Development of artificial viruses for nanomedicine and gene therapy.

Paolo Saccardo

PhD Thesis

2015

PhD program in Biotechnology

Development of artificial viruses for nanomedicine and gene therapy

PhD thesis 2015

Departament de Genètica i Microbiologia
Facultat de Biociències

Report presented by Paolo Saccardo in order to complete the requirements to be granted the degree of Doctor of Philosophy in Biotechnology by the Autonomous University of Barcelona.

Paolo Saccardo

Approval of the thesis directors:

Antonio Villaverde Corrales **Neus Ferrer-Miralles** **Escarlata Rodríguez-Carmona**

“O frati”, dissi, “che per cento milia
perigli siete giunti a l'occidente,
a questa tanto picciola vigilia

d'i nostri sensi ch'è del rimanente
non vogliate negar l'esperienza,
di retro al sol, del mondo sanza gente.

Considerate la vostra semenza:
fatti non foste a viver come bruti,
ma per seguir virtute e canoscenza”.

Dante Alighieri,
Divina Commedia - Inferno
canto XXVI, vv. 112-120.

A mio papá, mia mamma e Anna.

Contents

Introduction

Along the history of the humanity, self-awareness, survival instinct and curiosity have acted as driving forces for human evolution. Unwittingly discovery and rational observation of phenomena permitted to develop techniques and to acquire knowledge that is reflected in the actual everyday life comfort as technological advances.

Humans have always observed and used the most powerful and complete tool they have had within reach: nature. The rational handling of the biosphere components to obtain benefits, perhaps, started from seed selection for improving crops' yield and quality. Moreover, microbial enzymatic activity was first used in fermentation processes to obtain wine, beer and bread. This primitive's exploitation of biological tools in specific processes can be considered the first steps of a science brunch known as Biotechnology.

Nowadays, biotechnology is applied to a wide spectrum of fields like industrial processes, cultivation, material development, advanced medicine and chemistry. Last century has been spectator of the major advances in biotechnology being the main milestone the discovery of the DNA double helix structure by James Watson and Francis Crick in 1953[1]. Since the very first moment it was clear that the modeling of the DNA structure would be the first step that would open the way to deeply understand the living cell functioning and also for modern and breakthrough biotechnological applications.

In the last sixty years, remarkable investigations have provided extraordinary knowledge advances. As a cascade of discovery, DNA structure description permits the understanding of the molecular genetic inheritance mechanism[2], the cracking of the DNA code[3], the first experiments with recombinant DNA[4], the Sanger's DNA sequencing method[5], the implementation of PCR[6], and the production of first recombinant proteins licensed as drug[7], among others. In February 2001, the Human Genome Project announced that more than 90% of human genome was sequenced[8,9]. The wide and extended information obtained from genomic researches with bioinformatic tools, permit a fast evolution of proteomic science.

Proteins are recognized as main buildings blocks of living cells. They are obtained by the information encoded in the DNA sequence as different combinations of nucleotide triplets, meaning a linear amino acidic chain peptide. Once synthesis is completed, the polypeptide chain folds in a tridimensional structure, acquiring a specific biological activity.

To understand the molecular basis of cell biology is then necessary a deeply comprehension of the DNA expression regulation and the relationship between biological molecules at nano-scale level. It's now clear that gene dysregulation or anomaly in cellular environment can lead to a disease condition.

Once again, biotechnology, joined to nanomedicine, can help to find a solution for an increasing life quality of patients. It's well recognized that therapies should be effective, safe, non-invasive, and cell selective. In this context, recombinant protein engineering, production and characterization, play a key role being proteins a highly tunable biomaterial suitable for therapeutic purposes. Recombinant proteins, in fact, can act as delivery system for therapeutic molecules [10]. Since a good drug is not effective if it's not able to reach the cell target, protein nano-carriers development represents a really interesting and challenging research field.

1 Protein production

Recombinant protein production is a widely used technique in biotechnology. A lot of expression systems with an extensive set of features are now available allowing an optimized and tunable strategy of expression[11]. This is really important since every protein is different and choosing the correct condition for protein production is a key step for achieving a high quality result.

In nature, protein synthesis is finely regulated and depends on the cell's functional needs. External *stimuli* or intracellular requirements can lead to the activation or repression of the protein production process. In short, protein production consists in the transcription of the information encoded in the DNA, first into an mRNA molecule upon a fine regulated process involving transcription factors, and then from mRNA to an amino acidic sequence constituting the primary structure of the protein. Activators, repressors or chromatin structure modification act like switchers in order to activate or repress the transcription. In the initial step, the DNA molecule is unwound, permitting the RNA polymerase binding. Subsequently the coding region of the gene is transcribed in a mRNA chain by an enzyme called RNA polymerase, which use the DNA codifying strand as template. In the case of eukaryotes system, the mRNA is submitted to the splicing procedure in order to remove noncoding sequence and then is transported in the cytoplasm where it will be translated. Translation step involves a wide variety of factors that often work as multiprotein complexes. Their varied tasks result in a complex but well organized molecular machinery, which translates the information encoded in the mRNA triplets, into a polypeptide chain.

After translation, proteins are submitted to a post-translational modification, depending on protein function and specific cell factory characteristics. Eukaryotic and prokaryotic protein production mechanisms differ from each other. Eukaryotic protein synthesis involves generally more protein components and some steps are more complex. Principal differences concerning the synthesis location, the initiation process of transcriptions, the ribosomal composition, the mRNA processing and post translational modifications can be found elsewhere[12,13] (**Fig. 1**).

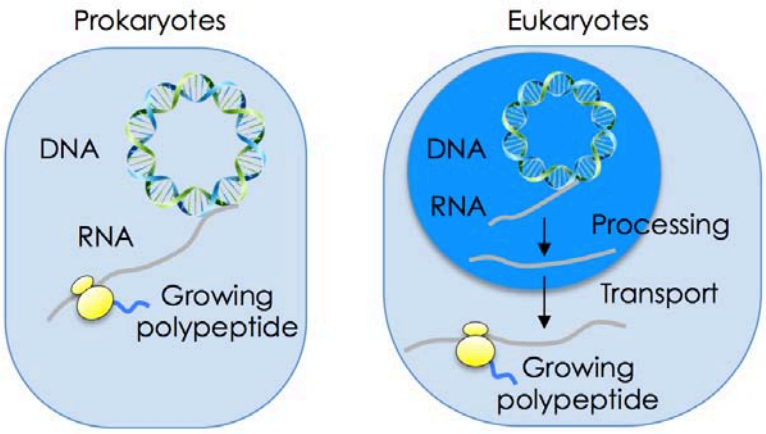


Fig. 1 General comparison of transcription and translation in prokaryotes vs. eukaryotes. Adapted from [13].

1.1 Cell factories

The first step in recombinant protein production is the choice of the cell factory. This decision defines and outlines the whole process, since a specific kind of equipment and molecular reagents are required for each kind of expression system. As mentioned before, recombinant protein features must be considered above all, but also production costs, yield of the system, time, and human efforts are factors to be carefully taken into account. For instance, proteins from eukaryotic organisms can also be expressed in prokaryotic systems[14].

Even though at theoretical level, recombinant protein production seems to be a quite simple and straightforward process, it can present some practical hurdles. Sometimes, recombinant gene expression can affect the growth of the host, lead to the formation of inclusion bodies or protein can be correctly produced but lacking biological activity, making the production process more difficult than expected. Other important aspect to evaluate is the downstream process and cellular contaminants that can be found in protein extracts as prokaryotic lipopolysaccharides (LPS) or genetic material[15,16].

1.1.1 Prokaryotic cell factories

Among all the expression systems, prokaryotes show some advantageous features, which make them the first option for recombinant protein production. They generally permit to obtain high amount of recombinant protein in a short time and the well known mechanisms of transcription and translation make the use of microorganism easier than other cell factories. Another important factor is the wide availability of bacterial mutants, which permit to select the optimal condition of production. Even though specific N-glycoproteins in bacteria have been reported[17], the major drawback

of prokaryotic systems is the inability to carry out complex post-translational modifications.

1.1.1.1 Escherichia coli

The prokaryotic expression system is based primarily on the bacterial species *Escherichia coli*, being the most widely used and best-characterized microorganism[18]. This makes possible the development and commercialization of many biomedical products. In fact, almost 30 % of approved recombinant therapeutic proteins are produced in *E. coli*[19,20]. The main advantage of this system is that cultivation is simple and inexpensive. Moreover, it shows fast growth kinetics and growth at high cell densities allowing high yields of recombinant protein production, in short time[21,22]. In addition, growth media can be prepared from readily available and cheap elements. It has being shown that the recombinant protein can account for up to 30 % of total cellular proteins[23] and the great flexibility of this system had also permitted to successfully transfer a N-linked glycosylation cassette from *Campylobacter jejuny* to *E. coli* cells, resulting in prokaryotic glycosylation of recombinant proteins[24].
Indeed, there are some key-points which should be accurately considered:

HOST STRAIN

E. coli host strain choice is highly important being the genetic background a key feature for protein production. Among the several genetic characteristics available, the capacity to stably maintain the expression plasmid is the most important. In the same way, the expression plasmid encoding the recombinant protein must be selected in order to be compatible with the genetic background of the strain.

BL21 *E. coli* strain, a non-pathogenic B834 derivative microorganism, is one of the most common host for recombinant protein production[25]. The main advantage of this strain is that it's deficient in OmpT and Lon proteases, which may interfere in the purification process, degrading the protein of interest. Several sub-strains are derived from BL21, each of them showing different features but still lacking the main *E. coli* proteases[26]. For example, BL21 trxB/gor negative mutants (Novagen Origami) promote the cytoplasmic disulfide bond formation. Moreover, Novagen Tuner Series BL21 *lacY* mutant strains permit tunable levels of protein production and BL21 RecA$^-$ strain improves the stability of expression plasmids. BL21 Rosetta strain is also designed to enhance the expression of eukaryotic proteins that contain codons rarely used in *E. coli*.

Despite the extended use of BL21 strains, the absence of proteases is not always helpful in the recombinant protein expression process, being these enzymes a relevant component of protein quality control system. Since proteases take part in the folding procedure performed by chaperones, deficient mutations in proteases can lead to an

unbalanced equilibrium between misfolded protein aggregation and solubilization. Insoluble polypeptides generally aggregate in the form of inclusion bodies, lowering the production yield of soluble recombinant protein.

E. coli **K12** and its derivatives are host strains generally used in recombinant therapeutic production in biotechnology. Since the National Institute of Health (NIH) provided guidelines for safety and adopted this strain as a standard, it is generally use in large-scale processes in biotechnology industry. Most common strains used are K12 derivatives Origami (*trxB*/*gor* mutant) and JM 83 for protein secretion to the periplasm[26]. Other *E. coli* host strains have been designed for supply special requirements, namely to enhance solubility of some protein of difficult expression, or to reduce acetate accumulation during cell growth. Most strains mentioned above are commercially available and may carry the lambda DE3 lysogen (IPTG inducible T7 polymerase) and the pLysS (T7 lysozyme) for reduced basal expression level of the gene of interest[19].

VECTOR DESIGN

As mentioned before, expression plasmid vectors should be designed in order to optimize protein production. It's well established that vector copy number and stability affect the protein production yield.

There are different genetic elements that compose an expression plasmid:

The **replication origin** determines the copy number of plasmids with either flexible or rigid control. It's a critical value since a too low or too high copy number can negatively affect protein production. If co-expression from different plasmids is required, compatible groups must be used[27].

The **marker resistance** allows screening for positive transformed cells and the maintenance of expression plasmid selection. Habitual resistance markers target ampicillin, tetracycline, kanamycin or chloramphenicol.

Transcriptional promoters and terminators are genetics regions, which control the transcription activation. They are the main factors responsible for gene expression. There are a lot of promoter systems in which gene expression can be induced either by thermal or chemical factors. The most common chemical inducer is isopropyl-beta-D-thiogalactopyranoside (IPTG) which enhances transcription from *lac, trc and tac* promoters[26]. The transcriptional terminators are placed downstream of the Gene Of Interest (GOI) to avoid unspecific transcription of following coding sequence.

Translation initiation and terminator regions mediate the ribosome binding on mRNA with a Shine-Dalgarno containing sequence, called ribosomal binding site (RBS). A terminator region is composed by a stop codon (UAA in *E. coli*) that controls the translation termination.

CODON USAGE

Codon bias refers to the fact that there are specific synonymous codons; being the 20 amino acids codified by a triplet of nucleotides, there are some amino acids that can be represented by more than one codon. Depending on strains, there are some different preferences in synonymous codon use. The set of codons preferred by each host is called codon usage.

The degeneration of genetic code can cause some problems in recombinant protein expression, since GOI should own the optimal codon usage depending on cell factory chosen for production. If the mRNA owns a rare codon usage, the tRNA corresponding to the rare codon couldn't be available for continuing the protein synthesis. Halting the elongation of polypeptide chain can cause some translational errors or premature termination of peptide synthesis[28,29]. To avoid this problem the best option is the codon optimization of expression plasmid. An alternative is co-transforming the host with an expression plasmid codifying for rare codon tRNA in order to supply the shortcoming. Moreover, there are some *E. coli* host strains, as Rosetta or Rosetta-Gami, which overcome bias in codon usage.

mRNA STABILITY

Generally, *E. coli* mRNAs half-life range between 30 seconds and 2 minutes[30,31], before being degraded by 3'-5' exonucleases enzymes (RNases). Despite the mRNA decay tremendously influences the protein production process, mRNA stability is a feature not always considered. mRNA hairpin structures can protect genetic material from degradation. Polyadenylation at 3' mRNA position also has been suggested to influence mRNA metabolism, favoring degradation. As well as there are host strain for rare codon usage, also exist *E. coli* poly(A)-deficient strains[32,33]. Moreover, RNaseE mutant strain (Invitrogen BL21 star) can help for improving production yield.

1.1.1.2 Other prokaryotic cell factories

Other bacterial hosts can also be exploited for recombinant protein production. Similarities in codon usage and other characteristics with the original organism of recombinant protein help in the good achievement of protein expression.

Theoretically, all bacteria could be used as cell factory. Unfortunately, information about regulation and molecular mechanism of most of them is missing. This lacking of knowledge, joint with a poor availability of commercial expression vectors, limits the variety of prokaryotic hosts in recombinant protein production. ***Streptomyces***[34], ***Methylobacterium***[35], ***Cyanobacterium***[36], ***Staphylococcus***[37] and ***Pseudomonas***[38] are some examples of rare host strains that might increase their relevance in near future.

Nonetheless, the second most popular organism after *E. coli* seems to be ***Bacillus*** system. This gram-positive host is GRAS (Generally Recognized As Safe) and provides a strong secretion pathway if compared with *E. coli*. Another important characteristic is that the outer membrane has no Lipopolysaccharides (LPS), which contains endotoxins that can cause fever in human and other mammals. The most relevant *Bacillus* species used in recombinant protein production are ***B. subtilis, B. licheniformis*** and ***B. brevis***[26].

1.1.2 Eukaryotic cell factories

Eukaryotic expression systems are widely used in both biotechnological industry and research. This family of cell factories includes unicellular organisms and cell lines derived from a variety of species. One of the major advantages of eukaryotic cells is that they are able to perform many post-translational modifications (PTMs) (**Fig. 2**). This fact is extremely important since the PTMs are the key mechanism to increase proteome diversity.

The most relevant protein modifications are summarized in **Annex 3**.

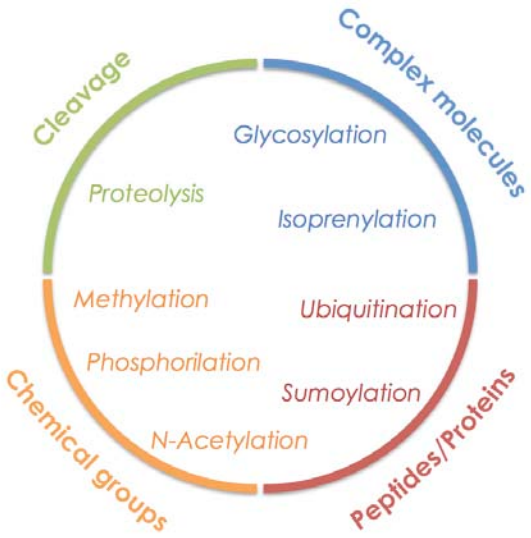


Fig. 2 Scheme of most common post-translational modifications. Adapted from [39].

1.1.2.1 Insect cells-baculovirus technology

Insect cells-baculovirus expression vector system (BEVS) to produce recombinant proteins was first developed in early 1980s[40]. During the last 30 years it has shown to be a versatile platform for producing proteins requiring complex post-translational modifications and high yield. Moreover, advances in DNA technology permit overcome system drawback and turn it a versatile and more robust expression system[41].

BEVS protein production is a two-step process. First, insect cells cultivation at desired concentration, and, secondly, baculovirus (BVs) infection to lead protein production. As other viral infections, BVs take control of the host cell expression machinery, replicate themselves and express the recombinant protein gene, previously cloned into the viral genome[42].

What enables BVs to be suitable for foreign protein productions are two proteins: polyhedrin and p10, which are involved in the horizontal transmission in larvae population. During the late infection cycle virions are coated with polyhedrin. This protective matrix can reach up to the 50 % of total protein amount at the end of BVs infection cycle. At the same time p10 has been also associated with several intracellular structures during BVs infection and it has being found that these two proteins are not required for virus replication in cultured insect cells[43,44]. Therefore, the two genes coding for these proteins, under control of *polh* and *p10* strong promoters, can be replaced by genes of interest, which will be expressed in the very late stage of infection. This permits to obtain a really high productivity with rarely observed formation of inclusion bodies[45].

