DNA-sustained, nanoparticulate architecture of artificial viruses build by multifunctional proteins

Joan Domingo-Espín [1, 2, 3], Esther Vazquez [1, 2, 3], Javier Ganz [4, 5], Oscar Conchillo [1], Elena García-Fruitós [2,3], Juan Cedano [1,6], Ugutz Unzueta [1, 2, 3], Valérie Petegnief [7], Nuria Gonzalez-Montalbán [1,2,3], Anna M Planas [7], Xavier Daura [1, 6, 8], Hugo Peluffo [4, 5], Neus Ferrer-Miralles [1,2,3], Antonio Villaverde [1, 2, 3, *]

[1] *Institute for Biotechnology and Biomedicine, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain*

[2] *Department of Genetics and Microbiology, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain*

[3] *CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, 08193 Barcelona, Spain*

[4] *Neurodegeneration Laboratory, Institut Pasteur de Montevideo, CP 11400 Montevideo, Uruguay*

[5] *Department of Histology and Embryology, Faculty of Medicine, UDELAR, CP 11800 Montevideo, Uruguay*

[6] *Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona. Bellaterra, 08193 Barcelona. Spain*

[7] *Departament d'Isquèmia Cerebral i Neurodegeneració. Institut d'Investigacions Biomèdiques de Barcelona (IIBB), Consejo Superior de Investigaciones Científicas (CSIC)-Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona.*

[8] *Catalan Institution for Research and Advanced Studies (ICREA), 08010 Barcelona. Spain*

*Corresponding author

Running title: Self-organizing, protein-based artificial viruses

Keywords: protein nanoparticles; gene therapy; protein engineering;
nanomedicine; innovative medicines, DNA-protein interactions

Abstract

We have produced in bacteria two chimerical recombinant peptides of 10.2 kDa each that contain four covalently joined, biologically active domains, and that act as building blocks of protein-based artificial viruses for gene therapy. These constructs bind and condense plasmid DNA, penetrate target mammalian cells through RGD cell binding domains and promote significant levels of the reporter transgene expression. To validate the artificial virus concept based on multifunctional proteins we have analyzed the molecular organization of the resulting DNA-protein complexes. While the protein alone aggregates as amorphous, polydisperse clusters of more than 1000 nm, the presence of DNA promotes the architectonic reorganization of these complexes as mechanically stable, monodisperse, nanometric spherical entities of around 80 nm, very efficient in the delivery of expressible transgenes. The structural analysis of the protein in such nanoparticles indicates a molecular conformation of the polypeptide chains with predominance of alpha helix and the absence of cross-molecular, beta sheet-supported protein interactions. The nanoscale organizing forces generated by DNA-protein interactions can be then observed as key elements for the rational design of artificial viruses based on multifunctional proteins, and as a critical but potentially tunable parameter for the successful rational design of non-viral vehicles for gene therapy.

Introduction

Strategies for non-viral gene therapy are under continuous exploration, pressured by the undesired side effects observed in viral-based gene therapy trials [1,2,3,4,5]. In this context, the “artificial virus” approach [6,7,8] implies the use of non infectious bio-safe entities that mimic relevant activities of the viral life cycle, as carriers in the cell-targeted delivery of therapeutic nucleic acids. Liposomes, carbohydrates and proteins are the most used scaffolds for the construction of bio-inspired artificial viruses, although the functionalization necessary for specific receptor binding, endosomal escape and nuclear trafficking, among others, is mostly provided by proteins (namely peptides, full-length proteins or antibodies). In fact proteins, as presented in a diversity of forms, are considered as excellent and fully biocompatible carriers for drug delivery [9]. In this regard, virus-like particles (VLPs) mainly formed by self-assembling capsid proteins from *Papillomaviridae* and *Polyomaviridae* viral families have been explored as gene therapy vehicles (once filled *in vitro* with nucleic acids) [10], either by keeping the original tropism of natural viruses or upon functionalization by the appropriate display of foreign functional peptides. These studies have been also extended to bacterial viruses, which might be more convenient regarding scaled-up production. For instance, phage MS2 VLPs loaded with antisense oligodeoxynucleotides and decorated with transferrin have been proven active on leukemia cancer cells [11]. However, despite the convenient size mono-dispersion exhibited by VLPs, their architectonic constraints limit the tolerance to extensive engineering and the possibility of functional tuning.

A more versatile type of protein-based carriers for therapeutic nucleic acids are multifunctional proteins, constructed by the combination of appropriate functional domains fused in a single polypeptide chain [12]. The integrated domains enable the whole construct to mimic the activities of the infective viral cycle that are relevant to the targeted delivery of nucleic acids (namely DNA or RNA binding, cell attachment and internalization, endosomal escape, proper cytoplasmic trafficking, eventual nuclear transport and nucleic acid release). The modular nature of such constructs permits the selection of functions using relevant peptides identified from nature or combinatorial libraries, and a

functional redesign in iterative improvement processes [13,14]. Diverse protein vehicles within this category have been successful in promoting significant transgene expression levels in vitro [15,16] and therapeutic effects in vivo [17,18], proving the potential of this approach in the clinical context.

Interestingly, non-viral vehicles based on multifunctional proteins have been scantily characterized from the morphologic point of view. Therefore, information about how these proteins might organize as building blocks of higher order structures, and how protein-DNA complexes are formed and shaped is in general not available. Therefore, particle size and molecular organization, nanoscale properties potentially critical for cell attachment, internalization and endosomal escape remain excluded from potential tailoring. To approach this issue, we have produced in *Escherichia coli* two different versions of very short structural proteins as subunits for artificial viruses based on alternative combinations of four functional domains (an integrin-binding motif, an endosomal escape domain, a nuclear localization signal and a DNA-binding, cationic peptide) joined in short peptide stretches. Significant levels of transgene expression driven by the complexes have been observed proving the appropriate selection of the functional domains. On the other hand, in absence of DNA, protein blocks self-organize as amorphous, polydisperse particulate entities ranging from a few nanometers up to around one micron. However, in presence of DNA, protein-DNA complexes appear as tight and rather monodisperse spherical-like nanoparticles of around 80 nm in diameter that resemble bacterial inclusion bodies, in which proteins remain attached by beta-sheet-based cross molecular interactions. However, both protein modeling and structural analysis of these complexes reveal an unexpected molecular organization that does not rely on protein-protein cross-molecular interactions but that is instead stirred by protein-DNA interactions. Such DNA-sustained organization seems to generate an optimal architectural pattern that supports virus-like performance of short multifunctional proteins as building blocks for non-viral gene therapy vehicles.

