotechnol.* 2004; **22:** 371-7.

434 34. Villaverde A. *Nanoparticles in translational science and medicine*. London:
435 Academic Press (Elsevier), 2011.

35. Sanvicens N, Marco MP. Multifunctional nanoparticles--properties and prospects for their use in human medicine. *Trends Biotechnol.* 2008; **26:** 425-33.

36. Rodriguez-Carmona E, Villaverde A. Nanostructured bacterial materials for innovative medicines. *Trends Microbiol.* 2010; **18:** 423-30.

37. Vazquez E, Ferrer-Miralles N, Mangues R, Corchero JL, Schwartz S Jr, Villaverde A. Modular protein engineering in emerging cancer therapies. *Curr.Pharm.Des* 2009; **15:** 893-916.

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

39. Vazquez E, Ferrer-Miralles N, Villaverde A. Peptide-assisted traffic engineering for nonviral gene therapy. *Drug Discov.Today* 2008; **13:** 1067-74.

40. Ferrer-Miralles N, Vazquez E, Villaverde A. Membrane-active peptides for non-viral gene therapy: making the safest easier. *Trends Biotechnol.* 2008; **26:** 267-75.

41. Zlotnick A. Are weak protein-protein interactions the general rule in capsid assembly? *Virology* 2003; **315:** 269-74.

42. Plascencia-Villa G, Saniger JM, Ascencio JA, Palomares LA, Ramirez OT. Use of recombinant rotavirus VP6 nanotubes as a multifunctional template for the synthesis of nanobiomaterials functionalized with metals. *Biotechnol Bioeng.* 2009; **104:** 871-81.

43. Vazquez E, Roldan M, ez-Gil C *et al.* Protein nanodisk assembling and intracellular trafficking powered by an arginine-rich (R9) peptide. *Nanomedicine (Lond)* 2010; **5:** 259-68.

Figure 1. Molecular architecture of R9-GFP-H6-DNA polyplexes. A) Size distribution of R9-GFP-H6 in absence of DNA, at different pH values. Some of the data shown here have been published previously [15]. B) Size distribution of R9-GFP-H6-DNA polyplexes formed at different pH values. DNA alone is shown as a control. C) DNA mobility assay (using pTurbo FP635 [11] as reporter DNA) of R9-GFP-H6-DNA polyplexes formed at pH 5.8. GFP-H6 is shown as a control, non-binding protein.

Figure 2. Microscopic analysis of R9-GFP-H6-DNA polyplexes. A) Left. Spherical-shaped green fluorescent signal in HeLa cells exposed for 24 hours to R9-GFP-H6-DNA polyplexes. Right. Spherical-shaped blue labels for the same field, corresponding to the embedded DNA. B) Left. Rod-shaped green fluorescent signal in HeLa cells exposed for 24 hours to R9-GFP-H6-DNA polyplexes. Right. The same field, showing blue fluorescence corresponding to the embedded DNA. C) Isosurface representation of polyplexes within a 3D volumetric x-y-z data field, showing the inner localization of the cargo DNA. Magnification increases in the bottom image. D) Superimposition of TMV nanodisks and a R9-GFP-H6 molecular model of a stable, planar oligomer [43]. Arginines in the TMV coat protein are located in a radial distribution surrounding the inner hole (shadowed in yellow, inset), in parallel to those of the R9 tail in R9-GFP-H6 monomers. E) TEM analysis of cell-free R9-GFP-H6 nanoparticles. F) R9-GFP-H6 alone internalized into cultured HeLa cells (upon exposure for 24 h) showing the absence of any associated DNA.

Figure 3. Functional and structural profiling of DNA-loaded nanoparticles. A) Remaining plasmid DNA after treatment with DNAse I, resulting from protection mediated by protein shells at alternative retardation units. Different modular proteins were tested as indicated. At the right, the digestion of protein-free DNA is shown under

493 the same conditions. T indicates time of digestion in min. B) Determination of the z-
494 potential of R9-GFP-H6 nanoparticles, with and without DNA.
495
496 **Figure 4.** Potential intermolecular contacts in R9-GFP-H6 protein oligomers and in R9-
497 GFP-H6-DNA polyplexes. A) Protein-protein model configurations were obtained by
498 docking simulations using HADDOCK at neutral pH, assuming a pentameric
499 composition that is in agreement with experimental size of protein-only particles. The
500 first model (left) was obtained using R9 residues as active and H6 residues as passive
501 [43] and it was used for the superimposition depicted in Figure 2 e. The remaining
502 three models derived from using R9 residues as active and no passive ones. No
503 significant differences in packing were obtained when performing the docking runs at
504 pH 5.8, i.e. with doubly-protonated His (not shown). B) Superposition of the 50
505 solutions with highest score from a (1:1) DNA-protein docking simulation. The structural
506 fitting is based on the DNA molecule, which is shown in red.
507
508
509