BACULOVIRUS

Baculovirus (BVs) structure

Baculoviridae (from latin *baculum*, "stick") family of viruses was first used as ecologically friendly biopesticides because of their natural host infectivity[46]. BVs are enveloped rod-shaped viruses with a covalently closed double strand DNA genome, ranging between 80 and 180 Kbp long depending on the species[47]. They infect arthropods, mainly insects, and the host range is generally restricted to a few species. Genome sequencing revealed the presence of at least 895 open reading frames, coding for structural and non-structural proteins. The virus particle size ranges between 30 to 60 nm in diameter with 250 to 300 nm in length, although it can increase to accommodate larger genomes[48]. Two members of the genus *Alphabaculovirus, Autographa californica multiple nucleopolyhedrovirus* (AcMNPV)[49] *and Bombyx mori nucleopolihedrovirus* (BmNPV)[50], both with a 130 Kbp genome length, are widely used for gene expression in insect cells and silkworm larvae, respectively. AcMNPV is the most widely used BVs for gene expression in insect cells. It is able to infect up to 25 different lepidopteran insects and, among all, *Spodoptera frugiperda* derived *Sf9* and *Sf21* cells and *Trichoplusia ni Tn5* cells (High Five™) are the most common used cell lines[51]. Also *Trichoplusia ni single nucleopolyhedrovirus* (TnSNPV) is employed as expression vector, but it's use is not as extended as that of AcMNPV[52] (**Fig. 3**).

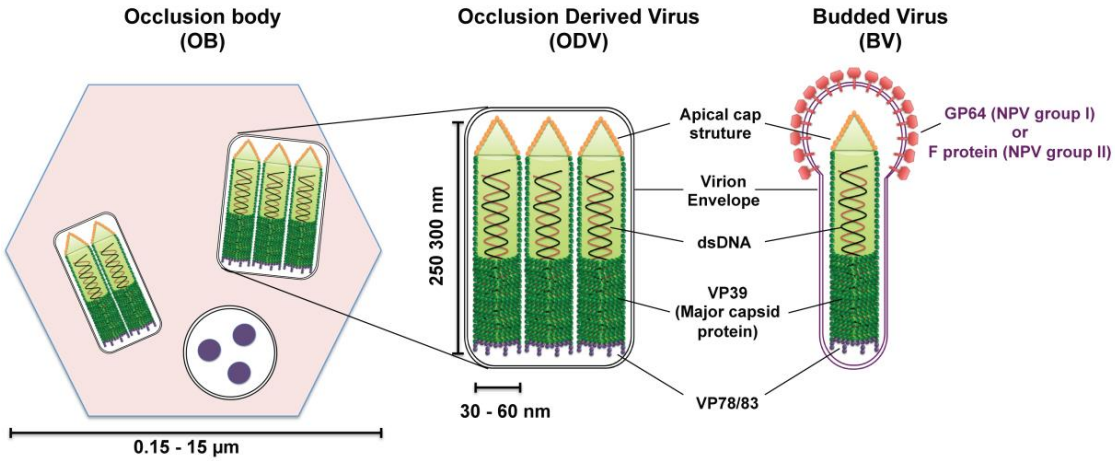


Fig. 3 Structures of baculovirus among their life cycle. ODVs and Bvs are genetically identical differing only in their envelope composition and tissue tropism. Note that ODV are embedded in polyhedrin forming OB. Adapted from [53].

Two different kinds of virions are produced during the biphasic replication of BVs in insect host: the occlusion bodies (OB) and the budded virions (BVs). OBs consist in one or more enveloped nucleocapsids (called occluded derived virus ODVs), embedded in a proteinaceous matrix composed of the very late expressed protein polyhedrin, in case of nucleopolyhedrovirus (NPV). They are adapted to maintain stability outside the host body, and also own an outer coat called *Calyx*[54].

On the contrary, BVs are non-occluded structures and are the responsible of cell-to-cell infection spreading within the host cells. A plasma membrane protein, composed also by major envelope glycoprotein gp64, surrounds them conferring the typical baculovirus profile[55] (**Fig. 4**).

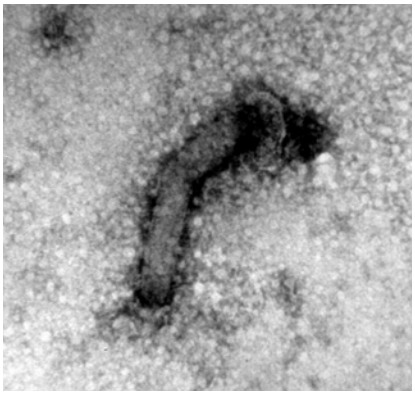


Fig. 4 Transmission Electron Microscopy (TEM) image of AcMNPV. Saccardo P. not published.

Baculovirus infection cycle

Infection cycle starts with the ingestion of OBs by the insect host. Due to the alkaline conditions in insect midguts, OBs are dissolved and polyhedrin is degraded by proteases[56,57]. In this way, free ODVs can infect the epithelial cells by virion-specific binding proteins called *per os infectivity factors* (pif)[58]. Viral entry occurs via non-endocytic pathway. Nucleocapsids are then released in the cytoplasm and, afterwards, reach the nucleus thanks to an actin-based motility activity, which drives the baculovirus transit[59]. Once into the nucleus, baculovirus start reprogramming cells for virus replication. In this way, after about 6 hours post infection (hpi) viruses subvert host cell activity (immediate-early phase)[60]. During the subsequent 12 hours (6-18 hpi, late phase) viral DNA is replicated together with a viral protein expression, essentials for the new nucleocapsid assembly.

During the late-very late phase of infection polyhedra protein is over-produced and it accumulates in the nucleus. At the same time both BVs and ODVs are assembled. Newly formed BVs, then, infect new cells in the secondary infection cycle. They reach new cells through hemolymph and secondary infection occurs by GP64 clathrin-mediated endocytosis[61]. Once enter to a new cell, during the late-very late phase of infection, polyhedra protein is over-produced and it accumulates in the nucleus. The dynamic is similar to primary infection events except for the fact that secondary cycle provokes an extensive infection of insect host, leading to cell lysis and subsequently insect larvae death. This fact permits to the ODVs to spread into the environment and start the infection again. As mentioned before, due to the protection of polyhedrin, these occlusion bodies can be stable outside the insect, until being ingested by other larvae[42,62] (**Fig. 5**).

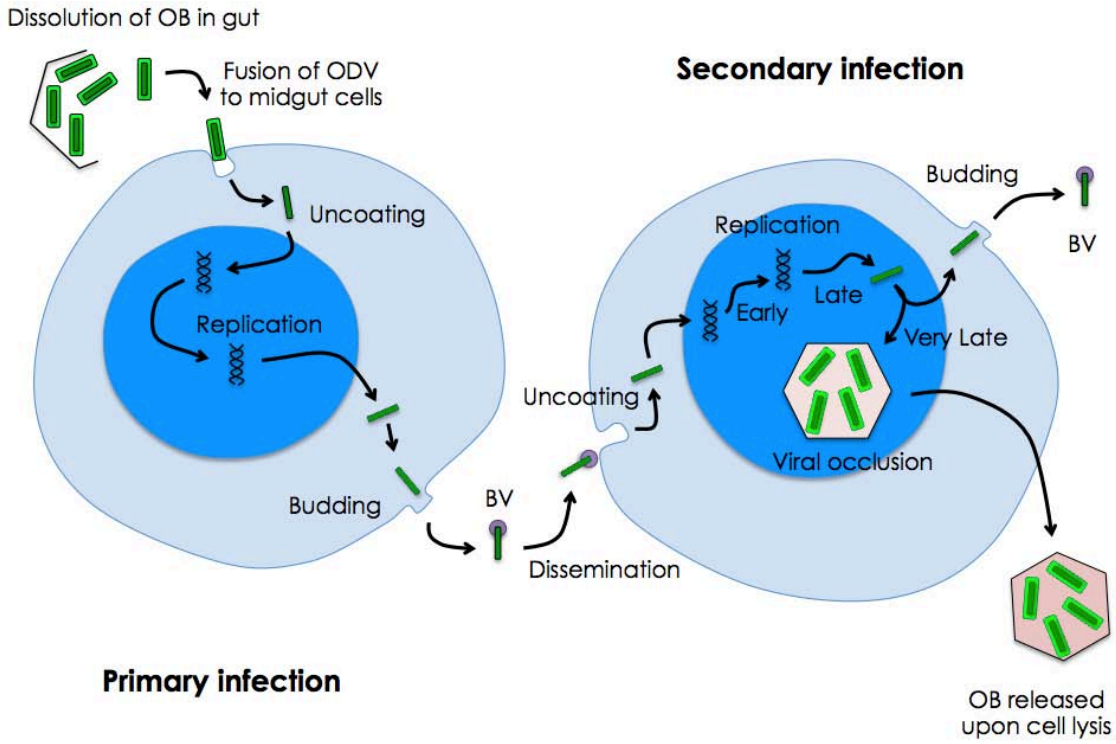


Fig. 5 Simplified representation of baculovirus infection cycle. Adapted from [63].

BACULOVIRUS AS A PLATFORM FOR PROTEIN PRODUCTION

As mentioned before, the key characteristic of BVs is the fact that polyhedrin protein is not essential for virus replication in cell culture, as well as P10 protein and all proteins involved in horizontal transmission of infection. Therefore, polyhedrin and p10 genes can be replaced with DNA sequence coding for protein of interest. Transcription is then controlled by the very late strong promoters *polh* and *P10*, ensuring high yield of recombinant protein production[64].

Diverse kits are commercially available for baculovirus expression vector system (BEVS) engineering. Differences between them are given by the methods of recombination used to obtain recombinant genomes:

Homologous recombination *in vivo*

The homologous recombination *in vivo* requires a "Transfer plasmid", in which the heterologous gene is flanked by homologous sequences of baculoviral polyhedrin locus. Cotransfection of cultured cells with transfer plasmid plus purified AcMNPV DNA results in a double crossover recombination of heterologous gene of interest (GOI) DNA into baculovirus genome and subsequent formation of engineered BVs. The weak point of this method is the higher rate of single recombination events in comparison to double insertion and subsequent clone selection, being only 0.1 % the recombinant BVs obtained in each cotrasfection[43,65] (**Fig. 6 A**).

Modification of the protocol has permitted to increase the BVs vector recovery up to 10-20 %, by just linearizing baculoviral genome with introduced Bsu36I unique restriction site in the polyhedrin locus. In this way, linearized Baculovirus vector is unable to replicate unless recombination with transfer plasmid occur[66] (**Fig. 6 B**).

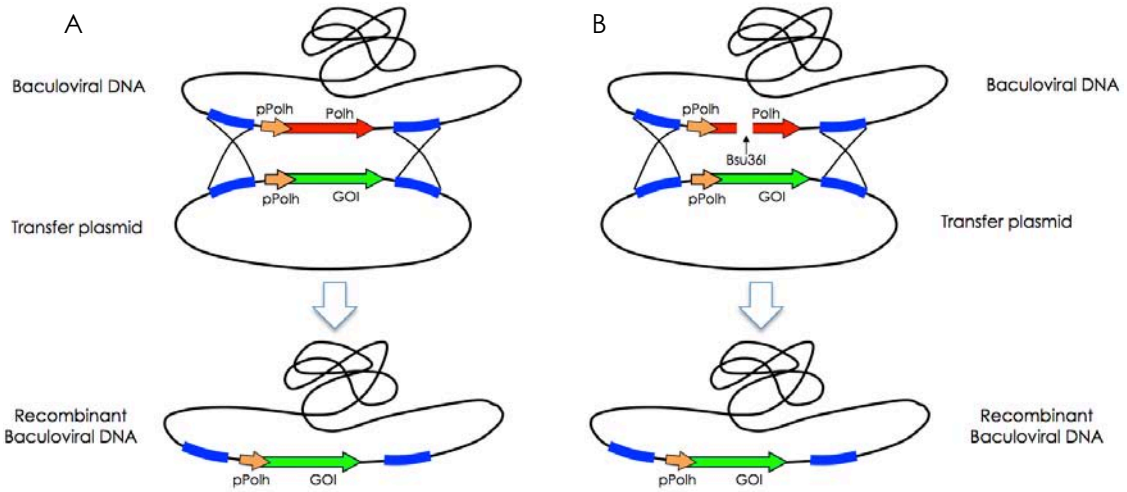


Fig. 6 Homologous recombination in vivo: Baculovirus expression vector obtained using A: simple homologous recombination (note that initial baculoviral DNA is able to produce viable empty baculovirus and that unstable Baculovirus genomes are formed by single recombination events in which bacterial DNA are inserted into the viral genome), B: linearized baculoviral DNA recombination (single recombinant events are not able to render viable recombinant baculovirus DNA). Adapted from [66].

Another modification of baculoviral genome had permitted to insert two Bsu36I restriction sites, in orf603 and orf1692 recombinant region. These regions flanking a reporter *lacZ* gene cloned under the *polh* promoter control. As mentioned before, only recombination with transfer plasmid allows the recovery of viable baculoviral genome. Moreover, if undigested or single-digested, repaired parental DNA produces progeny and it can be detected by colorimetric assay[66] (**Fig. 7 A**).

BacPAK™ (Clontech) commercial kit had permitted to obtain up to 95 % of recombinant baculoviruses thanks to a deletion in the *orf1629* recombinant region, which encodes an essential phosphoprotein of the nucleocapsid. A third *Bsu*36I restriction site was added in *lacZ* gene sequence. In these conditions, only recombination of GOI permit the Bacmid viability recovery since phosphoprotein sequence is restored by recombinant region in transfer plasmid[67-69] (**Fig. 7 B**).

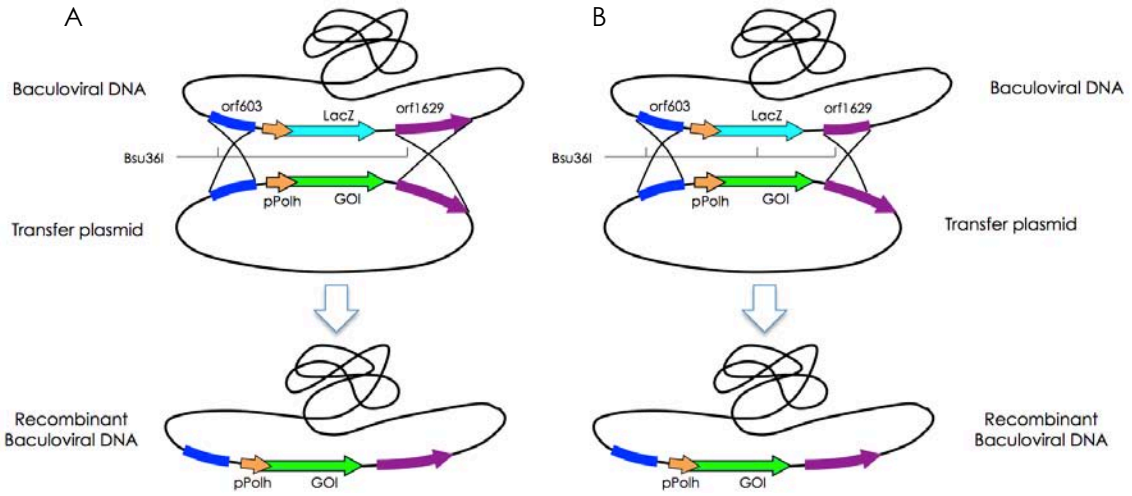


Fig. 7 Homologous recombination in vivo: Baculovirus expression vector obtained using A: orf603-orf1629 linearized genome B: orf603-orf1629-LacZ triple digestion (BacPAK™). Adapted from [66].

Homologous recombination with Bacmid technology

The homologous recombination *in vitro*, also known as Bacmid technology, has being the major step forward in the baculoviral engineering field for enhancing recombination efficiency. With this method, baculovirus genomes are capable to replicate in *E. coli* as a bacterial artificial chromosome because of the inserted *E. coli* fertility factor (F-factor) replicon (miniF). These Bacmidial (BAC) vectors can accept large inserts up to 300 kb with the help of a plasmid (helper plasmid) encoding for a transposase. Once selected the positive recombinant colony, Bacmid genome is extracted and transfected in Insect cells culture, leading to the formation of recombinant baculoviruses[49].

Bacmidic DNA can be maintained in *E. coli* cells as defective viral genome (orf1629 recombination site deleted portion). As mentioned above, only restauration of orf1629 sequence by recombination with the transfer plasmid can restore viable baculoviral genomes, allowing high percentage of recombinant baculoviral genome recovery (**Fig. 8 A**).

Commercial Bac-to-Bac™ (Invitrogene) kit provides a 100% of efficiency in recombinant baculovirus recovery. This kit exploits "mini-Att Tn7" sites, for the transposition of the GOI. This strategy requires a helper plasmid encoding for transposase: this enzyme is involved in the transposition of GOI from Transfer plasmid to the polyhedrin locus of bacmid, knocking out the *lacZ* sequence and allowing an easy blue/white screening of *E. coli* colonies (**Fig. 8 B**).