Materials and methods

Plasmid construction and protein sequence

Plasmid pET28aTEV, derived from pET28a (Invitrogen) in which the DNA sequence encoding the thrombin cleavage site was substituted by a DNA fragment encoding a Tobacco Etch Virus (TEV) protease cleavage site, was used to generate constructs pET28aTEV-HKRN and pET28aTEV-HNRK. HKRN and HNRK correspond to DNA sequences coding for selected modules in the specified order (Figure 1A). Plasmid construction was performed by classical cloning methods introducing synthetic oligonucleotides encoding the corresponding modules into selected restriction enzyme recognition sites of the multiple cloning site of pET28aTEV. The RGD motif used here derives from foot-and-mouth disease virus (serotype C_1) cell binding protein [19], and it is known to bind mammalian cells trough $\alpha_v\beta_3$ and $\alpha_5\beta_1$ integrins [20,21]. The nuclear localization signal of the Simian virus 40 (SV40) large T-antigen [22] has been universally used for the nuclear transport of delivered drugs and DNA [23]. The polysine tail (K10) is a cationic peptide extensively used as a DNA-condensation agent in artificial viruses [24], while the polyhistidine tail (H6) is both an efficient endosomal-escape peptide [14] and a convenient tag for one-step protein purification from bacterial cell extracts [25]. Finally, the biologically irrelevant central amino acid stretch in both HKRN and HNRK was added to enlarge the mass of the resulting modular peptides and to make them more stable in bacterial cells, according to our previous laboratory experience (unpublished data).

Protein production and purification

The production of both chimerical proteins was triggered by the addition of 1 mM IPTG to plasmid-containing BL21(DE3) *E. coli* cell cultures (at OD=0.4-0.6) growing in LB medium at 37ºC. Four hours latter, cells were harvested by centrifugation, washed with PBS and stored at -80 ºC until use. The pellet was resuspended in Lysis Buffer (20 mM Tris-HCl pH 8, 500 mM NaCl and 6 M ClHGu) and cells disrupted by sonication in presence of EDTA-free protease inhibitor cocktail tablets. The soluble fraction was separated by centrifugation at 15,000 g for 45 minutes at 4 ºC and filtered through 0.22 µm filters. Proteins

were purified in a single-step by Ni^{2+} affinity chromatography in an ÄKTA^{TM} FPLC (GE Healthcare) using a 20 CV linear gradient to 100 % of elution buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 6 M ClHGu and 1M Imidazole). Positive fractions were collected and passed through a PD-10 desalting column (GE Healthcare) with Heppes Buffered Saline (HBS) and quantified by Bradford's method. Finally, proteins were stored at -80ºC until use. Inclusion bodies used for Scanning electron microcopy were purified as described elsewhere [26].

Retardation assay

Different protein amounts were incubated with 300 ng of pcDNA3-tdTOMATO plasmid DNA resulting in 0.05, 0.1, 0.25, 0.5, 0.75, 1, 1.5, 2.5, 5 and 7.5 protein/DNA ratios (w/w). Mixtures were incubated in microcentrifuge tubes at room temperature for 1 hour in HBS, and complex formation was detected in 0.8% agarose gels.

Electron microscopy

Complexes of protein and DNA were observed using transmission electron microscopy (TEM) with the aid of negative staining. One drop of the mixture was applied to glow-discharged carbon-coated copper grids (SPI) for 5 min and then drained off with filter paper. Subsequently, one drop of 2 % uranyl acetate was placed on the grid for 2-3 min before being drained off. The grid was then placed in a transmission electron microscope (Jeol JEM 1400) operating at an accelerating voltage of 120 kV. Images were acquired using a CCD camera (Gatan) and saved as 8-bit images. A series of micrograph images were obtained tilting the sample from -60º to +60º with a 914 High Tilt Holder.

Inclusion bodies were analyzed by Scanning Electron Microscopy (SEM) by standard procedures using Quanta FEI 200 filed-emission gun (FEG) environmental scanning electron microscope (ESEM).

Protein structure modeling

The three-dimensional structures of the chimeric peptides were modeled with modeller 9v7 [27] using the coordinates of the original protein segments (when available) as templates. Thus, the structures of the NLS and RGD modules were based on chain B of 1Q1S [28] and chain 5 of 1QGC [29],

respectively. The poly-lysine module was modeled, on the sole basis of the force field, as an unstructured segment, in line with the structural diversity reported for poly-Lys peptides [30] and the disorder of the poly-Lys tail in the structure with PDB code 1KVN [31]. The central region was modeled using the chain A of 1HA0 [32] as a template (61.9% similarity). Hexa-histidine peptides have become one of the most popular tag for protein purification, but the abundance of His-tagged protein models contrast with the lack of structure in which this tag has been successfully solved. This fact clearly indicates that this region tends to be intrinsically unstructured and it was not suitable for modeling under our approach, being then absent in the models.

Dynamic light scattering

Volume-size distribution of DNA-protein complexes at different weight ratios was determined in a dynamic light scattering (DLS) device (Zetasizer Nano ZS, Malvern Instruments Limited) using DTS (Nano) ver. 5.10 software for data evaluation.