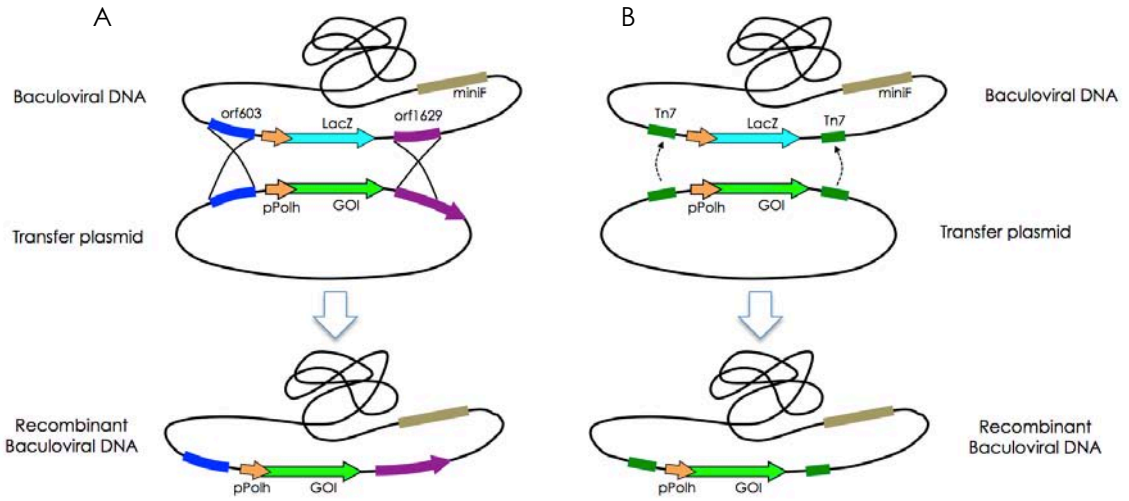


Fig. 8 Homologous recombination using the bacmid technology: Baculovirus expression vector obtained using A: partially deleted orf1629 sequence, B: mini-Att Tn7 recombination sites (Bac-to-Bac™). Adapted from [66].

Commercial kits as FlashBAC™ (Oxford Expression Technologies) and BaculoGold™ (BD Bioscience) expression not only restore the *orf1629* sequence but also knock out the bacterial replicone, generating baculoviral genomes with negligible background. This strategy cross-hybridise the linearisable baculoviral DNA and bacmid approaches[65]. The defective baculoviral genome maintained in *E. coli* cells contains the bacterial replicon in the polyhedrin locus and a deletion in the *orf1629* gene. The bacmid DNA and the transfer plasmid are used to cotransfect insect cells. Thus, homologous recombination allows the *orf1629* sequence recovery, the insertion of GOI in the polyhedrin locus and knocking out the bacterial replicon at the same time (**Fig. 9**).

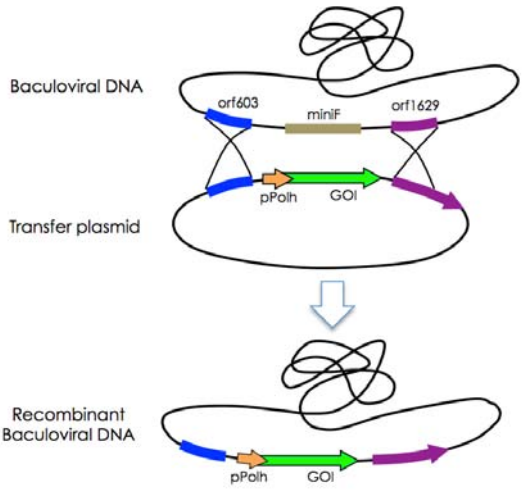


Fig. 9 Homologous recombination using the bacmid technology: Baculovirus expression vector obtained using FlashBAC™ and BaculoGold™ method, which permits knocking out the bacterial replicon. Adapted from [66].

Homologous recombination *in vitro*

BaculoDirect™ kit strategy takes advantage from the different approaches described above. GOI is cloned in transfer plasmid, flanked by recombination site attL1 and attL2. Baculoviral genome contains a reporter *lacZ* gene and the herpes simplex virus thymidine kinase gene flanked by site-specific recombination sites *attR1* and *attR2*, replacing the polyhedrin coding sequence. The pre linearized viral genome and the transfer plasmid are then mixed *in vitro* with a purified recombinase. The mix with recombinant baculoviruses is then transfected into insect cells. Recombinant baculoviruses can be recovered using gancyclovir as selective agent against replication of parental viral DNA and colorimetric screening assay[70] (**Fig. 10**).

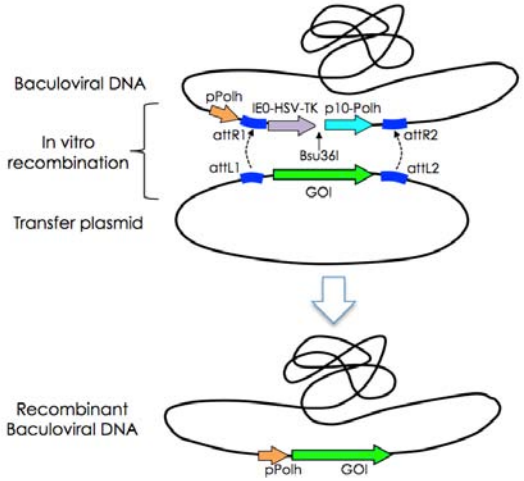


Fig. 10 Homologous recombination using homologous recombination in vitro: Baculovirus expression vector obtained using BaculoDirect™method, which permits viral recombinant selection by colorimetric assay and Gancyclovir. Adapted from [66]..

INSECT CELLS FOR PROTEIN PRODUCTION

Insect host also play an important role in protein production with BEVS. About 400 cell lines, derived from more than 100 insect species have being used for an extensive range of protein expression[71,72]. Insect cells morphology is generally spherical, with a diameter ranging from 10 to 20 µm[73]. Traditionally, the insect cell lines most commonly used have been lepidopteran. Grace reported the first description of established lepidopteran cell line in 1962[74]. Among the three most frequently used cell lines susceptible to AcMNPV vectors, *Sf9* and *Sf21* are derived of IPLB-SF-21 pupal ovarian tissue of the fall armyworm *Spodoptera frugiperda*[75], meanwhile the BTI-Tn-5B1-4 cell line, commercially known as High Five™, derived from the ovarian tissue of cabbage looper *Trichoplusia ni*[76].

Generally, insect cells grow at 27°C both in adherence or in suspension culture and can be adapted to serum free media[77,78]. No CO_2 is required for cell growth, and, since they

are loosely adherent, Trypsin and EDTA are not required for cells subculture. Other advantage of these cell lines is that they can be easily scaled up achieving high densities. Production in large scale bioreactors allow then to obtain high recombinant protein production yields[79,80].

Starting from 1998, engineered subcloned Sf9 cells are available[81]. Most relevant modifications regarding the glycosylation pathway involving the post-translational process of newly synthesized proteins in order to obtain a glycosylation pattern as similar as possible to human (MIMIC™ cells) have been described[82]. Other modifications regarding the prolongation of cell viability in post infection period (SuperSf9 OET) are also available[83].

As alternative to cell lines, insect larvae cultivation can be exploited. *Spodoptera frugiperda* and *Trichoplusia ni* larvae cultivation is cost effective and easily to scale up. On the other side, recombinant protein purification is disturbed by insect parts as contaminants products and protein aggregation has also being observed. Nonetheless, *Bombix mori* larvae are commonly used as hosts due to growth difficulties in large cell line culture volumes[54].

One of the most appreciated features of insect cells for recombinant protein production is that baculovirus system allows expressing simultaneously up to 4 coding region under strong p10 or polh promoters. Moreover, multiple expression can be achieved co-transfecting cells with two or more engineered baculoviruses[84].

1.1.2.2 Other eukaryotic cell factories

Eukaryotic organisms offer a wide spectrum of cell factories for recombinant protein production. **Yeasts** are particularly exploited as cell factory because they combine the versatility of prokaryotic *E. coli* cells to grow in high density and in low cost media with the post translational modification pathways and the ability to secrete recombinant products, helping then in the downstream purification process. Most used yeasts strains are *S. cerevisiae* and *P. pastoris*, among the wide catalogue of strains available[85].

Mammalian cells are also suitable for protein production, especially of those requiring specific post-translational modifications for being active. Chinese Hamster Ovary (CHO) cells mouse myeloma (NS0), Baby Hamster Kidney (BHK), Human Embryonic Kidney (HEK-293) and human retina-derived (PERC6) cells, are the most used cells lines[86].

Filamentous fungi have being also exploited for many years for production of fungal enzymes and some other low molecular weight products. Even it's not a widely diffuse expression system, some species as *Trichoderma reesei* can reach extremely high yield of product[87,88].

Microalgae are characterized by high productivity, fast generation of transgenic organism and cost effective cultivation. They are recognized as GRAS organisms since

there is no gene flow by pollen in the environment. In this way transgenic microalgae can be easily handled even at high volume of cultivation[89]. *Dunaliella, Chlamydomonas, Haematococcus, Chlorella* and *Volvox* are the most common used microalgae[90,91].

2 Protein folding

For almost 80 years, protein folding has being considered an important cross-disciplinary field and an essential issue for understanding the biological activity of living cells. The awareness of how newly synthesized proteins adopt their three dimensional structure complete the comprehension of the last step of the central dogma DNA → RNA → Polypeptide chain → Functional conformation. The first major advance in this field occurred in 1951, when Linus Pauling first discovered and described the α-helix and the β-sheet conformations, found in almost all proteins[92]. Eleven years later, in 1962, Max Perutz and John Kendrew were awarded with a Nobel prize for their pioneering work in determining the structure of globular proteins[93] and, in 1973, Anfinsen postulated that, in a given environment, the protein folding to the native 3D conformation is determined by the 1D amino acidic string sequence[94]. Nonetheless it was still unclear how and why does polypeptide sequence fold onto a unique 3D structure.

Since the native protein conformation usually have the lowest Gibbs free energy and is thermodynamically stable, the simply assumption that protein folding is guided by thermodynamic driving forces does not explain the question posed by Cyrus Levinthal, in 1968. He raised the issue, known as Levithal's paradox, of how proteins reach the native conformation in an extremely short time (seconds or fraction of second) despite the enormous number of spatial conformations that the polypeptide can assume[95].

Levinthal and Wetlaufer suggested a kinetic model in which folding is a cooperative process guided by the simultaneous formation of structure nuclei along the protein, restricting the general conformational freedom and, subsequently, the number of possible conformations that a given polypeptide can assume, thus, driving the polypeptide to the native conformation[95,96]. That initial stage of folding was described by Dill as a hydrophobic collapse, lead by the repulsion between hydrophobic region of amino acidic chain and water, followed by a rearrangement in protein conformation. These nuclei could then cooperatively associate, forming micro domains and molten globule structures, in which the hydrophilic residues are exposed on the protein surface area and secondary structures are formed.

These successions of events finally drive the protein folding process to the native conformation in an extremely short lapse of time[97].

The solution to Levinthal's paradox and the current view of the protein folding is based on studies with foldable polymers models[98,99]. The studies of the chains' entropies and the concept of the statistical nature of folding, lead Wolynes and coworkers to illustrate the protein folding energy landscape as funnel-shaped profile[100] (**Fig. 11**). This model describes both kinetic and thermodynamic behavior of the process; moreover it is very fast and requires cooperativity between the folding nuclei.

Harrison and Durbin also purposed the existence of multiple folding pathways with a sequential formation of intermediates. The same protein may follow completely different routes, and some paths can be followed by a larger population than others. Towards the bottom of the funnel the number of possible conformations is limited and the folding process speeds up to the same protein native conformation. If subunit association is required for the biological activity, association of oligomeric proteins occurs as a last step, inducing conformational readjustments which confer the biological functional properties[101].

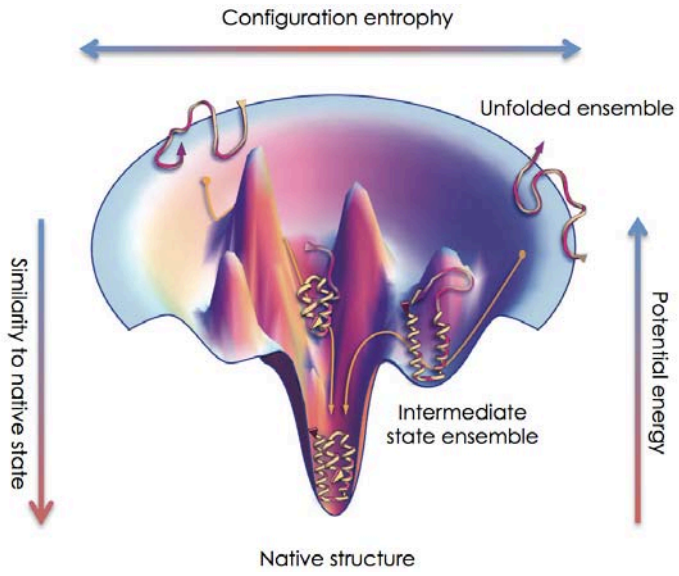


Fig. 11 Scheme of the folding funnel according to Wolynes et al[100]. The depth of the funnel represents the energy (minimum at native structure conformation, red color) and the width represents the configuration entropy (maximum at non-native structure conformation, blue color). Yellow arrows represent different folding routes. Partially folded species decreases as the protein glide down to the minimum of the energy landscape. Adapted from [102].

Formation of intermediates occurs when the cooperativity is not so high, and their distribution does not depend on protein concentration[103]. Then, detection and characterization of intermediates can be impeded by high cooperativity and subsequently by the rapidity of the process which difficult the analysis. For these reasons, the structural characteristics still remain matter of discussion.

In case of transcription or translation error, the polypeptide sequence can be altered by a mutation or deletion of one or more amino acids. In this case, some of the intermediates may be retained in a stable minimum enclave, generating an unbalanced ratio between the correct folded and misfolded aggregated protein. In this singular situation of persistent non-native interactions, the protein biological function

can be affected, and aggregation is favored by the exposition of hydrophobic residues, otherwise buried in the native conformation[104,105].

In humans, altered proteins' biological function can provoke many different pathologies including cancer[106] and cystic fibrosis[107] among others. Moreover, protein aggregation generally renders deposits in brain, spleen and heart, causing disorders as Parkinson's and Alzheimer disease, type II diabetes and spongiform encephalopathy[108-111].

Protein folding in living cells and cell factories also presents a more complex scenario. Ribosomes continuously synthesize and release proteins in matter of seconds, establishing a really high concentration of macromolecules in cytoplasm[112]. This situation can cause a non-desired interaction between unfolded proteins or intermediates generating aggregation and precipitation of insoluble proteins. For this reason, there are some folding modulators called chaperones, which assist nascent, stress-destabilized or translocated proteins in the folding process[113] (**Fig. 12**).

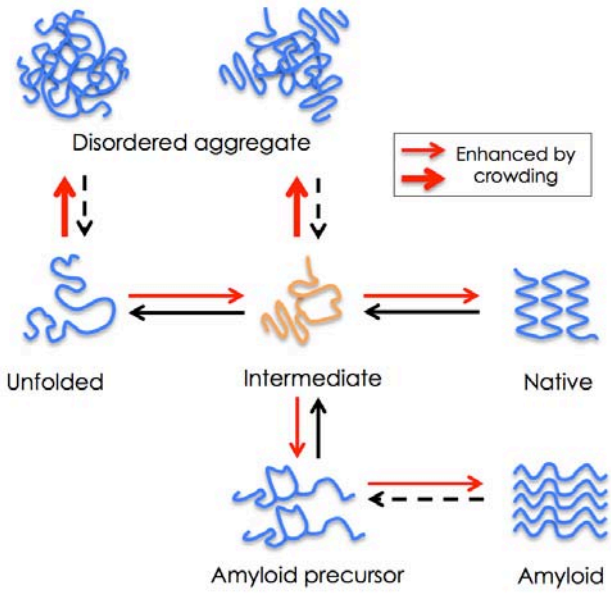


Fig. 12 Scheme of effects in non-native proteins' aggregation by the crowded environment of the cells. Red arrows show the reactions more (bold arrow) or less (normal arrow) influenced by macromolecules concentration in cytoplasm. Adapted from [114,115].

2.1 Quality control

Chaperones assist the *in vivo* folding of proteins in accord to the Anfinsen's postulate: nor adding conformational information to the folding process, neither interacting covalently with nascent proteins. They not only limit the accumulation of misfolded proteins but also help in protein removal from aggregates[116]. Chaperones don't work alone; these molecular assistants take part of an interwoven network called the protein quality control system, which is composed also of proteases. If a protein is dramatically

damaged and becomes unable to fold even with the help of the chaperones, it will be degraded by proteases, avoiding formation and accumulation of misfolded structures. Chaperones and proteases then act together in coordination to ensure a correct balance of protein folding and degradation, avoiding dangerous situations for the cell as protein misfolding, inactivity or aggregation[117] (**Fig. 13**).

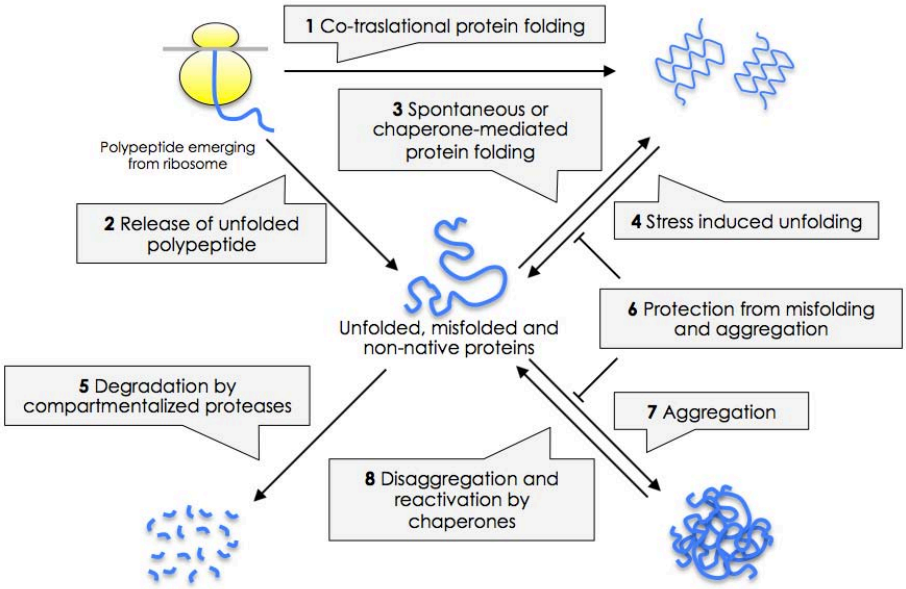


Fig. 13 Schematic representation of protein control quality network. Chaperones support co-translational protein folding (1), folding of polypeptides completely released from ribosomes (2) and remodeling non-native proteins (3) unfolded by cellular stress conditions (4-6). If correct fold is impossible, proteins can be degraded by chaperone associated proteases (5). If the production system is affected by cellular stress, non-native proteins can aggregate (7) in a reversible process (8). Adapted from [117].

2.2 Chaperones

Molecular chaperones are, then, a functionally related set of proteins, first described in 1978[118], which have a primary role in the quality control system. All these proteins can be classified into more than 20 families, according to their molecular weight. Chaperones can be found in almost all organisms and cells may be able to express several different versions of the same chaperone. Sequence homology and related structural functionality can be found between chaperones of the same family[119] (**Fig. 14**).

Despite protein synthesis is the principal source of unfolded polypeptide chains, there are other processes that can lead to unfolded protein as well, such as non-physiological high temperature. Since chaperone sets were found to be overexpressed to overcome the cellular stress due to high temperature, they are also called heat shock proteins (Hsp), even though not all Hsp are chaperones and vice versa[120].

Two distinct mechanisms govern the *de novo* folding in cytosolic cellular compartment: Hsp70 and trigger factor hold newly and nascent polypeptide chains until the complete synthesis is finished, meanwhile other chaperones complexes form a defined compartment in which the unfolded protein can be harbored and the hydrophobic residues are buried in the protein core, leading to the folding process[121]. These two macro classes of proteins cooperate in a topologically and timely ordered manner[114,122,123].

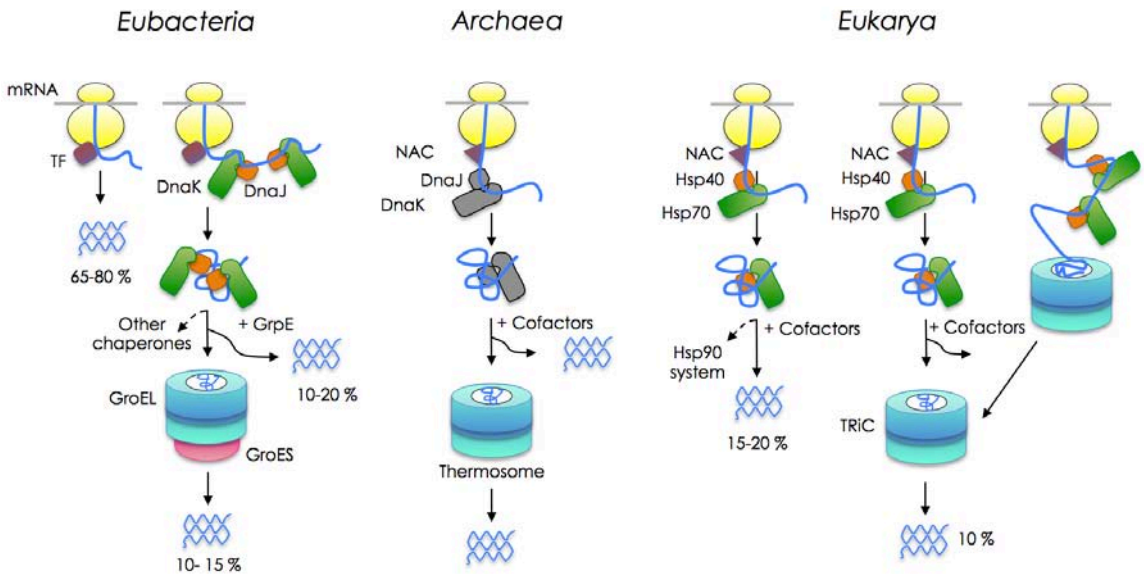


Fig. 14 Schematic and simplified representation of chaperone-assisted folding in Eubacteria, Archaea and Eukarya cytosols. Percentages represent the total fraction of protein reaching the native state for each pathway. Adapted from [115].

In biotechnology, recombinant protein expression is often impaired by protein aggregation in Inclusion Bodies. In some cases this problem can be solved by the combined overexpression of chaperones. Thus, chaperones are recognized as useful tools for tuning protein production and then optimize biotechnological processes. Being protein quality control system involved in an extended and populous net of interactions, in the following sections only the most relevant chaperone proteins, taking part in the cytosolic quality control system, are described.

2.2.1 Trigger factor

Trigger factor (TF) is a 48 kDa eubacterial protein which binds the large subunit of ribosomes, near the nascent polypeptides exit site. It interacts with polypeptides as short as 57 residues are synthesized[124] and the target recognition is mediated by short

hydrophobic sequences of aromatic amino acid residues. After releasing from ribosome, protein-TF complex dissociates in ATP independent manner[124,125].

Peptidyl-prolyl cis/trans isomerase activity (PPIase) has being observed *in vivo* even though proline is non-essential for TF activity in protein quality control system[126]. TF then stabilizes nascent polypeptides maintaining them in a "flexible state" thanks to the cis/trans isomerase activity in cooperation with Hsp70 group of chaperones. At 37°C, *E. coli* cells lacking TF do not apparently show folding defects. Moreover, TF enhances viability of cells at low temperatures[127].

The eukaryotic cytosol lacks TF but shows a ribosomal heterodymeric complex called nascent chain-associated complex (NAC), which lacks of PPIase activity[115].

2.2.2 Hsp70

Proteins of the Hsp70 family, DnaK in *Eubacteria* and *Archaea*, are expressed in the cytosol of eubacteria, eukarya and also in mitochondria and endoplasmatic reticulum of eukaryotic organisms. In higher Eukaryotes, both the constitutively expressed Hsp70 homologs Hsc70, and the stress inducible form Hsp70, are produced and act together with the Hsp40 (DnaJ) family.

Hsp70 proteins share common and conserved functional domains: the 44 KDa N-terminal ATPase domain, and 27 kDa C-terminal ATPase activation domain with substrate-binding domain and other chaperones binding function[128,129]. The Eubacterial Hsp70 protein DnaK, its cochaperone Hsp40 DnaJ and the nucleotide exchange factor GrpE characteristics and cooperation will be discussed in this section (**Fig. 15**).

DnaK targets are small hydrophobic regions distributed along the polypeptide nascent chain, statistically every 40 residues approximately. DnaK chaperone can recognize this region with an affinity ranging from 5 nM to 5 M[128].

Chaperone-Polypeptide binding occurs when DnaK is in ATP-bound state (or open state), and the stable holding of unfolded protein (close state) involves ATP hydrolysis in ADP. The switch between the two states is regulated by chaperone DnaJ.

DnaJ is a 41 KDa Hsp40 protein composed by the N-terminal domain, which binds DnaK and promotes the ATP hydrolysis, and the C-terminal domain which can recognize hydrophobic residues and then targeted them to DnaK[115,130,131].

GrpE completes the reaction cycle, inducing the release of ADP by DnaK upon ATP rebinding, allowing the formation of the DnaK open state, favorable to unfolded protein recognition[132]. Interestingly, not all Hsp70 proteins are GrpE dependent since the rate-limiting step in the DnaK ATP/ADP cycle is normally the ATP hydrolysis[115].

DnaK-protein closed state time may vary between 1 minute in case of rapid folding after complete synthesis, to more than 10 minutes, depending on polypeptide length and folding complexity. Moreover, since DnaK generally associates with 20-30 KDa

nascent polypeptides, it's assumed that its action is subsequent to the Trigger factor one[133]. Moreover, it has been observed that deletion of DnaK in TF-deleted cells causes the aggregation of newly synthesized large proteins[134].

Hsp70 proteins can also bind unfolded proteins under stress conditions or slow-folding intermediates that still expose hydrophobic residues, preventing aggregation or intramolecular misfolding. The released peptide can also be folded by GroELS chaperones which act downstream in the protein folding system[134].

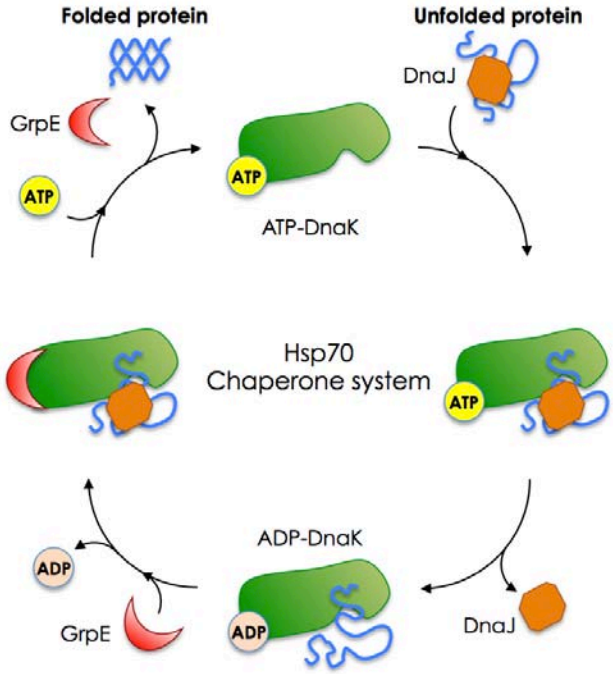


Fig. 15 Schematic and simplified representation of Hsp70 chaperone system. Adapted from [135].

2.2.3 Hsp60:

The Hsp60 family, GroEL in Eubacteria, constitutes a conserved class of an ATP-dependent folding system, which is formed by a 800 KDa double ring/barrel shaped structure composed by back-to-back rings of identical symmetric subunits of 60 kDa each. Each ring can switch between two major states: a binding-active or open state, which is characterized by the exposure of hydrophobic residue, and the folding-active or close state, in which the protein substrate is buried inside the ring in a folding favourable environment (**Fig. 16**).

The folding activity, again, is carried out through an ATP hydrolytic dependent cycle. Newly synthetized polypeptide chain binds the GroEL binding active ring in presence of ATP. In this way the substrate is protected from undesired reactions with other cellular components. Switch from binding-active to folding-active state is driven by the collaboration of a Hsp10 co-chaperonine, **GroES**, which acts as a lid-like barrel and

triggers the hydrolytic catalysis of ATP to ADP. This causes a conformational change inside the barrel, favouring the native folding state of the substrate. Simultaneous binding to the back ring by ATP/co-chaperonine/protein substrate induces the dissociation of the opposite complex, composed by newly folded protein/ADP/co-chaperonine cap. In this manner, rings' activity acts as an ATP dependent cyclical activity (**Fig. 16**).

Because of the limited size of the structure, the common substrates of Hsp60 complex are small proteins between 25 and 60 KDa[115,128].

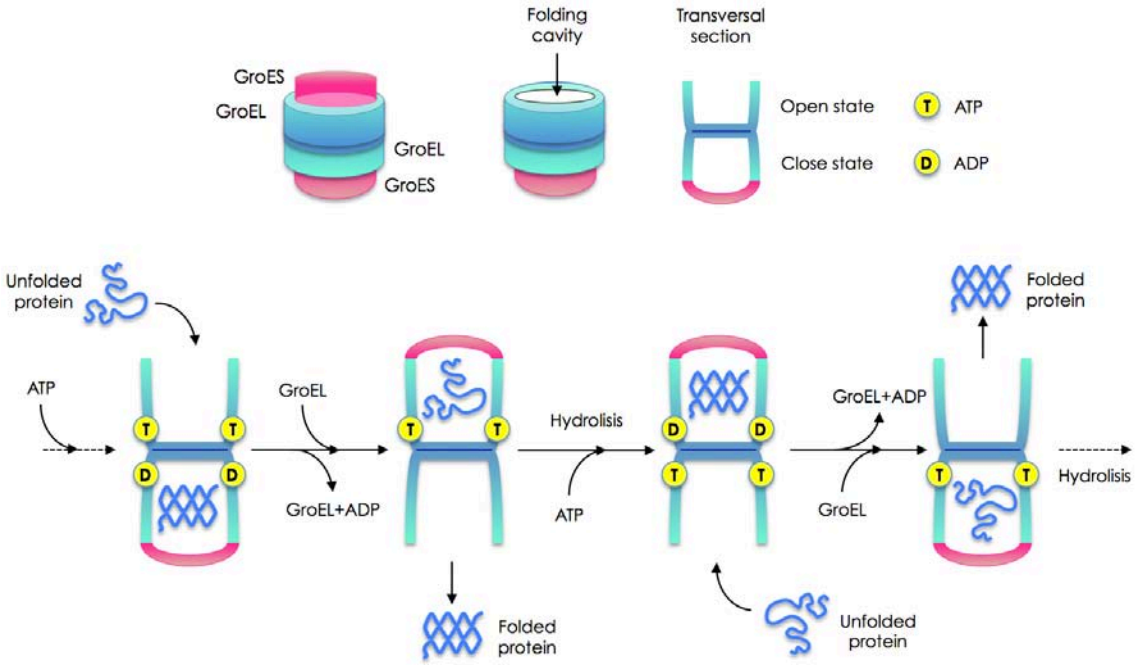


Fig. 16 Schematic and simplified representation of Hsp70 chaperone system. Adapted from [128].

2.2.4 Hsp 90

Hsp90 is a highly abundant and conserved chaperone involved in many pathways of protein folding[136-139]. Despite its abundance in the prokaryote cytosol, its mechanism of action is poorly understood if compared with other chaperone families. Like Hsp70 and Hsp60, Hsp90 is an ATP dependent chaperone that can bind non-folded proteins to prevent aggregation, and its action is developed at the late stages of substrate folding. The bacterial form of this protein is not indispensable for viability while the eukaryotic form is. This class of chaperone is the responsible for the maturation of proteins involved in signaling in cell division like steroid hormones receptors, kinases and the p53 tumor suppressor protein among others (**Fig. 17**).

Hsp90 is formed by a homodimer of elongated subunits, composed by three domains: N-terminal nucleotide binding domain, a middle domain and a dimerization site at C-terminal that changes its conformation responding to ATP-ADP hydrolysis stimuli.

The same 3 domains architecture is conserved among all homologs[140,141].

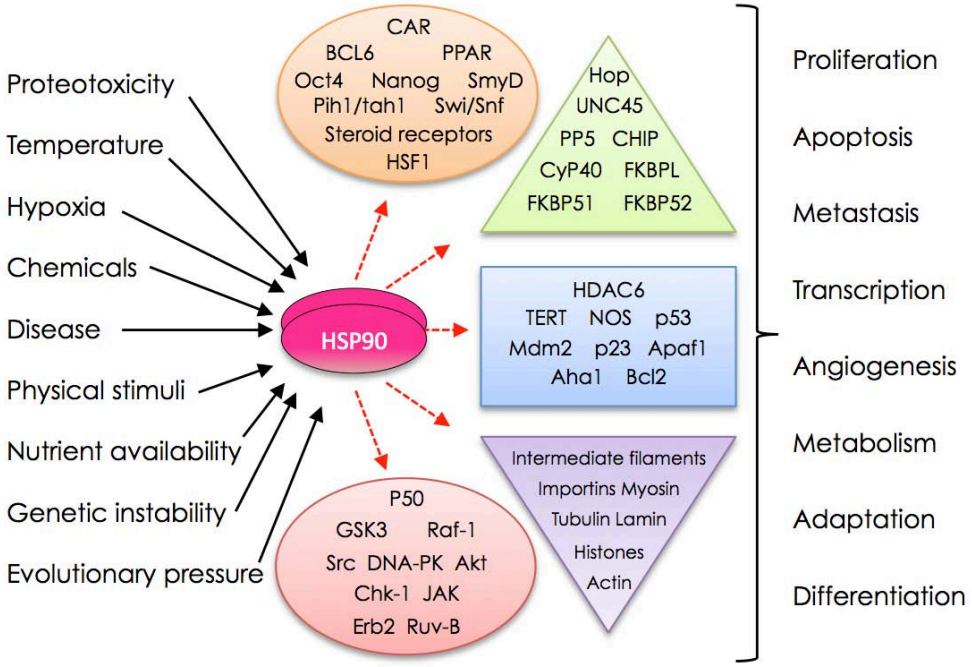


Fig. 17 Schematic representation of Hsp90 with client proteins. Hsp90 modulate hundreds of factors in order to regulate homeostasis against several factors (left list). These factors can be transcriptional factors (orange), TRP-domain proteins (green), structural proteins (purple), protein-kinases (red) or miscellaneous (blue). Depending on the pathway, cell can respond in different ways, reported in the right list. Adapted from [140].

2.3 Proteases

Another class of proteins, called proteases, degrades aberrant polypeptides, minimizing the accumulation of misfolded proteins in cytosol and, by the other hand, recycles amino acids. Proteases can process kinetically trapped folded intermediates and partially folded or truncated proteins, through degradation pathways involving ATP-dependent and ATP-independent proteases[142].

In eukaryotes, the most representative ATP-dependent protease is the 20S proteasome meanwhile, in bacteria, several proteases members of Lon and caseinolytic protease Clp family perform the proteolytic task. Lon and Clp are key ATP-dependent proteases involved in the protein quality control system of *E. coli*. Unlike Clp, Lon orthologs are found in many eubacteria, archaea, and eukaryotes[143,144] (**Fig. 18**).