Transfection, flow cytometry analysis and fluorescent microscopy

HeLa (ATCC-CCL-2) cell line was maintained in MEM (GIBCO) supplemented with 10 % fetal calf serum (GIBCO) and incubated at 37 ºC and 5 % CO_2 in a humidified atmosphere in 24 wells plates at a cell confluence of 70-80 %. The vectors pcDNA3-tdTOMATO and pEGFP-C1 (Clontech), carrying the gene of the fluorescent proteins tdTOMATO and EGFP respectively were used to monitor DNA transfection. DNA-HKRN or DNA-HNRK complexes were prepared incubating different amounts of protein in 50 µL OPTIPRO medium and different amounts of DNA in 50 µL OPTIPRO (GIBCO) medium. After 5 minutes DNA-protein complexes were generated by mixing DNA and protein at specified protein/DNA ratios at room temperature for 1 hour. 100 µL of OPTIPRO (GIBCO) was then added to the mixture and then to the cells. Transfection and gene expression was monitored by flow cytometry in a FACSCalibur system (Becton Dickinson) at 24 h and confirmed at 48 hours post-transfection in a fluorescence microscope (Nikon ECLIPSE TE2000-E) and. As controls we used non-treated cells, cells exposed only to the protein and cells exposed only to plasmid DNA.

Primary cell cultures

Cortical neuron-enriched cultures were prepared from 18-day-old Sprague–Dawley rat embryos (Charles River Laboratories), as described previously [33]. Animals were anaesthetized and killed by cervical dislocation. All procedures were approved by the Ethical Committee for Animal Use (CEEA) at the University of Barcelona. Cells were seeded on 24-well plates at a density of 1580 cells/mm^2 in Neurobasal medium supplemented with 2 % B27 Supplement, 0.5 mM glutamine and 0.1 mg/ml gentamycin. Partial medium changes were performed on 4 and 7 day *in vitro* (DIV). Transfection was performed on 10 DIV as for HeLa cells, except that the transfection medium was Neurobasal:conditioned medium (2:1). Gene expression was confirmed at 24h post-infection in a fluorescence microscope (Olympus IX71).

Luciferase gene expression

HKRN or HNRK were incubated at room temperature for one hour with pGL3-BOS-luciferase reporter plasmid (kindly provided by Marta Barrachina) at the indicated ratios of protein/DNA in 20-30ul of OPTIMEM medium. Subconfluent HEK293 cells were washed once with OPTIMEM and then incubated with the protein/DNA complexes for 4 hours. Then, the medium was removed and cells maintained in DMEM+10%FBS for another 48 hours. The measurement of luciferase activity was performed according to the manufacturer's instructions (Luciferase Reporter Gene Detection Kit, SIGMA Cat. LUC1-1KT). As a control reference, cells were transfected with Lipofectamine 2000 (Invitrogen, 2 µg lipofectamine + 1 µg DNA/well on 24 well plate) and data were expressed as % relative light units per µg of protein in the samples compared to lipofectamine 2000.

Results

The chimerical genes encoding the multifunctional proteins HKRN and HNRK were constructed by ligation of partially overlapping and complementary oligonucleotides encoding four selected protein domains, and in which the codon usage had been optimized for *E. coli*. Both polypeptides, containing the same functional motifs displayed in alternative positions (Figure 1A), were successfully produced in *E. coli* BL21 (DE3) pLysS, in full-length forms and at reasonably high yield (around 4 µg of protein per ml of culture). Western blot analyses of purified proteins revealed the absence of truncated protein versions and the minor occurrence of high molecular mass immunoreactive species, especially in HKRN, what might indicate a tendency to form supramolecular structures (Figure 1C). When HKRN and HNRK were challenged in DNA retardation assays, HKRN showed a higher capability (one retardation unit corresponding to a protein/DNA mass ratio of 1.5) than HNRK (one retardation unit corresponding to a protein/DNA mass ratio of 2.5) to impede the mobility of plasmid DNA (Figure 1B). This divergence could be accounted by either a different oligomerization potential or by a different performance of the DNA binding domain (K10) as alternatively positioned in HKRN and HNRK. In the first case K10 was placed in an internal position within the amino terminal protein moiety and in HNRK, this peptide overhanged as a carboxy terminal end.

The resulting protein-DNA complexes (artificial viruses) were tested in HeLa cell cultures for their ability to promote expression of a plasmid-harbored reporter transgene. Although the design of non-viral vehicles for gene therapy is a rather trial-an-error process, we expected that the combination of the FMDV integrin-binding motif, the SV40 nuclear localization signal, the His-based endosomal escape peptide and the Lys-based DNA binding stretch could summarize the main viral functions required for cell uptake and trafficking of the cargo DNA and result in significant levels of nuclear gene delivery and expression. In agreement with this presumption, flow cytometry analysis of cultured cells 48 h after exposure to HKRN- and HNRK-DNA complexes revealed the occurrence of significantly prevalent cell subpopulations expressing the reporter *td tomato* gene. In this context, more than 10 % of HeLa cells transfected with HNRK-

based vehicles emitted red fluorescence, indicating the proper nuclear delivery and release of the carried DNA. However, being still significant, DNA delivery mediated by HKRN resulted in rather moderate transgene expression that was detected only 0.5 % in the cell population (Figure 2A). To discard that this value could be due to experimental noise we examined in situ the cultures treated with HKRN-based complexes by fluorescence microscopy, detecting clear fluorescence emission in individual cultured HeLa cells when using two different reporter genes, namely *EGFP* and *td tomato* (involving more than 10 % of cells, at 24 h; Figure 2B, top, medium). Furthermore, in primary cultures of neurons and glia several cells strongly expressing *tdTomato* were observed 24 h after transfection with the DNA-HKRN complex. A cell with neuronal morphology strongly expressing the *tdTomato* gene in the cell body and neurites is shown in the inset of Figure 2B, bottom, demonstrating that neurons can be effectively transfected and the transgene transcribed and translated into protein. An additional transfection experiment on Hek293 cells with a third reporter luciferase (*luc*) gene confirmed the consistent transgene expression mediated by HKRN (Figure 2C). These data demonstrate the stability, robustness and good performance of both HKRN and HNRK as non-viral gene vectors and the appropriateness of the selected protein modules to mediate DNA delivery, being the modular distribution in HNRK more convenient for the proper mimicking of viral functions.