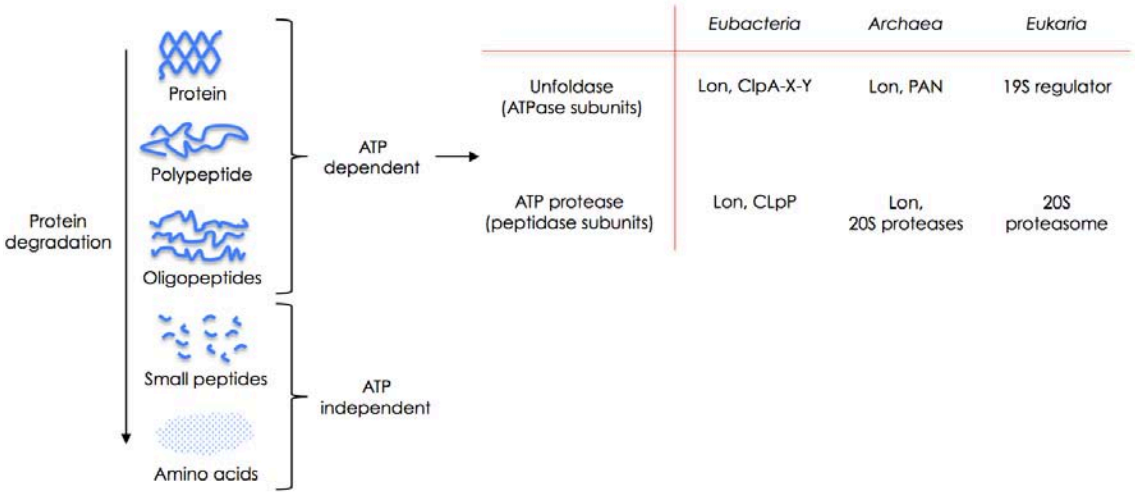


Fig. 18 General scheme of proteases involved in protein degradation. The degradation pathway (from top to bottom arrow) leads to the disruption of proteins to amino acids both in ATP-dependent and ATP-independent steps. ATPase subunits and peptidase subunits for different kingdoms are reported in the table. Adapted from [144].

2.3.1 Lon

Lon protease is structured as an oligomer of identical subunits of approximately 87 Kda forming a ring shaped hexamer[145]. ATPase and proteolytic active sites are both positioned in the same polypeptide chain[146], as well as the sensor and substrate discrimination domain[147]. Lon, first described in *E. coli*[148], is an AAA protein, which means that it is implicated in various cellular activities. Although the recognizing mechanism is not well understood, this protease seems to preferentially recognize some key substrate domains like the C-terminal histidine of cell division inhibitor SulA, or the N terminal domain of transcription factors SoxS and MarA. On the other hand, substrate discrimination seems to be mediated by the exposure in protein surface of hydrophobic residues or structural motifs usually buried in the protein core[149].

2.3.2 Clp

Clp is an ATP-dependent chaperone-protease complex organized in a "stack-of-rings" architecture rendering a toroidal particle. This structure is composed by two functional elements: a tetramer with cylinder-like shape acting as proteolytic core nucleus and the hexamer ATPase-active chaperone ring. The chaperone ring is the responsible for the recognition and the threading of unfolded protein into the proteolytic core, in which multiples actives sites of serine or threonine hydrolyze proteins into 5-10 amino acid peptides[150]. The two-component architecture makes possible a series of different protease complexes, depending on the interaction of the CLpP protease core with different chaperones rings as ClpA, ClpC, ClpE, ClpX, ClpY. Moreover, Clp chaperones

can be divided in two classes: class I composed by ClpA, ClpC and ClpE and containing two consecutives ATP protolytic sites, and class II, composed by ClpX and CplY containing only one ATP protolytic site. Recognition of substrate is non-specific and is generally due to the poor content of tertiary structure in non-completely folded proteins[151].

3 Gene Therapy

The use of biotechnology for producing recombinant molecules for therapeutic or diagnostic applications has emerged in the past decades as a really promising tool. Protein and nucleic acid bioengineering permitted to generate a wide range of innovative pharmaceutical compounds that permit not only to fight disease but also improve quality of life of patients. Moreover, the understanding of the pathogenesis of diseases had increased, thanks also to the investigation on the genetic function/regulation of genes involved in pathological processes. The information regarding the genetic disease mechanisms had allowed to develop novel nucleic acid drugs, which can act before the protein synthesis, substituting the gene involved, or silencing/regulating original gene expression.

In gene therapy, one of the main challenges arises from the nucleic acid delivery. Due to their chemical/physical properties which make them not stable enough when administered in a naked formulation and to the possible side effects of insertional events, nucleic acid-mediated therapy needs a smart delivery system to reach the specific cellular compartment to act as therapeutic agent[152].

3.1 History and challenges

Gene therapy, then, is based on the delivery of therapeutic nucleic acids in precise target cells. In theory is a simple concept but it has presented considerable obstacles during its development. Interestingly, the first approved gene therapy performed on humans was tried in 1990, the same year when the Human Genome Project began, provoking excitement in the scientific community. However, Martin Cline conducted the very first experiments on two patients affected by thalassemia in 1979/1980, without official approval[153]. Since the trials were criticized for scientific and procedural reasons, National Institute of Health (NIH) established a Recombinant DNA Advisory Committee (RAC) for human clinical trials approval. In 1989, then, W. French Anderson set the first approved gene therapy trial to investigate its applicability and side effects[154], and in 1990 they tried to cure two patients with severe combined immunodeficiency (SCID), a serious monogenic disease, by *ex vivo* introduction of functional adenosine deaminase (ADA) gene using retrovirus, without good results[155]. In that first trial, a monogenic disease was chosen because it's caused by mutation in a single gene and appears to be an easier target for gene therapy.

These first experiments accelerated the pace and in the decade of '90s, 400 protocols approved by NIH/RAC with over 3,000 patients had been registered, but, unfortunately, only some anecdotal reports of success were reported[156-158]. Despite the initial hopeful enthusiasm, the reality showed a series of bad results crown in an eighteen-year-old

man death in 1999: the first death directly imputing to Gene Therapy. This caused a mistrusted general opinion about gene therapy in non-scientific community and lead scientists to revise trial design and ethical standards in order to avoid similar accidents.
Moreover, in 2002, two of ten children affected by SCID develop a leukemia-like condition after insertional mutagenesis near a proto-oncogene, stopping all clinical trials for an exhaustive revision[159]. Negative results mentioned before have influenced public perceptions of gene therapy. It indeed generates concern about genetic engineering and nucleic acid manipulation for genetically modified organisms. Nonetheless, from 2002 to 2011, first gene therapy successes were observed. In 2003, Gendicine was launched in China as the first commercial gene therapy product[160] meanwhile positive results were reported for X-linked SCID[161], ADA-SCID[162], Parkinson's disease[163], adrenoleukodystrophy[164] and chronic myelogenous leukemia (CLL)[165] in clinical trials.
In November 2012, the European commission, under European Medicines Agency recommendations, approved the commercialization and application of uniQure's Glybera treatment under exceptional circumstances. This treatment is directed to adult patients affected by familiar lipoprotein lipase deficiency (LPLD), and utilizes non-replicating AAV1 vectors for LPL gene substitution[166,167]. Since 1989, more than 2000 clinical trials had been approved up to June 2014, and nowadays only two protocols are in phase IV (**Fig. 19**).

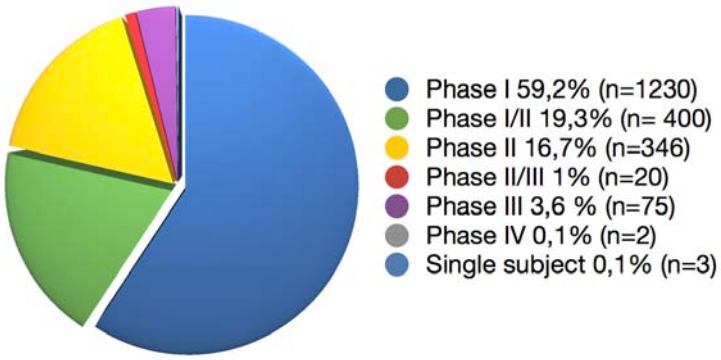


Fig. 19 Statistics on phases of gene therapy clinical trials up to June 2014. Adapted from [168].

3.2 Properties of gene therapy vectors

There are two major families of nucleic acid delivery systems: viral and non-viral vectors. Viruses are recognized as highly efficient delivery system but, unfortunately, viral components can cause severe adverse immunological reaction. On the contrary, non-

viral vectors are generally safe, allows a higher degree of structural modifications and show less delivery efficiency if compared with viruses.

There are three different ways to deliver nucleic acids to the targeted cells (**Fig. 20**): the first one involves an *in vitro* treatment of cells previously removed from the patient. After the manipulation, cells are transferred back to the patient's body. In the second one, nucleic acids are delivered directly to the target tissue (i.e. by injection). Third, the *in vivo* treatment: therapeutic nucleic acids are injected into the blood and reach the targeted cells by itself through the blood torrent. Although the last one is the most "patient friendly" method, it's the more difficult to develop. In all cases, a nucleic acid transport vehicle is required.

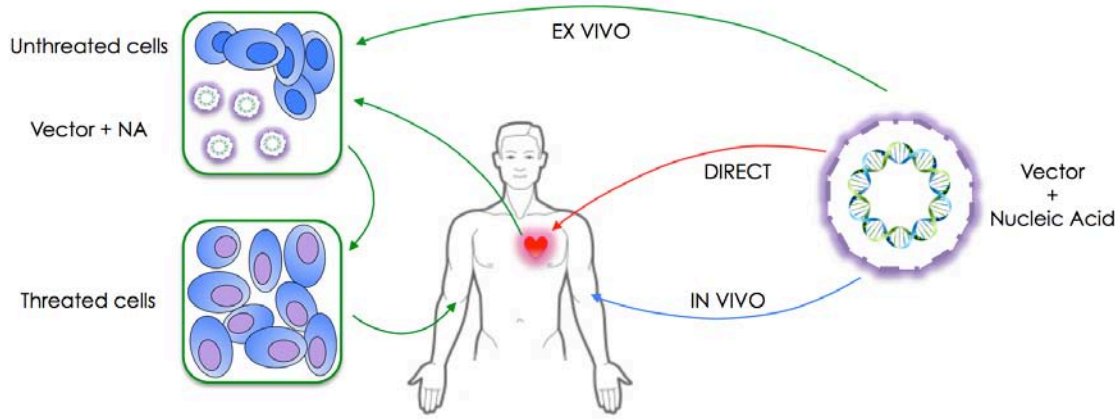


Fig. 20 Schematic representation of different approaches for gene therapy. Adapted from [169].

It's now then clear that clinical useful protocols are needed to develop a system that allows to deliver nucleic acids to the targeted cells without side-effects or with acceptable commitment. The delivered therapeutic genetic material should not cause any undesired change in the genome and, more importantly, its activity should be highly controlled and adjustable.

Developing **safe and efficient delivery systems** is a key step for successful gene therapy progress. A good delivery system should own not only a high medical standard quality, but also have to take care about patient acceptance, being commercially attractive and easily approved by regulatory authorities.

The vector should be **easy and cheap to produce**. Time of production, costs and yield are the most relevant factors that manufacturing company takes into consideration before investing for a commercial product. Moreover, product should be easily delivered and stored without showing degradation or handling troubles.

The vector components should **not elicit immune response**. This is particularly important if several therapeutic boosts are planned. Adverse immune reaction could nullify the therapeutic action and provoke severe complication or even patient's death[170].

In all kind of treatments, but especially if targeted cells are part of heterologous population or are widespread throughout the body, a **good tissue targeting** is also a key vector feature. A non-selected delivery can lower the efficacy of the treatment and be potentially dangerous since can elicit immune response.

Since genetic material harbored into vector could be of variable length and also could need regulatory sequences, the **vector size capacity** must be taken into account. The ideal vector should not show any size limit to deliver nucleic acids. Moreover an ideal vector should be able to **infect both dividing and non-dividing cells**[171].

3.3 Biological barrier

Physicochemical properties of gene therapy vehicles and their interactions with biological components determine their biodistribution. Therefore, they must be modified and designed in order to overcome several biological barriers that can condition the therapeutic efficiency[172,173]. The number and nature of barriers encountered by a vector in an *in vivo* administration in humans depend on the route of administration. Among different strategies like gene gun[174] or direct injection in tissue[171], intravenous administration are recognized as the less invasive and well accepted by patients.

The biological barriers can then be classified in two classes: extracellular and intracellular barriers (**Fig. 21**).

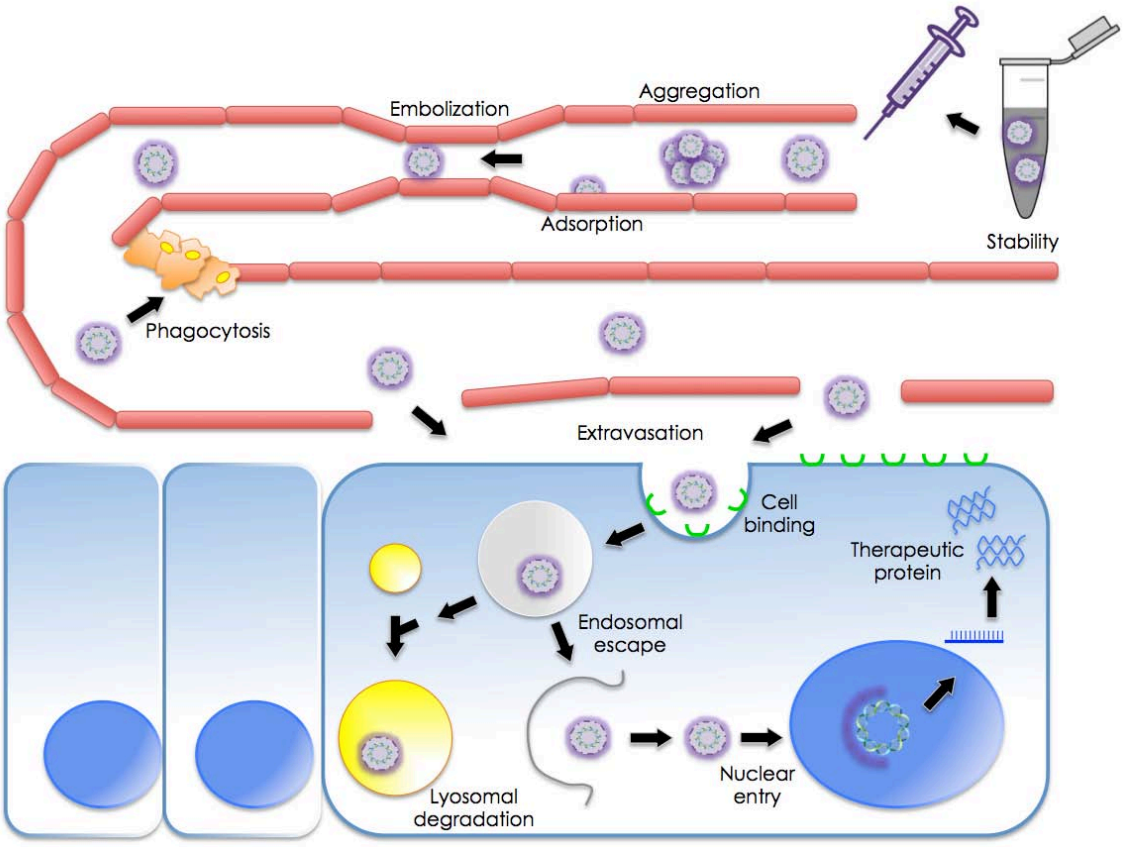


Fig. 21 Schematic representation of biological barriers in gene delivery therapy. Adapted from [175].

3.3.1 Extracellular barriers

Extracellular barriers can be encountered from the very beginning of delivery as the point of injection, to the surface of the cellular target.

Above all, the delivery vector system itself should show low toxicity, low or null immunogenicity and *in vivo* stability. The molecular architecture of non-viral carriers can be tuned in order to modulate the toxicity but every modification should be carefully considered, since the molecular architecture is the essence of the delivery system itself. Especially if target cells are spread across the body, gene delivery particles' stability is a key factor required for extended circulation time, helping to reach their specific cell target. Stability can be affected by the charges of particles, in example, strong positive charges can facilitate non-specific interactions with the extracellular matrix, cell surface or plasma proteins. On the other side, strong negative charges can cause scavenging by macrophage phagocytosis.

Along the blood circulation vessels, carriers can encounter several agents and its interactions are difficult to control and sometimes unpredictable. Serum proteins and blood cells are the first factors influencing the vector biodistribution. Again, excess of

positive or negative charges can lead to a decrease of general efficiency of the delivery or to the release of the transported nucleic acid[176].

To reach the cell target, vector must cross the endothelial cells from blood vessel to target cell in a process called extravasation. There are diverse types of vascular walls but the main characteristic that helps extravasation is a small carrier size. Nonetheless, alteration of vascular permeability by using vasodilator can help the vector to spread out in tissues[175,177].

3.3.2 Intracellular barriers

Intracellular barriers then, can be found in cellular binding, uptake, vesicles escape, trafficking and entry into the nucleus, nucleic acid release and vector degradation. In systemic *in vivo* application, vectors are generally functionalized with ligands to confer target specificity. Cell entry generally can occur in several ways: via receptor-mediated endocytosis (clatrin/caveolae mediated or independent), via macropinocytosis or via a non-endocytic pathway. Usually, peptide vectors are internalized by endocytosis into endosomal vesicles[178]. These vesicles, in the late stages of cellular trafficking, tend to fuse with lysosome for degradation by low pH and nucleases. Endosomal escape is then essential for saving the nucleic acid cargo. One of the most popular strategy is disrupting endosomal membrane with aminoacid "proton sponge" domain, which cause endosomal collapse by osmolitic swelling and subsequent release of the endosomal content in cytoplasm[179,180]. Other solution could be the insertion of viral fusiogenic peptides: a mix of hydrophobic and negatively charged amino acids with regular pattern. The low endosomal pH causes a conformational change in this domain, leading a pore formation in endosomal membrane and subsequent exit of the content[181].