Intriguingly, the morphology and structure of protein-DNA complexes in non viral gene therapy has been historically neglected, and for protein-based vehicles other than VLPs, the concept of artificial virus refers exclusively to functional (instead of nanoscale physical) properties. Therefore, at this stage, we were especially interested in evaluating the architectonic properties of both constructs as building blocks of artificial viruses, and how these multifunctional protein subunits could organize to hold the plasmid DNA. To explore the molecular organization of the artificial viruses we approached their structural analysis from different angles. Interestingly, the TEM images of both peptides alone indicated the occurrence of amorphous, highly dispersed protein clusters or around 1 μm without any apparent morphological pattern and internal organization (Figure 3A). However, the artificial viruses formed by HKRN and

HNRK organized as regular, pseudo-spherical nanoparticles or around 80 nm in diameter (Figures 3A), morphologically reminding the inclusion bodies commonly observed in recombinant bacteria [34,35,36] (although these last particles can be slightly larger, up to 450 nm in diameter [34]). The molecular reorganization of the protein building blocks induced by the addition of DNA occurred at 0.5 but not 2 RU (Figure 3B), and it did not prevent the parallel emergence of larger protein clusters (Figure 3B). These micron sized particles, as seen by DLS, are probably clusters of the 80 nm particles promoted by DNA, as only these smaller structures were clearly identified by TEM (Figure 3A). Despited the absence of nanosized particles at 2 RU, the size variability of DNA-protein complexes was strongly reduced when comparing with proteins alone (Figure 3B), indicating that the presence of DNA promoted conformational alterations on the holding proteins with impact in the oligomeric organization. The regularity of size in the protein-DNA complexes as compared with the protein alone also indicates protein-condensing abilities of plasmid DNA that reduce the molecular stickiness (their aggregation tendency) of HKRN and HNRK proteins. This fact, strongly suggested that the cationic poly-lysine stretches, responsible for DNA binding in multifunctional proteins [24] and whose charge is expected to be neutralized in the complexes, effectively drive the unspecific formation of higher order, protein-alone clusters shown in Figure 3A. Taken together, all these data indicate that HKRN and HNRK, apart from exhibiting the functions associated to their forming protein domains, act as efficient building blocks for the construction of artificial viruses under the architectonic scope of this term.

In the regard of inclusion body resemblance, a ±60º TEM scan of HNRK revealed a slightly flattered ellipsoid form of the protein-DNA complexes (Figure 4A), again very similar to the images of inclusion bodies formed by other proteins seen by Atomic Force Microscopy (AFM) [34]. In fact, HKRN and HNRK themselves are both partially found as inclusion bodies in the cytoplasm of the producing bacteria (Figure 4C). As determined by conformational analysis through FTIR [37,38,39], inclusion bodies gain their mechanical structure and shape by cross-molecular protein-protein interactions sustained by a beta-sheet-based, amyloid-like architecture [40,41]. At this stage, we wondered if the

architecture of the 80 nm-artificial viruses formed by HKRN- and HNRK-DNA could be also supported by protein-protein interactions, and therefore, we approached the FTIR analysis of artificial viruses and their protein building blocs taking HNRK as a model.

The conformational status of HNRK inclusion bodies was found to be similar to those described previously as formed by other recombinant proteins, and characterized by the presence of extended, crossmolecular β-pleated sheet elements peaking at 1621 cm^{-1} (Figure 4 B, top) [42,38,41]. These self-organizing structures coexist with other secondary elements such as native alpha-helices and unordered structures (corresponding to the overlapped region between 1640 and 1660 cm^{-1}). In solution, HNRK was instead characterized by the presence of alpha-helix elements peaking at 1654 cm^{-1} (Figure 4 B, center, green line), in agreement with that observed by its *in silico* modeling (Figure 4D). Note that although as expected for short peptides, both HKRN and HNRK are in general unstructured, some locally structured regions inherited from their templates were however noted in the models, namely a 3-10 helix spanning residues 44 to 47 in HNRK and 45 to 48 in HKRN, apart from some additional turns and bends (Figure 4D). Interestingly, upon lyophilizing, HNRK seemed to evolve in a more lightly loose and unordered structure, as it can be seen by the broad peak between 1640 and 1660 cm^{-1}, corresponding to alpha-helices and unordered structures overlapped region (Figure 4B, center, black line).

To test the molecular organization of HNRK-DNA particles and the potential molecular resemblance to HNRK when forming inclusion bodies, these artificial viruses were lyophilized and analyzed as above. Interestingly, and in agreement with the structural impact of DNA on the complexes suggested by the DLS data (Figure 3), the presence of the plasmid DNA had a critical effect on the peptide structure (Figure 4 B, bottom), preventing the smooth deconstruction of alpha-helices observed during the liophilization of HNRK alone. Also, HNRK alpha-helices gained looseness along with the increase of DNA: HNRK ratio, as can be seen by the slight shift from lower wavenumber, from 1653 cm^{-1} in the lyophilized sample without DNA (Figure 4B, bottom, black line) to 1651 and 1650 cm^{-1} in the HNRK 2 RU and 0.5 RU (Figure 4 B, bottom, red line and blue

line respectively). This minor but significant shift might suggest that the binding of DNA to the protein shells is not a random but organized event possibly involving the central alpha-helix region of the peptide. Such interaction could account for the architectonic organization emerging in the artificial viruses and absent in the protein building blocs alone.

In summary, the FTIR analysis discarded any inclusion-body like organization of artificial viruses and demonstrated that the architecture of these particles is not based on cross-molecular protein-protein contacts but that it is instead sustained by charge-dependent, but potentially sterospecific DNA-protein interactions. These interactions generate artificial viruses able to transfect expressible DNA, with morphologies and sizes within the nanoscale and compatible with those found optimal for efficient cell interaction and further uptake (in the range of those exhibited by natural virus particles) [43,44,45]. A further evidence of the architectonic role of DNA in the organization of artificial viruses is that, upon treatment with DNAse, the HNRK-based artificial viruses disassemble in smaller entities whose lower range sizes (a few nm), are compatible with those of individual polypeptides (Figure 3B).

Discussion

Artificial viruses are manmade constructs designed to mimic viral activities relevant to the cell targeted delivery of therapeutic nucleic acids [7], thus representing safer alternatives to viral gene therapy [6,2]. Lipids and polysaccharides with different molecular organizations are commonly used as shells for nucleic acids acting as particulate cores of artificial viruses. On the other side, because of the ability of proteins to interact with specific ligands, these vehicles are often functionalized with antibodies, peptides or whole proteins in an attempt to reach the target cell type or tissue. Although tissue targeting in drug delivery can also be effectively achieved by distally applying magnetic force on paramagnetic drug carriers [46], the versatility of protein engineering offers unique opportunities for the fine tailoring of the biological properties of artificial viruses to attain, for instance, complex biodistribution maps.

In the context of the tunable nature of proteins, artificial viruses can be efficiently constructed by uniquely using these macromolecules, provided all the functions required by nucleic acid condensation and intracellular delivery are embraced. In this regard, an intriguing approach to protein-based artificial viruses is the design of multifunctional proteins [12], that contain, in a single polypeptide chain, functional peptides from different origins. Upon selected and combined in a modular way, these functional units confer proper cellular specificity and intracellular traffic to the DNA-protein complexes [13,14]. Those functional peptides can either be inserted in permissive sites of a scaffold protein, or sequentially fused as a new, non natural peptide or short protein [47]. Examples of constructs generated by these alternative strategies can be found elsewhere [12].

From the material science point of view, the organization of protein-based cages has been classified according to rather general schemes [48,9], but the precise architecture of proteinaceous artificial viruses other than based on VLPs remains poorly explored. In fact, multifunctional proteins based on large scaffold proteins such as for instance *E. coli* beta-galactosidase [49,50], organize as amorphous polydisperse protein clusters whose properties seemed to be

defined by protein features (the enzyme is a tetramer of around 460 kDa, [51]) rather than by the presence of DNA [52]. Upon addition, plasmid DNA does not modify the morphology of the complexes. In the same context, arginine rich peptides (Rn), when displayed on the surface of a chimerical GFP, provide self-assembling properties to the fusion protein (rendering planar 20 nm-particles) also irrespectively of the presence of DNA [15]. Here we have explored the nanoscale organization of two short multifunctional proteins, namely HKRN and HNRK (Figure 1), which are shown to be competent in gene delivery by using both cultured cell lines and primary cell culture models (Figure 2 A,B). The reached transgene expression levels and stability are comparable or higher than those observed with previous prototypes of artificial viruses based on multifunctional proteins [49,53,15,52,50], being in the case of the less active building block, namely the construct HKRN, around 18 % than those achieved by using lipofectamine (Figure 2 C). The slight differences in the ability to retain and deliver expressible DNA are obviously due to the alternative disposition of functional motifs, and the end terminal location of the cationic K10 peptide seem to be specially convenient for the performance of the whole vehicle. These proteins have been designed on the basis of a mere sequential fusion of functional domains without any scaffolding protein, a strategy that resulted favorable regarding their productivity in bacteria, when comparing with the moderate yield in which high molecular mass engineered beta-galactosidases had been obtained ([49,52]).

The building blocks alone tend to passively aggregate as amorphous clusters with averages sizes of around 1 µm (Figure 3). However, the presence of DNA dramatically modifies the organization of the protein and at 0.5 RU, it induces the formation of protein-DNA nanoparticles of around 80 nm from which DNA molecules eventually overhang (Figures 3A and 4A). These artificial viruses, having optimal size regarding their potential interaction with mammalian cells and further uptake [43], are able to promote the transgene expression in targeted cultured cells, as observed by several models (Figure 2), more efficiently than other vehicles constructed by us in the past, based on larger scaffold proteins and showing amorphous organization [49,52].

Interestingly, the organization of these regular protein-DNA complexes based on the short HKRN and HNRK multifunctional peptides is not dependent on protein-protein interactions but on the sticky, glue-like potential of DNA (Figure 3B, Figure 4B), that seems to show some avidity for the internal alpha-helix exhibited by both proteins (Figure 4D). The architectonic properties of DNA in creating regular nanoparticles, based on charge-dependent interactions [24], strongly depend on the protein-DNA ratio (Figure 3B) and are probably more relevant when interacting with short peptides than with large proteins, since in this last case, no DNA-induced architectonic changes in similar but larger protein building shells have been previously reported [15,52]. In this context, the particle size (80 nm) observed here by using two short chimerical proteins has resulted very similar to that observed when associating other short peptides with plasmid DNA, namely in adenoviral core peptide mu-DNA complexes (80-120 nm, [54,55]) and in intermediates in toroid formation by histidylated polylysine-DNA complexes (80-100 nm, [56]). Polylysine- and polyornithine-DNA polyplexes have rendered, however, slightly larger particles (150-200 nm, [57]). These organizing forces are probably depending on the ability of DNA to alter the conformation of the shell proteins (Figure 4). In this context, it has been previously proved, by elegant analysis, that short peptides affect the local and distal secondary and tertiary structure of bound DNA [58], but according the data presented here the conformational changes in protein-DNA artificial viruses are mutually induced.

The nanometric organizing abilities of DNA-multifunctional protein complexes, reported for the first time in this study, opens intriguing possibilities for the design and development of artificial viruses intended for gene therapy. The small size of the protein counterpart seems to be a key element that facilitates the DNA-promoted particle self-organization, while the position of the functional domains within the multifunctional protein seems to have only a moderate impact on the vehicle performance, manly linked to the ability of the protein to interact with the DNA trough the cationic motif (Figure 1). The possibility to generate monodisperse, viral-sized artificial viruses by engineering DNA-protein interactions is offering a promising alternative to the engineering of VLP for gene therapy, whose main advantage is their architectonic regularity. The

functional plasticity instead offered by the multifunctional protein approach, combined with the particle size adjustment that a further comprehension of DNA-protein morphogenetic forces should permit, open a road of investigation to generate chemically hybrid, novel and improved bio-nanoparticles for nucleic acid but also conventional drug delivery.