Once in the cytoplasm, when dealing with expressible DNA-containing vectors, the vector must reach the nucleus. In dividing cells, nucleic membrane is disrupted, facilitating the nucleic acid cargo to reach the target. In non-dividing cells, nuclear membrane is intact and only 45 Kda/10-30 nm molecules are able to cross through nuclear pore system by passive diffusion. Larger molecules require an active transport system, mediated by a specific aminoacidic sequence. For these reasons a Nuclear Localization Signal domain (NLS) from simian Virus SV40 or GAL4, Tat, protamines, are generally incorporated in vectors. This permits a receptor-mediated entry of delivery particles, which can then release the nucleic acid cargo into the nucleus[182].

3.4 Therapeutic nucleic acids

Depending on the nature of the nucleic acid to be delivered, gene therapy approaches can follow different strategies of action:

3.4.1 Gene delivery

Deletion, mutation or compromised function of a single or multiples genes can cause a wide spectrum of diseases as cystic fibrosis, muscular dystrophy, hemophilia and also cardiovascular diseases or cancer[183]. Substitution of the gene or genes involved in the disease and restoring the normal phenotype is then the main goal of gene delivery therapy (**Fig. 22**). In an *in vivo* administration, the half-life of naked plasmid DNA in blood torrent is approximately of 10 minutes. Then, in order to protect it from proteases and subsequently to enhance circulation time, DNA is generally entrapped in nanoparticulate carriers. As mentioned before, packaging into nanoparticles permit the DNA to reach the nucleus, crossing the nuclear membrane. Once into the nucleus, DNA can be expressed without being incorporated in the host genome (transient expression) and function is then lost in dividing cells or gene can be incorporated into the genome and then maintain the function in dividing cells with a permanent expression.

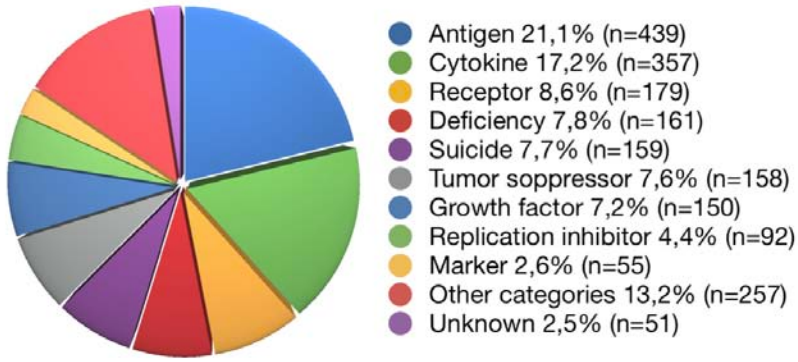


Fig. 22 Statistics of gene types transferred in gene therapy clinical trials up to June 2014. Adapted from [168].

3.4.2 Antisense oligonucleotides

The use of antisense oligonucleotides is adopted for "switching off" an undesired gene activity. Silencing the expression of growth factors, growth factor receptors, or genes directly involved in cell cycle progression in cancer is an example of it, among others[184]. This approach consists in the delivery of DNA or RNA that will bind to nuclear double stranded DNA (antigene strategy) or mRNA (antisense activity) inactivating the gene expression. Gene silencing can also be obtained by a therapeutic DNA/mRNA complex that stimulates the mRNA digestion by RNAse H.

3.4.3 RNA interference (RNAi)

RNAi is an eukaryotic gene silencing method that causes the destruction of mRNA involved in the disease process. It is based on the delivery of specific double stranded RNA molecules that, once in the nucleus, are cleaved by an enzyme called Dicer. Shorter RNA fragments generated by Dicer cleavage are called Small Interference RNAs (SiRNAs). SiRNA are then the responsible of mRNA silencing by forming a SiRNA/mRNA complex, which is recognized by RISC/Argonaute protein cleavage complex, causing the degradation of the target mRNA[185].

Interestingly, low doses of SiRNA are required for satisfactory gene knockdown since it stimulates the mRNA cleavage in a catalytic manner. Moreover, expression of Small Hairpin RNAs (ShRNAs) by a DNA vector has being purposed as an alternative of RNA delivery. DNA-directed RNAi (ddRNAi DNA) plasmid codify for ShRNA that are processed by Dicer enzyme and SiRNAs generated acting as previously described[186] (**Fig. 23**).

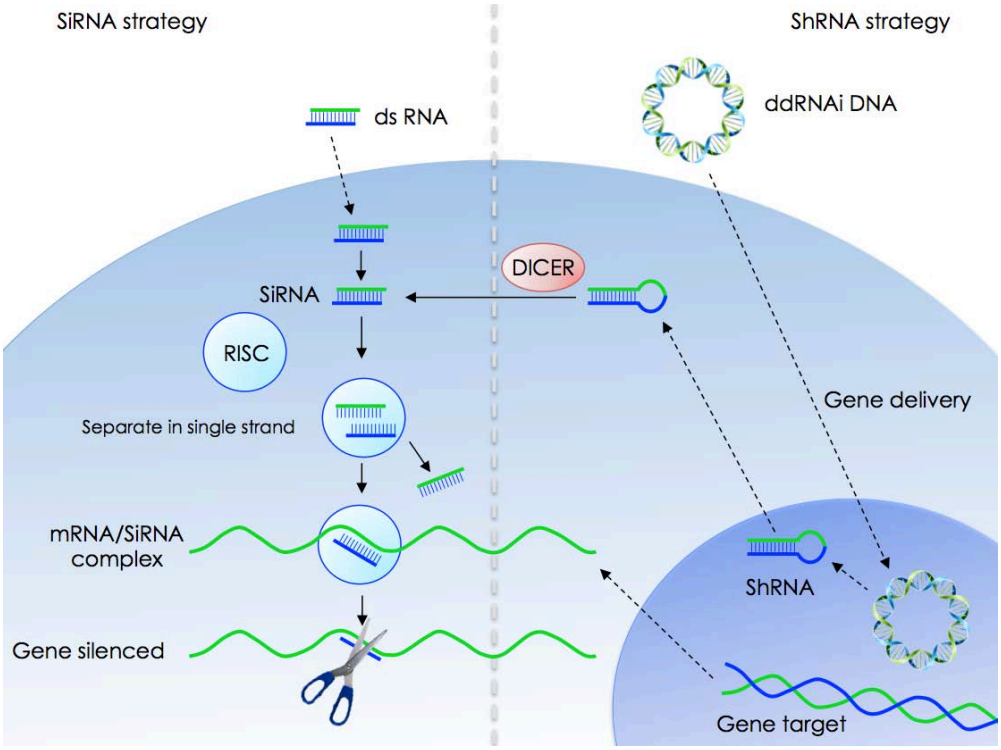


Fig. 23 Schematic representation of SiRNA and ShRNA gene therapy approaches. Adapted from [187].

3.4.4 Splice-Switching Oligonucleotides

In higher eukaryotes as humans, almost all genes are predicted to undergo alternative splicing[188]. This is a biological process that enables a single pre-messenger RNA to

generate multiple protein isoforms with diverse functions and up to 50 % of human genetic diseases involves gene mutations that influence this process.

Alternative splicing, then, can be resumed as a competition among splice sites and splicing elements for the spliceosome.

Splice-Switching Oligonucleotides (SSOs) are 15 to 20 base in length nucleic acids, that bind mRNA sequence elements and block the access of spliceosome and splicing factors, pausing the splicing events. Moreover, complexion with mRNA avoids RNAseH cleavage and prevents mRNA degradation.

SSOs can be applied to:

- Restore correct splicing in wrong splice transcripts
- Switch a splice variant to another
- Produce a novel splice variant

SSOs delivery, contrarily to RNAi or antisense, can enhance a gene activity and is then purposed for disease like Duchenne's muscular dystrophy (DMD), β-thalassemia, or arteriosclerosis, among others[189].

4 Gene therapy vehicles

It is well established, then, that the major issue for a successful gene therapy is the adoption of safe and efficient gene delivery vehicles. A lot of different strategies have been developed to ensure the therapeutic gene entrance into the targeted cells. Gene therapy vectors can be classified in two large groups: viral based carriers and non-viral based carriers. The majority of the vehicles used in gene therapy trials are viruses (**Fig. 24**).

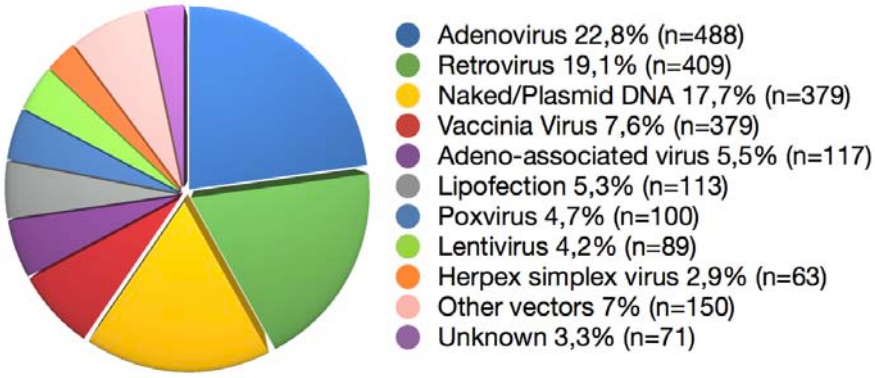


Fig 24 Statistics on vectors used in gene therapy clinical trials up to June 2014. Adapted from [168].

Viruses are naturally evolved nucleic acids carriers, representing a specific system for gene transfer and expression. They were first used as gene vectors in mammalian cells, showing their major and best characteristics and drawbacks: high efficiency and low biosafety, respectively[190]. The toxicity of viral proteins, the possibility of random integration of the vector DNA into the host genome and their high production cost are challenging aspects of viral vectors which have leaded to develop alternatives for better standard quality of gene therapy delivery.

On the other side, non-viral vectors show high tunable architecture adaptable to specific therapeutic features, with the unique disadvantage of relative low efficiency.

Major advantages and disadvantages of viral and non-viral approaches are summarized in **Table 1**.

Table 1: Advantages and disadvantages of gene therapy vectors.

Type of vectors	Advantages	Disadvantages
Viral vectors	High transfection efficiency	- Immunogenicity, oncogenesis - Viral protein contamination - Unfavorable pharmacokinetics - Costs and safety problems in large scale

		production - Low storage stability
Non-viral vectors	- Adjustable immunogenicity - Opportunity for chemical/physical manipulation - Favorable pharmacokinetics and large scale production - Storage stability - Plasmid independent structure	- Relatively low transfection efficiency

Adapted from[191].

4.1 Viral vectors

Viruses are biological structures that are able to enter the cell and exploit the cellular machinery to express their own genetic material, replicate, and then spread to other cells[192]. To be used as gene transfer vectors they must be engineered, in order to carry the selected therapeutic nucleic acid and control their immunogenicity.

The main viruses families studied as gene delivery systems are: Adenovirus, Adeno-associated virus, Retrovirus, in particular Lentiviruses as HIV, Vaccinia virus and Herpex simplex virus. Each viral system has its own advantages and disadvantages, summarized in **Table 2**.

Table 2: Advantages and disadvantages of viral vectors.

Vector system	Advantages	Disadvantages
Adenovirus	- High transduction efficiency *ex vivo* and *in vivo* - Large host spectrum - Transduces proliferating and non-proliferating cells - Production easy at high titers - Non-enveloped - Wild type cause a mild disease - Episomal virus genome	- Immunogenic - Reversal to wild type - Short period of gene expression in dividing cells - Leakage of viral proteins - Remains episomal - Transient expression - Requires packaging cell line - Immune-related toxicity with repeated administration - No targeting - Limited insert size: 4–5 kb
Adeno-associated virus	- Large host spectrum, including mitotic as well as post-mitotic cells - No associated to human disease - Integration on human chromosome 19 to establish latent infection - Prolonged expression - Small genome, no viral genes	- Immunogenic - Not well characterized - No targeting - Requires packaging cell line - Potential insertional mutagenesis - High titers (1010 pfu/mL) but production difficult

		- Limited insert size: 5 kb
Retrovirus	- High transduction rates *in vitro* - Vector proteins are not expressed in the host - Integration into cellular genome - Broad cell tropism - Prolonged stable expression - Requires cell division for transduction - Relatively high titers - Larger insert size: 9–12 kb	- Immunogenic - Risk of insertional mutagenesis - Inactivation by complement fractions in the serum - Low delivery rates *in vivo* - Inefficient transduction - Insertional mutagenesis - Requires cell division for transfection - Requires packaging cell line - No targeting - Potential replication competence
Retrovirus-Lentivirus	- Transduces proliferating and non-proliferating cells - Transduces hematopoietic stem cells - Prolonged expression - Relatively high titers	- Immunogenic - Difficult to manufacture and store - Limited insert size: 8 kb - Clinical experience limited - Low efficiency *in vivo*
Vaccinia virus	- Potential for the development of a variety of gene vaccines	- Use is restricted to individuals not previously vaccinated - Use in the immunocompromised is not indicated
Herpex simplex virus	- Episomal virus genome - Large insert size: 40–50 kb - Neuronal tropism - Latency expression - Efficient transduction *in vivo* - Replicative vectors available	- Immunogenic and cytotoxic - Variable selectivity - EBV is oncogenic - Activation of latent virus - Low transduction efficiency - System is under development - No targeting - Requires packaging cell line - Transient expression, does not integrate into genome - Moderate titers

Adapted from[193,194].

As mentioned before, the major drawbacks of viral vectors are well-established clinically including important side effects in target individuals, ranging from inflammation to death[190]. Moreover, evidence indicates viral carrier involvement in leukemia development in treated patients[159]. Nonetheless some good results were achieved as in the case of the 18 year old male patient with ß-thalassemia, successfully treated with lentiviral vector to transduce the human ß-globin gene into purified blood and marrow[195] or in goals achieved using Adeno-associated viral vectors for Leber's congenital amaurosis treatment[196].

4.2 Non-viral vectors

As an alternative to viral vectors, non-viral approaches show high potential due to safety and customization advantages in vector architecture. In general, they can transfer larger nucleic acids, they are less toxic, easy and cheap to produce or apply, and they can be associated with ligands, conferring them a high cell targeting specificity. Anyway, they generally show less transfection efficiency and poor transgene expression[197].

Non-viral nucleic acid delivery systems can be classified into two subgroups according to the strategy to get into the cell:

Physical methods: nucleic acids are delivered to their target cell by physical forces to weaken the cellular membrane in order to facilitate the entrance[198] (**Table 3**).

Table 3: Advantages and disadvantages of physical methods.

Physical method	Advantages	Disadvantages
Nucleus microinjection	- High transfection rates - Avoids cytoplasmic and lysosomal degradation of the injected material	- *Ex vivo* use only - The technique is labor-intensive and requires well-isolated cells - Potential for use in germ line - Technical and ethical issues
Electroporation	- High transfection rate	- *Ex vivo* use only - Low efficiency - Highly damaging for cells
Hydrodynamic injection	- Simplicity - 2-19 Kb capacity - Good for use in gene vaccines	- Low transgene expression - Use limited to skin, thymus and striated muscle
Gene gun	- High transfection rates - Precise dosage delivery	- Transient expression - Cell damages in shooting region
Ultrasound	- Non-invasive method - Can be combined with micro bubble technology - Cost effective	- Modest efficiency - Under development

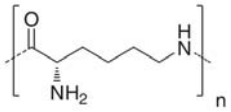
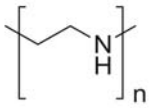
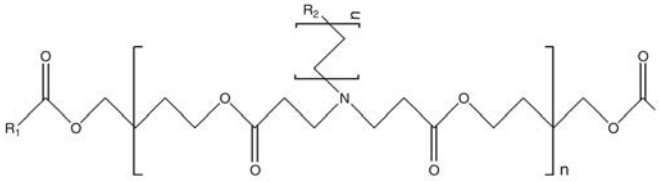
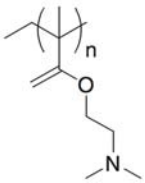
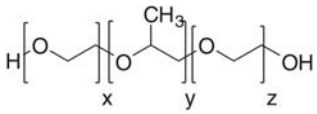
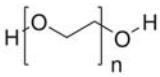
Adapted from[193,194,199].

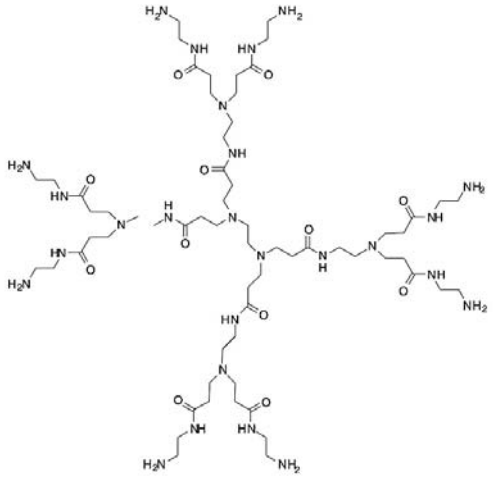
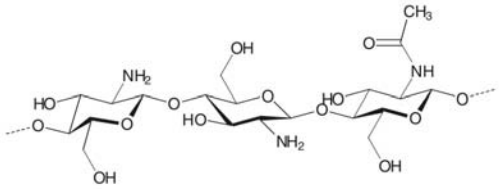
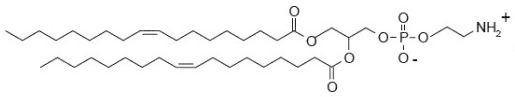
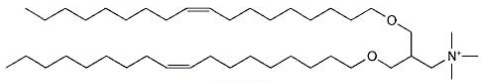
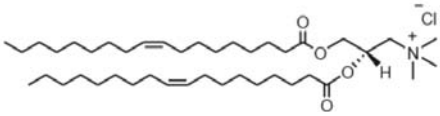
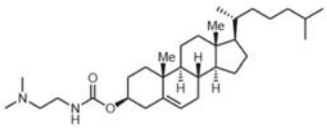
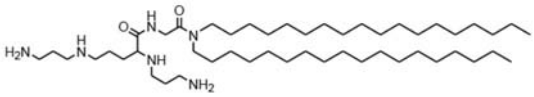
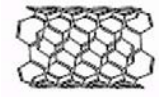
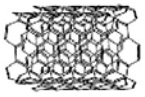
Some examples of physical methods applications are the dystrophin plasmid-based gene delivered through injection via the radialis muscle in Duchenne/Becker muscular dystrophy patients, which have shown good results in phase I clinical trials[200] or the

intratumoral injection of plasmidic DNA encoding antiangiogenic factor FLT-1 for phase I experimental therapy of vulvar cancer neoplasms[201].

Chemical methods: nucleic acids are delivered to target cells by synthetic or naturally occurring carrier compounds[198] (**Table 4**).

Table 4: Classifications of most common chemical carriers.

Name	Basic structure and/or example	Category
Poly L-Lysine (PLL)		**Polyplexes**
Polyethylenimine (PEI)		
Poly(β-aminoesters)		
Dimethylaminoethyl Methacrylate (DMAEMA)		**Polyplexes**
Pluronic		
Polyethylene Glycol (PEG)		

Dendrimers (PAMAM)		
Chitosan		
DOPE		**Lipoplexes**
DOTMA/DOPE		
DOTAP/DOPE		
DC-Chol		
DOGS		
Single walled CNT		**Carbon nanotubes**
Multi walled CNT		

Adapted from: [193,194,197,202-205].

Promising results were obtained by direct intratumoral injection with a plasmid DNA encoding the genes HLA-B7 and β2-microglobulin complexed with a DMRIE/DOPE cationic lipid mixture, for metastatic melanoma treatment[206].

4.3 Protein nanoparticles

Protein nanoparticles for nucleic acid delivery are being revealed as an interesting tool for the improvement of diseases treatment. Self-assembling protein nanoparticles shown to be valid potential candidates for one of the major challenges of gene therapy: the development of highly tunable, efficient and commercially appealing vectors. Protein-only nanoparticles, in fact, can be engineered to contain the entire functional characteristics to overcome biological barriers for nucleic acids delivery. There are two main architectural principles for protein-only delivery nanoparticles: the first is based on conventional genetic engineering of functional modules, leading to a recombinant production of multi functionalized proteins, thanks to specific domains with different biological functions. The second one is based on recombinant production of the structural proteins of viruses, which leads to the formation of empty viral particles called Virus-Like Particles (VLPs).

Both approaches take advantage of recent proteomic studies, deep knowledge on physicochemical properties of polypeptides and the assembly of different peptidic modular domains.

4.3.1 Multifunctional proteins

Multifunctional proteins can be designed following two main strategies: *de novo* protein rational design or modular engineering (**Fig 25**).

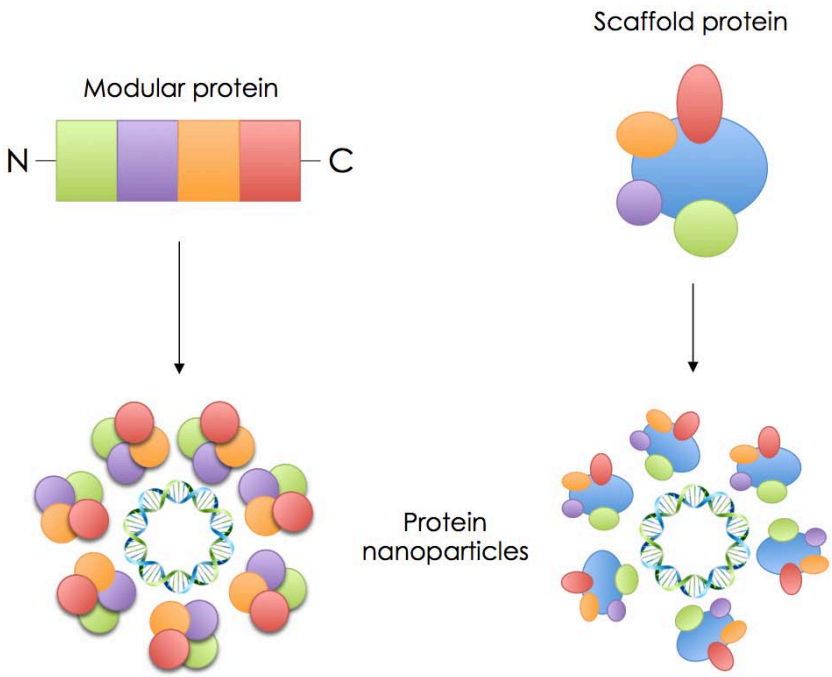


Fig. 25 Schematic representation of modular protein and de novo scaffold protein approaches for multifunctional proteins as non-viral vectors for gene therapy. Adapted from [207].

***De novo* rational protein design** is based on the insertion of amino acidic sequences in specific sites of a protein, which acts as a backbone scaffold[208]. These amino acidic blocks should be inserted minimizing the impact of the original three-dimensional structure of the scaffold protein and, at the same time, alter or confer new specific functions, lacking in the original protein. In order to obtain the best results for nucleic acid delivery, this strategy requires a deep knowledge of the functional domains, structure and activity of backbone protein properties. Modern genetic engineering techniques and bioinformatic tools make this option affordable and feasible. Unfortunately, since modification in amino acidic primary sequence can cause a different three-dimensional final structure, inserting foreign domains is not always possible or sometimes leads to unpredicted effects on modified protein. In fact successful introduction of therapeutic genes in *in vivo* gene delivery to the postnatal rat central nervous system (CNS) was achieved by 249AL multifunctional protein. This protein was engineered starting from a β-galactosidase in which a deca-lysine (K_{10}) DNA binding domain and Foot and Mouse Disease Virus-RGD cell recognition domains were inserted in the original sequence, conferring gene delivery properties to the wild type protein[209].

Modular protein engineering is based on different modular domains combined in a single polypeptide. Peptidic blocks are then combined together by conventional genetic engineering, and each module confers a specific biological function to the final construct. The optimal order and number of building blocks cannot be predicted in advance and, since for the same biological function there are different functional domains from different origins, selecting the appropriate modules and the order in which the modules are positioned is a time-consuming and trial-and-error procedure. Since the first steps, research activity in this field show the enormous potential of these two approaches[210,211].

The designing process of modular protein or modification of a protein scaffold requires a multidisciplinary view of the gene delivery process. As mentioned in previous sections, vehicles should own specific physicochemical properties favorable for overcoming biological barriers, for being easily, safely and cost-effective produced, and being stable outside and inside the cells.

The different modules can derive from different origins like natural sources, protein/peptide libraries or artificial sequences. Both in modular proteins and in *de novo* rational design, the election between different available domains depends on the required vector characteristics. Moreover, once established the optimal functional building blocks, it's important to set the order in the final construct. The domain order, in

fact, can alter the final construct's characteristics, or inactivate the biological activity of building blocks themselves[212]. Moreover, modular domains can have also a therapeutically effect themselves in the target tissue. An example of this "therapeutic module" is the RGD cell attachment motif, widely used in non-viral protein vehicles[213], which is involved in cell proliferation triggering and nervous cell differentiation[214].

The modules that can be used for protein vehicle design for gene delivery are:

- Nucleic acid condensation peptides (i.e. K_{10} domain, **Annex 2**)
- Blood brain barrier (BBB) crossing peptides (i.e. Angiopep-2 peptide)[215]
- Cell-penetrating peptides (CPP)(i.e. R9 domain)[216]
- Receptor-specific ligands (i.e. T-22 domain)[217]
- Endosomal escaping fusogenic peptides (i.e. GALA domain)[181]
- Endosomal escaping histidine rich peptides (i.e. H6 domain)[218]
- Nuclear import peptides (i.e. SV40 NLS domain)[219]
- Therapeutic modules (i.e. RGD domain)[213]

Inside each of these categories, a lot of peptides can be chosen, giving multiple possibilities of functional block combinations.

4.3.2 Virus-like particles (VLPs)

Virus-like particles (VLPs) are self-assembling, non-replicating viral particles formed by one or several viral structural proteins, lacking the viral genome (Fig. 26). The grade of similarity of VLPs to the original virus structure can vary, depending on the original virus morphologic complexity and depends mainly on the number of proteins involved in VLP formation[220].

More than 30 years ago, first experiments with Murine Polyomavirus (MPyV) VLPs, demonstrated that VLPs could package nucleic acids and transduce cells *in vitro*[221,222], leading to the development of this new transport system.

The main purpose of this delivery strategy is, then, to produce a delivery nanoparticle showing the advantage of viral vectors, mainly in terms of nucleic acid transport, cell tropism, nuclear entry, but lacking of immunogenicity, viral genome integration and production difficulties.

VLPs' key characteristic is the possibility to disassemble and re-form the nanoparticles structure *in vitro*, in order to eliminate possible residual nucleic acids from protein structures and fill VLPs with the therapeutic genes. Moreover, lacking of any kind of viral genome excludes any possibility of replicative competence.

VLPs can be produced in different cellular systems including bacterial[223,224] plant[225], insect[226,227], yeast[228,229] and mammalian cells[230,231], through transfection with a single plasmid encoding only viral structural proteins.

VLPs structures can be then generated from many families of viruses, including polyomaviruses[232], papillomaviruses[233], hepatitis B and E viruses[234,235], lentiviruses[236], rotaviruses[237], parvoviruses[238] and noroviruses[228]. Genetic engineering also permits to modify VLPs forming proteins, in order to confer or change the specific vector tropism[239].

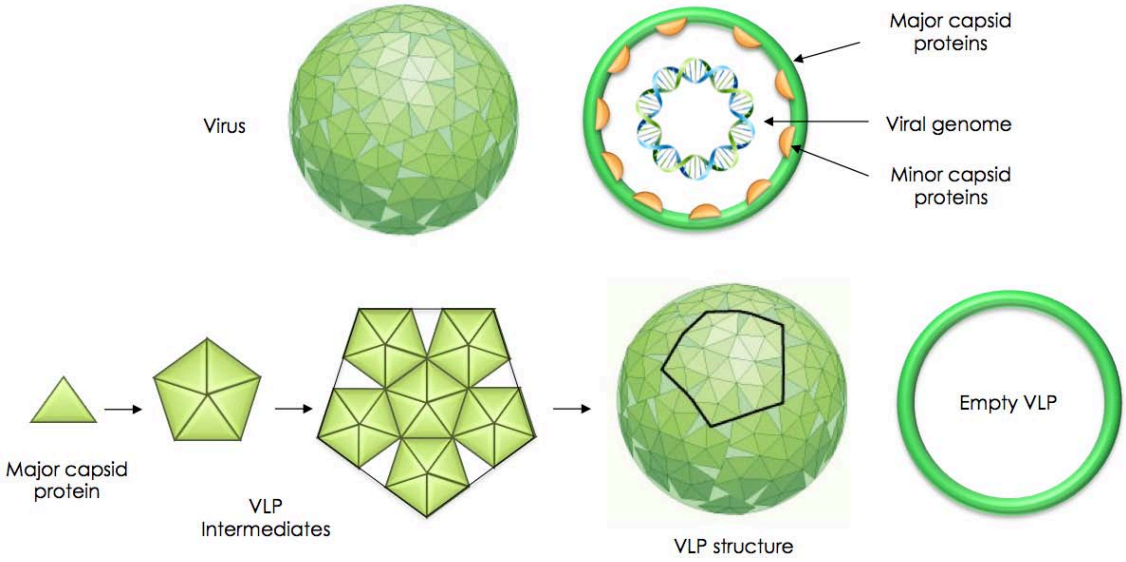


Fig. 26 Schematic representation of the model of Virus-Like Particles structure. In the upper part the wild type virus and virus section. In the lower part the formation process of VLP and particle section. Adapted from [240].

5 Physicochemical characterization of multimeric protein-only nanoparticles

The design, production, characterization, handling and controlling protein nanoparticles at the atomic and molecular scale belongs to a brunch of science called Nanotechnology. Essential parameters of protein-based gene delivery vectors are, then, particle size, molecular organization and nanoscale properties (**Fig. 27**). The ideal gene carrier should be a well-defined nanoparticle of less than 100 nm in size that enables all the functions required for gene transfer. The physicochemical and functional characterization of nanoparticles is essential to understand and optimize their behavior at molecular and cellular level. Imaging tools, like transmission electron microscopy (TEM) and atomic force microscopy (AFM) can give the first visual information about size and morphology of nanoparticles. Dynamic light scattering (DLS), Zeta-potential measurement, electrophoretic mobility shift assay, nucleic acid retardation assay (EMSA), nuclease protection and hemagglutination assay provide additional information about size, stability, surface charge, condensation and nucleic acid protection.

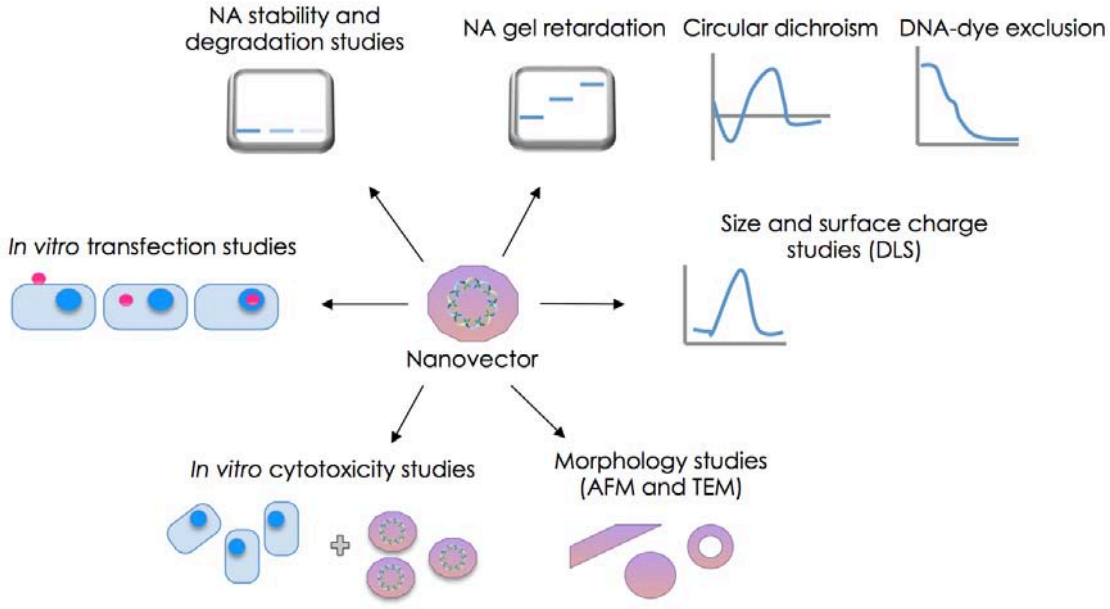


Fig. 27 Schematic representation of in vitro physicochemical characterization assays of nanovectors. Adapted from [241].

5.1 Size and Surface charge

Nanoparticle's size parameter exert significant effects on general deliver efficiency, since can affect biodistribution, cellular uptake, endosomal escape, DNA packaging/unpackaging and nuclear internalization[242]. Dynamic light scattering (DLS)

permits to estimate size and Zeta-potential value of small particles in suspension or polymers in solution. Surface modifications by multifunctional groups can change the vector size and surface charge distribution. Exposed hydrophobic residues and altered surface charges are the major responsible of nanocarrier opsonization and clearance from the blood torrent by phagocytes. Zeta-potential is of value to study surface properties on the nanostructure surface. Zeta-potential is then influenced both by the composition of nanoparticles and the medium in which they are dispersed. Since the interactions between particles affect their colloidal stability, this parameter can give important information about repulsive forces and, subsequently, nanoparticle stability. Generally, nanoparticles showing Zeta-potential values above/under +/- 30 Mv have enough surface charges to prevent aggregation and remain stable in suspension. Moreover, Zeta-potential can reveal whether a nucleic acid is encapsulated inside the nanoparticle scaffold or just adsorbed onto the surface[243-245].

5.2 Self-assembly

Nanoparticle self-assembly refers to the process by which small-scale building blocks are spontaneously organized in thermodynamically stable and ordered structures. Spontaneous self-assembly is a common event in biology. Several non-covalent interactions are involved in this fact, mainly due to hydrophobic interaction, electrostatic energy, and Van der Waals forces. Moreover, nucleic acid–peptide interactions, salt concentration, mix order and nucleic acid and protein ratio can also strongly influence the condensation process[246]. As mentioned in section 4.3.2, VLPs naturally tend to self-assemble in highly ordered structures because of their viral origin. Interestingly, changing the storage buffer composition can control VLP assembling and disassembling. This simplifies the nucleic acid loading inside the nanoparticles and permits clearance from contaminants[247]. On the other hand, recombinant proteins can also self-organize in nanoparticles. Apart of bacterial micro-compartment proteins, little cationic peptides are recognized to form nanoparticles suitable for non-viral gene therapy. For example, arginine-rich cationic peptide used as building block in R9-GFP construct is described to induce the formation of protein nano-disks, condense DNA and deliver it in the nucleus for gene expression[248].

5.3 Nucleic acid-protein interactions

Other of the major concerns in gene delivery system is the nucleic acid stability. A stable nucleic acid-protein formulation is essential for the development of a therapeutic product. Even though in a sterile aqueous solution at 4°C, pH7.5, plasmid DNA can be stable up to 2 years, at room temperature stability is compromised and not always suitable for long-term storage. Moreover, degradation by nucleases, radical

oxidation and UV light, can compromise the integrity of DNA[249]. The efficiency of protein-based nanocarriers is then related to the protection of DNA, which is condensed and stabilized, resembling the viral performance.