Conclusions

We have biologically produced short, mainly unordered multifunctional peptides as building blocks of protein-based artificial viruses, which have shown an excellent performance in transgene delivery under different biological models. Interestingly, the artificial viruses resulting from protein-DNA association are pseudo-spherical entities with regular particle sizes or around 80 nm, at specific protein-DNA rations in the range of those promoting high transgene expression levels. A deep structural characterization of the protein components in these artificial viruses has revealed that the global architecture of the particles is not supported by protein-protein interactions but on the contrarily, unexpectedly sustained by the DNA. The nucleic acids acts as a compacting, molecular glue that affects the conformation of the protein building blocks, altering the alpha helix structure of the central region, minimizing their aggregation tendency and promoting an ordered, self-organization of the complexes in sizes compatible with an efficient receptor-mediated cell uptake and proper intracellular trafficking to the cell nucleus. This first description of the architectonic properties of DNA at the nanoscale opens intriguing opportunities for a better rational design of artificial viruses for gene therapy regarding their molecular and physical organization.

Executive summary

- Short chimerical proteins produced in bacteria, that contain four functional domains relevant to intracellular trafficking promote high transgene expression levels when used as artificial viruses.
- The presence of DNA promotes conformational changes in the protein moiety of the artificial viruses that affects the minor alpha helix region exhibited by those rather unstructured peptides.

- The resulting artificial viruses are pseudo-spherical stable particles of around 80 nm, fully sustained by DNA-protein interactions rather than by protein-protein crossmolecular beta-sheet interactions, which at difference from protein-only aggregates, are undetectable.

Acknowledgments: We are indebted to the Cell Culture and Citometry Units of the Servei de Cultius Cel·lulars, Producció d'Anticossos i Citometria (SCAC), and to the Servei de Microscòpia, both at the UAB. We also appreciate the financial support received for the design and production of artificial viruses for gene therapy from MICINN (BIO2007-61194 and ACI2009-0919), AGAUR (2009SGR-108) and CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN, Spain), 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. We also thank ANII, Ministerio de Educación y Cultura, Uruguay, for financial support. AV has been distinguished with an ICREA ACADEMIA award.

Figure legends:

Figure 1. HKRN and HNRK organization and main features. (A) Distribution of functional modules in HKRN and HNRK and amino acid sequence of the whole protein constructs. In both cartoons and sequence, the histidine (H) tail is labeled in red, the lysine (K) tail in blue, the FMDV cell binding (RGD) in purple and the SV40 nuclear localization signal (NLS) in green. The irrelevant central region is depicted in grey. In the amino acid sequences, residues resulting from the cloning process are underlined, a TEV protease target site introduced between H and the immediate carboxy module is indicated in boldface and the RGD motif within the FMDV peptide is shown in italics. Sizes of the modules in the cartoons are not intended to be representative of the actual length in the protein segments. **(B)** Up-shift of pcDNA 3.1 (harboring the *td tomato* gene) mobility in agarose gel electrophoresis as induced by HKRN and HNRK. The

protein/DNA ratios at which migration of DNA is fully impeded (one retardation unit, RU), are indicated by an asterisk. **(C)** Comassie-blue (Cb) staining and Western-blot (Wb) analysis (using an anti-His antibody) of HKRN and HNRK upon purification. Molecular masses of the markers (M) are indicated in the central column.

Figure 2. Transgene expression mediated by HKRN- and HNRK-based artificial viruses. (A) Fluorescence emission determined by flow cytometry of cultured HeLa cells 48 h after exposure to 24 µg of pcDNA 3.1 or in absence of foreign DNA (**no protein**). Cells were also exposed to HKRN-and HNRK-DNA complexes and to these proteins alone (**protein**), and the P2 section in the plots corresponds to the red fluorescence emitted by the td tomato protein. The percentages of fluorescent cells are indicated above each plot. **(B)** Fluorescence microscopy of cultured cells 24 h after exposure to HKRN- DNA complexes formed at 2 RU. **(C)** Light emission of HEK293 cells 48 h after exposure to HKRN-pBOS complexes formed at several RU (28,8 µg HKRN/well). pBOS alone at a concentration equivalent to 2 RU (5.75 µg/well) or Lipofectamine 2000 (2 µg/well + 1 µg pBOS) were used as a controls.

Figure 3. Size and morphology of HKRN and HNRK and their derived artificial viruses. (A) Representative transmission electron microscopy images of both HKRN and HNRK alone and as complexes with plasmid pcDNA3.1. **(B)** Effective size of protein particles alone or protein-DNA complexes (HNRK in top and HKRN in bottom) determined by dynamic light scattering. The size of HNRK-based artificial viruses treated with DNAse (7 µg/ml) for 30 min at 37 ºC is also shown (medium).

Figure 4. Molecular organization and conformation of HKRN and HNRK building blocks. (A) Selection of TEM micrograph images from a series of images acquired at inclinations from -60º to +60º. **(B)** Second derivative ATR-FTIR absorption spectra in the Amide I region ranging from 1750 to 1550 cm^{-1}. Vertical lines correspond to aggregated β-strands, unordered structures and α-helices peaking at 1621, 1648 and 1653 cm^{-1} respectively. **Top**, Second

derivative spectrum of NHRP inclusion bodies. **Medium**, Second derivative of lyophilized (solid black line) and soluble NHPR (solid grey line). **Bottom**, Second derivative spectra of lyophilized NHPR (solid black line), NHPR 2 RU (solid grey line) and NHPR 0.5 RU (dashed black line). **(C)** HNRK- and NHRP-producing *E. coli* cells showing cytoplasmic inclusion bodies (left), and those inclusion bodies as observed by SEM upon purification (right). **(D)** HNRK- and NHRP models in which the different modules are colored according to the color pattern of Figure 1 A. Note that the His segments are not shown.

References

1. Smaglik P: Clinical trials end at gene-therapy institute.. *Nature* 405(6786), 497- (2000).

2. Edelstein ML, Abedi MR, Wixon J: Gene therapy clinical trials worldwide to 2007--an update. *J.Gene Med.* 9(10), 833-842 (2007).

3. Marshall E: Gene therapy. Second child in French trial is found to have leukemia. *Science* 299(5605), 320- (2003).

4. Marshall E: Gene therapy. What to do when clear success comes with an unclear risk? *Science* 298(5593), 510-511 (2002).

5. Smaglik P: Tighter watch urged on adenoviral vectors...with proposal to report all 'adverse events'. *Nature* 402(6763), 707- (1999).

6. Douglas KL: Toward Development of Artificial Viruses for Gene Therapy: A Comparative Evaluation of Viral and Non-viral Transfection. *Biotechnol.Prog.* (2008).

7. Mastrobattista E, van der Aa MA, Hennink WE, Crommelin DJ: Artificial viruses: a nanotechnological approach to gene delivery. *Nat.Rev.Drug Discov.* 5(2), 115-121 (2006).

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

9. Maham A, Tang Z, Wu H, Wang J, Lin Y: Protein-Based Nanomedicine Platforms for Drug Delivery. *Small* (2009).

10. Petry H, Goldmann C, Ast O, Luke W: The use of virus-like particles for gene transfer. *Curr.Opin.Mol.Ther.* 5(5), 524-528 (2003).

11. Wu M, Sherwin T, Brown WL, Stockley PG: Delivery of antisense oligonucleotides to leukemia cells by RNA bacteriophage capsids. *Nanomedicine.* 1(1), 67-76 (2005).

12. Aris A , Villaverde A: Modular protein engineering for non-viral gene therapy. *Trends Biotechnol.* 22(7), 371-377 (2004).

13. Vazquez E, Ferrer-Miralles N, Villaverde A: Peptide-assisted traffic engineering for nonviral gene therapy. *Drug Discov.Today* 13(23-24), 1067-1074 (2008).

14. Ferrer-Miralles N, Vazquez E, Villaverde A: Membrane-active peptides for non-viral gene therapy: making the safest easier. *Trends Biotechnol.* 26(5), 267-275 (2008).

15. Vazquez E, Roldán M, Diez-Gil C, Unzueta U, Domingo-Espin J, Cedano J, Conchillo-Sole O, Ratera I, Veciana J, Ferrer-Miralles N, and Villaverde A. Protein nanodisk assembling and intracellular trafficking powered by an arginine-rich (R9) peptide. *Nanomedicine*5, 259-268 (2010)

16. Xavier J, Singh S, Dean DA, Rao NM, Gopal V: Designed multi-domain protein as a carrier of nucleic acids into cells. *J.Control Release* 133(2), 154-160 (2009).

17. Peluffo H, Acarin L, Aris A *et al:* Neuroprotection from NMDA excitotoxic lesion by Cu/Zn superoxide dismutase gene delivery to the postnatal rat brain by a modular protein vector. *BMC.Neurosci.* 7, 35- (2006).

18. Peluffo H, Aris A, Acarin L, Gonzalez B, Villaverde A, Castellano B: Nonviral gene delivery to the central nervous system based on a novel integrin-targeting multifunctional protein. *Hum.Gene Ther.* 14(13), 1215-1223 (2003).

19. Benito A, Mateu MG, Villaverde A: Improved mimicry of a foot-and-mouth disease virus antigenic site by a viral peptide displayed on beta-galactosidase surface. *Biotechnology (N Y)* 13(8), 801-804 (1995).

20. Feliu JX, Benito A, Oliva B, Aviles FX, Villaverde A: Conformational flexibility in a highly mobile protein loop of foot-and-mouth disease virus: distinct structural requirements for integrin and antibody binding. *J Mol Biol* 283(2), 331-338 (1998).

21. Villaverde A, Feliu JX, Harbottle RP, Benito A, Coutelle C: A recombinant, arginine-glycine-aspartic acid (RGD) motif from foot-and-mouth disease virus binds mammalian cells through vitronectin and, to a lower extent, fibronectin receptors. *Gene* 180(1-2), 101-106 (1996).

22. Yoneda Y: How proteins are transported from cytoplasm to the nucleus. *J.Biochem.* 121(5), 811-817 (1997).

23. Collins E, Birchall JC, Williams JL, Gumbleton M: Nuclear localisation and pDNA condensation in non-viral gene delivery. *J.Gene Med.* 9(4), 265-274 (2007).

24. Saccardo P, Villaverde A, Gonzalez-Montalban N: Peptide-mediated DNA condensation for non-viral gene therapy. *Biotechnol.Adv.* 27(4), 432-438 (2009).

25. Block H, Maertens B, Spriestersbach A *et al:* Immobilized-metal affinity chromatography (IMAC): a review. *Methods Enzymol.* 463, 439-473 (2009).

26. Garcia-Fruitos E, Seras-Franzoso J, Vazquez E, Villaverde A: Tunable geometry of bacterial inclusion bodies as substrate materials for tissue engineering. *Nanotechnology.* 21(20), 205101- (2010).

27. 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)

28. Fontes MR, Teh T, Toth G *et al:* Role of flanking sequences and phosphorylation in the recognition of the simian-virus-40 large T-antigen nuclear localization sequences by importin-alpha. *Biochem.J.* 375(Pt 2), 339-349 (2003).

29. Hewat EA, Verdaguer N, Fita I *et al:* Structure of the complex of an Fab fragment of a neutralizing antibody with foot-and-mouth disease virus: positioning of a highly mobile antigenic loop. *EMBO J.* 16(7), 1492-1500 (1997).

30. Davidson B , Fasman GD: The conformational transitions of uncharged poly-L-lysine. Alpha helix-random coil-beta structure. *Biochemistry* 6(6), 1616-1629 (1967).

31. Pakhomova ON, Deep S, Huang Q, Zwieb C, Hinck AP: Solution structure of protein SRP19 of Archaeoglobus fulgidus signal recognition particle. *J.Mol.Biol.* 317(1), 145-158 (2002).

32. Chen J, Lee KH, Steinhauer DA, Stevens DJ, Skehel JJ, Wiley DC: Structure of the hemagglutinin precursor cleavage site, a determinant of influenza pathogenicity and the origin of the labile conformation. *Cell* 95(3), 409-417 (1998).

33. Petegnief V, Friguls B, Sanfeliu C, Sunol C, Planas AM: Transforming growth factor-alpha attenuates N-methyl-D-aspartic acid toxicity in cortical cultures by preventing protein synthesis inhibition through an Erk1/2-dependent mechanism. *J.Biol.Chem.* 278(32), 29552-29559 (2003).

34. Garcia-Fruitos E, Rodriguez-Carmona E, Diez-Gil C *et al:* Surface Cell Growth Engineering Assisted by a Novel Bacterial Nanomaterial. *Advanced Materials* 21(42), 4249-4253 (2009).

35. de Marco A , Schroedel A: Characterization of the aggregates formed during recombinant protein expression in bacteria. *BMC.Biochem* 6(1), 10- (2005).

36. Martinez-Alonso M, Gonzalez-Montalban N, Garcia-Fruitos E, Villaverde A: The functional quality of soluble recombinant polypeptides produced in Escherichia coli is defined by a wide conformational spectrum. *Appl.Environ.Microbiol.* 101(6), 1353-1358 (2008).

37. Doglia SM, Ami D, Natalello A, Gatti-Lafranconi P, Lotti M: Fourier transform infrared spectroscopy analysis of the conformational quality of recombinant proteins within inclusion bodies. *Biotechnol.J.* 3(2), 193-201 (2008).

38. Gonzalez-Montalban N, Natalello A, Garcia-Fruitos E, Villaverde A, Doglia SM: In situ protein folding and activation in bacterial inclusion bodies. *Biotechnol.Bioeng.* 100(4), 797-802 (2008).

39. Ami D, Natalello A, Taylor G, Tonon G, Maria DS: Structural analysis of protein inclusion bodies by Fourier transform infrared microspectroscopy. *Biochim.Biophys.Acta* 1764(4), 793-799 (2006).

40. Wang L, Maji SK, Sawaya MR, Eisenberg D, Riek R: Bacterial inclusion bodies contain amyloid-like structure. *PLoS.Biol.* 6(8), e195- (2008).

41. Carrio M, Gonzalez-Montalban N, Vera A, Villaverde A, Ventura S: Amyloid-like properties of bacterial inclusion bodies. *J.Mol.Biol.* 347(5), 1025-1037 (2005).

42. Natalello A, Ami D, Brocca S, Lotti M, Doglia SM: Secondary structure, conformational stability and glycosylation of a recombinant Candida rugosa lipase studied by Fourier-transform infrared spectroscopy. *Biochem.J.* 385(Pt 2), 511-517 (2005).

43. Jiang W, Kim BY, Rutka JT, Chan WC: Nanoparticle-mediated cellular response is size-dependent. *Nat.Nanotechnol.* 3(3), 145-150 (2008).

44. Chithrani BD, Ghazani AA, Chan WC: Determining the size and shape dependence of gold nanoparticle uptake into mammalian cells. *Nano Lett.* 6(4), 662-668 (2006).

45. Jiang W, Kim BY, Rutka JT, Chan WC: Advances and challenges of nanotechnology-based drug delivery systems. *Expert.Opin.Drug Deliv.* 4(6), 621-633 (2007).

46. Corchero JL , Villaverde A: Biomedical applications of distally controlled magnetic nanoparticles. *Trends Biotechnol.* 27(8), 468-476 (2009).

47. Vazquez E, Ferrer-Miralles N, Mangues R, Corchero JL, Schwartz S Jr, Villaverde A: Modular protein engineering in emerging cancer therapies. *Curr.Pharm.Des* 15(8), 893-916 (2009).

48. Uchida M, Klem MT, Allen M *et al:* Biological containers: Protein cages as multifunctional nanoplatforms. *Adv.Mater* 19, 1025-1042 (2007).

49. Aris A, Feliu JX, Knight A, Coutelle C, Villaverde A: Exploiting viral cell-targeting abilities in a single polypeptide, non-infectious, recombinant vehicle for integrin-mediated DNA delivery and gene expression. *Biotechnol Bioeng* 68(6), 689-696 (2000).

50. Aris A , Villaverde A: Engineering nuclear localization signals in modular protein vehicles for gene therapy. *Biochem.Biophys.Res.Commun.* 304(4), 625-631 (2003).

51. Jacobson RH, Zhang XJ, DuBose RF, Matthews BW: Three-dimensional structure of beta-galactosidase from E. coli. *Nature* 369(6483), 761-766 (1994).

52. Aris A , Villaverde A: Molecular organization of protein-DNA complexes for cell-targeted DNA delivery. *Biochem.Biophys.Res.Commun.* 278(2), 455-461 (2000).

53. Peluffo H, Gonzalez P, Aris A *et al:* RGD domains neuroprotect the immature brain by a glial-dependent mechanism. *Ann.Neurol.* 62(3), 251-261 (2007).

54. Tagawa T, Manvell M, Brown N *et al:* Characterisation of LMD virus-like nanoparticles self-assembled from cationic liposomes, adenovirus core peptide mu and plasmid DNA. *Gene Ther.* 9(9), 564-576 (2002).

55. Keller M, Tagawa T, Preuss M, Miller AD: Biophysical characterization of the DNA binding and condensing properties of adenoviral core peptide mu. *Biochemistry* 41(2), 652-659 (2002).

56. Midoux P, LeCam E, Coulaud D, Delain E, Pichon C: Histidine containing peptides and polypeptides as nucleic acid vectors. *Somat.Cell Mol.Genet.* 27(1-6), 27-47 (2002).

57. Ramsay E, Hadgraft J, Birchall J, Gumbleton M: Examination of the biophysical interaction between plasmid DNA and the polycations, polylysine and polyornithine, as a basis for their differential gene transfection in-vitro. *Int.J.Pharm.* 210(1-2), 97-107 (2000).

58. Tecle M, Preuss M, Miller AD: Kinetic study of DNA condensation by cationic peptides used in nonviral gene therapy: analogy of DNA condensation to protein folding. *Biochemistry* 42(35), 10343-10347 (2003).

Figure 1

Figure 2

Figure 3

Figura 4