DNA is a high molecular weight polyanion with highly organized chemical structure. Random coil DNA volume depends only partially on the length. Flexibility and stiffness are due to intramolecular electrostatic interactions that control the nucleic acid condensation in a given environment. In polypeptide-based nanoparticles, cationic residues reduce negative charge repulsion between adjacent phosphates[250]. It has been observed that cationic charges have to be clustered in groups of three or more cationic groups for DNA[251]. An exhaustive overview of proteins used in non-viral gene therapy is reported in **Annex 2**.

5.4 Functional characterization

As described above, physicochemical characterization is extremely helpful for revealing structural features of nanoparticles. Nonetheless, a key step in gene vector development is the functional characterization. A delivery test with *in vitro* cell lines is the first step to understand the functional properties and pharmacokinetics for a novel vector. It permits determining the performance of the carrier either for nuclear or cytoplasmic delivery. Moreover, *in vitro* assays allow to establish whether the vector uptake is receptor specific or not, and if it interferes with other cellular pathways. A spectrum of detailed settings needs to be considered for a valid *in vitro* assay, from which the most important are[191]:

- Choice of the cells to transfect
- Cell maintenance and cell confluence parameters
- Particle formulation
- Dose and incubation time
- Transfection efficiency measurement
- Toxicity measurement
- Cell induced changes detection
- Positive/negative controls and eventually a competitor choice

So far, a standard protocol for *in vitro* assay is missing. Due to the many setting variables and efficiency, comparison between vectors developed by different research groups can result troublesome. Multiple experiments with a group of candidates and different cell lines permit to screen and select the best candidates for subsequent *in vivo* assays.

Like in *in vitro assays*, there is not a standard protocol for *in vivo* experimentations. The only firm ethical rules for use of animals in *in vivo* assays is the triple Rs principle of replacement, reduction and refinement[252].

6 Considerations

The development of a novel protein nanocarrier for non-viral gene therapy is an exciting, challenging and multidisciplinary process. It involves a high number of steps, each of them being critical to obtain a high efficient, safe and commercially attractive product. Protein nanoparticles are being envisioned as an extremely powerful tool for gene delivery since they can form highly complex ordered structures to harbor therapeutically nucleic acids, and, moreover, own a high level of versatility. Since the very beginning of recombinant protein production, cell factories offer an extremely varied range of options, which must be carefully settled in order to obtain best yield and quality product. Deep comprehension of cellular mechanisms allows to module cellular metabolism to exploit as much as possible the protein production system. In this way, bacterial protein quality control system offers a set of molecular tools, also known as folding modulators, which can be used to adjust cell factories recombinant protein production.

One of the most important nanovector features is its capacity to bind nucleic acids. Unfortunately, high affinity to genetic material can lead to nucleic acid host contamination in the production process. It is considered a problem since it can reduce transgene expression efficiency or cause immune adverse reactions. Nonetheless, obtaining an efficient and safe nanocarrier is not only a matter of production; downstream procedures offer an interesting possibility to handle and modify carriers' characteristics in order to solve production or contamination troubles. Thus, the use of molecular tools in up and downstream processes and physicochemical characterization of protein nanocarriers for gene therapy are discussed in this work.

Objectives

The aim of the first part of this thesis is to explore and control the conformational quality of *hJCV* VP1 Virus-Like Particles (VLPs) produced in bacteria and insect cells, upon manipulation of the chaperone network. For this purpose we set the following objectives:

- To explore the best conditions for production of soluble VP1 VLPs in wild type expression systems, both in *E. coli* and insect cells.
- To set up a VLP purification protocol and the best storage condition.
- To study the variation of protein yield in soluble and insoluble fractions, when VP1 protein is produced in different molecular chaperone backgrounds.
- To characterize the conformational quality of VLPs produced in different chaperone backgrounds.
- To manage and analyze the VLPs open/close conformational state.

The second part of this work focuses on the characterization of recombinant multi-domain proteins that self-assemble as viral mimetics. In this regard, nucleic acid-protein interaction and its effects on protein self-assembling are studied in addition to on nuclease protection derived from protein self-assembling. Moreover, a downstream process for improving the efficiency of the protein-based viral mimetics is discussed. For this section we purpose the following objectives:

- To explore the effects of DNA and environmental conditions on protein self-assembling.
- To estimate the extent of DNA protection of different viral mimetics against nucleases.
- To propose a structural model of the studied DNA-protein complexes.
- To study nucleic acid contamination in production processes and set up a downstream procedure for improving the performance of artificial viruses.

Results

Article 1

Effect of the DnaK chaperone on the conformational quality of *JCV* VP1 virus-like particles produced in *Escherichia coli*

Saccardo P, Rodríguez-Carmona E, Villaverde A, Ferrer-Miralles N.

Biotechnol Prog. 2014 May;30(3):744-8.

Virus-like particles for therapeutical gene delivery can be produced in microbial cell factories as self-assembling structures favoring scalable processes, purification and lowering costs of production. Different genetic backgrounds for protein quality control system in *E. coli* have shown to alter the protein production yield and conformational quality of artificial virus assembly. Being VP1 protein of *human JC virus* able to self-assemble in a virus-like particle structure, we discuss in this work the effects of the prokaryotic DnaK chaperone on VP1 production yield and VLPs conformation quality. For this purpose we used three genetic backgrounds including, wild type expression, over-expression and absence of expression of DnaK molecular chaperone. Surprisingly, in the absence of the molecular chaperone the production yield of VP1 is enhanced but has negative effects on VLPs assembly. Moreover we tested different buffer formulations in order to establish the optimal salt concentration and pH for VLP organization, stabilization and conformation.

AIChE

Effect of the DnaK Chaperone on the Conformational Quality of JCV VP1 Virus-Like Particles Produced in *Escherichia coli*

Paolo Saccardo, Escarlata Rodríguez-Carmona, Antonio Villaverde, and Neus Ferrer-Miralles
Inst. de Biotecnologia i de Biomedicina and Dept. de Genètica i de Microbiologia, Universitat Autònoma de Barcelona and CIBER en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, Barcelona 08193 Spain

DOI 10.1002/btpr.1879
Published online February 15, 2014 in Wiley Online Library (wileyonlinelibrary.com)

Protein nanoparticles such as virus-like particles (VLPs) can be obtained by recombinant protein production of viral capsid proteins and spontaneous self-assembling in cell factories. Contrarily to infective viral particles, VLPs lack infective viral genome while retaining important viral properties like cellular tropism and intracellular delivery of internalized molecules. These properties make VLPs promising and fully biocompatible nanovehicles for drug delivery. VLPs of human JC virus (hJCV) VP1 capsid protein produced in Escherichia coli elicit variable hemagglutination properties when incubated at different NaCl concentrations and pH conditions, being optimal at 200 mM NaCl and at pH range between 5.8 and 7.5. In addition, the presence or absence of chaperone DnaK in E. coli cells influence the solubility of recombinant VP1 and the conformational quality of this protein in the VLPs. The hemagglutination ability of hJCV VP1 VLPs contained in E. coli cell extracts can be modulated by buffer composition in the hemagglutination assay. It has been also determined that the production of recombinant hJCV VP1 in E. coli is favored by the absence of chaperone DnaK as observed by Western Blot analysis in different E. coli genetic backgrounds, indicating a proteolysis targeting role for DnaK. However, solubility is highly compromised in a DnaK$^-$ E. coli strain suggesting an important role of this chaperone in reduction of protein aggregates. Finally, hemagglutination efficiency of recombinant VP1 is directly related to the presence of DnaK in the producing cells. © 2014 American Institute of Chemical Engineers *Biotechnol. Prog.*, 30:744–748, 2014
Keywords: human JC virus VP1 virus-like particles, chaperone, hemagglutination, protein nanoparticle

Introduction

The Human JC polyomavirus (*hJCV*) belongs to the family *Polyomaviridae* and is widespread throughout the human population. It causes progressive multifocal leukoencephalopathy in immuno compromised individuals[1,2] and it has been also associated with cancer.[3] The capsid of the viruses of this family is composed of three structural proteins, VP1, VP2, and VP3. VP1 is the major capsid protein forming the outer shell of the virus shell through the spatial assembly of 72 VP1 pentamers. VP2 and VP3 are involved in the interaction between the core of the viral particle and the capsid and are essential in the virus life cycle.[4–7] However, capsids formed exclusively of VP1 (*hJCV* VP1 virus-like particles-VLPs) can be obtained in heterologous expression systems including *Escherichia coli*, yeasts, mammalian cells, and insect cell-baculovirus expression systems.[8–14] An interesting feature of VLPs is their ability to self-assemble, which can be controlled experimentally allowing the internalization of dyes, nucleic acids, drugs, or proteins in vitro.[10,15–17] In addition, VLPs can be functionalized with cell ligands allowing the specific delivery of the cargo to target cells.[18,19] Interestingly, many VLPs have been proved to elicit sustained immune response in vaccination regimes.[20] However, the administration of *hJCV* VP1 VLPs without adjuvant is not able to stimulate any immune response,[9] making this type of VLPs suitable as nanovehicles for biotechnological and nanomedical applications.

Expression of viral proteins of eukaryotic hosts in prokaryotic expression systems copes with several difficulties. On the one hand, the prokaryotic expression system lacks many of the eukaryotic posttranslational modifications such as glycosylation or the formation of disulfide bonds in the oxidizing cytosol. In addition, the finely tuned chaperone-protease pathways of the protein quality control system are believed to be limiting in expressing cells. In that sense, it has been demonstrated that over expression or depletion of chaperones has a significant effect over recombinant protein yield and conformational quality, with these parameters not necessarily coincident. Assembly of polyomavirus VLPs both in vivo and in vitro seems to be favored by Hsp70 family chaperones in both eukaryotic and prokaryotic systems and is inhibited by the presence of Hsp60 family chaperones.[21,22] In this work, we have studied the effect of the bacterial DnaK chaperone on the production, solubility, and hemagglutination activity of *hJCV* VP1 VLPs, and the influence of this chaperone on the supramolecular organization of

Correspondence concerning this article should be addressed to N. Ferrer-Miralles at neus.ferrer@uab.cat.

polyomavirus VLPs in vivo. DnaK is a folding chaperone that promotes the correct protein folding in cooperation with its co-chaperone DnaJ, and assisted by the nucleotide exchange factor GrpE.[23,24] The results obtained show that, in the presence of DnaK, the total amount of recombinant *hJCV* VP1 recombinant protein is negatively affected although the solubility and biological activity is significantly improved. Then, the increase in solubility is translated into an enhanced conformational quality as the protein obtained in DnaK$^-$ genetic background shows higher titer in hemagglutination assays.

Material and Methods

hJCV VP1 gene cloning and E. coli strains

The *human JCV* VP1 gene (*Jvgp4*, NC_001699.1) was codon optimized for *E. coli* expression and provided by Geneart (Regensburg, Germany). The artificial gene (KF488587) was cloned into pTrc99A expression vector (Amersham Pharmacia Biotech, catalog no. 275007-01) using *Nco*I-*Bam*HI (Roche) restriction sites and transformed in *E. coli* DH5α strain (Invitrogen). Expression of *JCV* VP1 gene was under transcriptional control of the isopropyl-beta-D-thiogalactopyranoside (IPTG) inducible promoter Ptrc.

Plasmid pTrc99a-VP1 was transformed by heat shock pulse into three different *E. coli* expression strains: the pseudo wild type MC4100 [F$^-$ *araD139* Δ(*argF-lac*) *U169 rspL150 relA1 flbB5301 deoC1 ptsF25 rbsR*, Strep^R],[25] its DnaK$^-$ derivative JGT20 [MC4100 DnaK thr::tn10, Step^R, Tc^R, CGSC#: 6152],[25] and MC4100 bearing pBB535[26] that harbors IPTG-inducible, $P_{A1/lac\text{-}O1}$-controlled *DnaK-J* chaperone genes.

hJCV VP1 production and purification

The transformed *E. coli* cells were cultured overnight in LB medium at 37°C. The starter culture was then diluted at 1/20 with LB to a final volume of 500 mL and the mixture cultured again at 37°C until the optical density at 550 nm reached 0.5. The gene expression was induced by the addition of IPTG to final concentration of 0.5 mM. Cells were incubated at 30°C for 4 h, and 65 mL of the culture were then separated and cells harvested (15,000*g* for 15 min at 4°C), washed with phosphate buffered saline (PBS) and harvested again. Pellets were suspended in 6.5 mL of PBS with ethylenediaminetetraacetic acid-free (EDTA-free) protease inhibitor cocktail Complete (Roche, catalog no. 11873580001). Cells were disrupted by two 10 min-sonication cycles at 40–50% amplitude with a Labsonic U sonicator, 8-mm probe (B. Braun Biotech International, Melsungen, Germany). Soluble and insoluble fractions were separated by centrifugation.

Protein detection was performed by Western blotting in protein samples resolved in 10% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) transferred onto a nitrocellulose membrane. After blocking with 5% milk O/N, the nitrocellulose membrane was incubated with mouse monoclonal to *Human Polyomavirus JCV* capsid protein VP1 primary antibody (Abcam, catalog no. ab34756) at a dilution of 1:1,500 in PBS. For detection, a 1:2,000 diluted goat antimouse IgG (H+L)-HRP conjugate antibody (Bio-Rad) was used. Recombinant *hJCV Polyomavirus* Major Capsid VP1 Protein (Abcam, catalog no. ab74569) was used as protein standard in protein quantification experiments; for that, protein samples were loaded containing increasing amounts of recombinant protein: 15, 30, 60, 90, 120, and 240 ng. The protein amount (μg mL^{-1} culture was normalized according to optical density at 550 nm).

Hemagglutination Assays. Detection of red blood cell (RBC) hemagglutination was performed by mixing 0.1 μg of *hJCV* VP1 (obtained from *E. coli* lysates and adjusted to a final volume of 50 μL with PBS) with 50 μL of 0.5% chicken RBCs, kindly provided by Gerard Eduard Martin Valls from CReSA (Universitat Autònoma de Barcelona), in a serial twofold dilution in 96 well plates. Samples were incubated for approximately 48 h at 4°C and then observed. Different buffers were used in hemagglutination assays: Tris 20 mM, 0/40/110/150 mM NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, pH 7.5, and also Tris 20 mM, 200 mM NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, pH 5.8/7.5/9. The hemagglutination titer for each sample has been calculated as the median value from three different plates taking into account that 50 μL (containing 0.1 μg of protein) is the initial volume. Briefly, the individual titer for each point in the dataset corresponds to the highest dilution factor that produced a positive reading. This value is then corrected to mL by a factor of 20 giving hemagglutination (HA) units/mL.

Transmission electron microscopy

E. coli MC4100/pBB535 soluble cell fraction was loaded on 40% sucrose cushion and centrifuged at 100,000*g* at 4°C for 4 h (Beckman SW27 swinging bucket rotor). Resulting pellet was dissolved in Tris 20 mM, 150 mM NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, pH 7.5 buffer. Sample was applied on carbon coated grids and negatively stained with uranyl acetate 2% [w/v] aqueous solution. Photographs were taken with JEM-1400 transmission electron microscope at 250,000 X nominal magnification.

Statistical Analysis. Significance of differences between means of recombinant protein production in the insoluble, soluble, and total cell fractions were evaluated by a Student's t-test. Each experiment was performed in triplicate in intra and inter experiments. Briefly, each data point presented in the graphs corresponds to the results obtained in three independent experiments and each experiment was performed in triplicate. Therefore, nine different measures have been analyzed for each data point.

Results and Discussion

In vitro stabilization of VP1 VLPs produced in wild type E. coli

Production of recombinant *human JCV* VP1 has been described in several expression systems including *E. coli*, yeast, insect cells-baculovirus, and mammalian cells. In all cases, the presence of the VLPs has been demonstrated by using the hemagglutination assay or by transmission electron microscopy.[8,9,12,14] In this study, we wanted to determine the effect of the incubation buffer on the efficiency of VP1 VLPs assembly in cell extracts. For that reason, chicken RBCs were incubated with VP1 VLP containing samples produced in an *E. coli* wild type genetic background under different buffer conditions (Figure 1a). It has been described the importance of the presence of Ca^{2+} and oxidizing conditions in the stability of the SV40 and *hJCV* VLPs. In fact, the formation of the

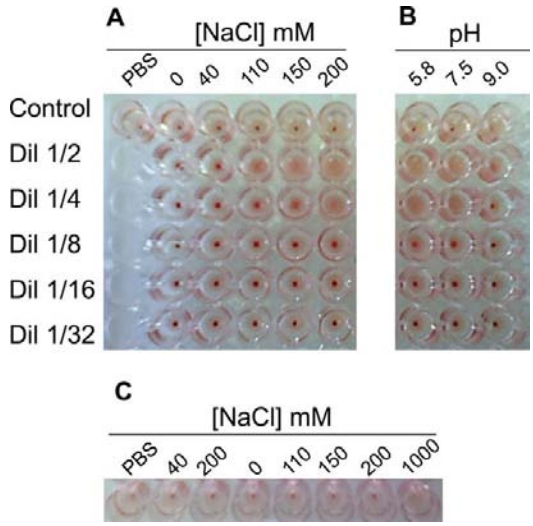


Figure 1. Hemaggutination assays using 0.1 microg of hJCV VP1 VLPs from *E. coli* soluble cell fraction.

Hemagglutination assay under different NaCl concentrations (A), under different pH conditions in 200 mM NaCl in Tris buffer (B). Control wells show the hemagglutination activity of a negative *E. coli* cell extract under different buffering conditions. Effect of NaCl concentration on RBC precipitation in the absence of *E. coli* cell extract (C).

VLPs depends on the interaction of Ca^{2+} within the GH loop and probably on the presence of disulfide bonds between CD loops of neighboring capsomeres.[8] Therefore, a buffer containing Ca^{2+} and Mg^{2+} was selected and different monovalent ion concentration and pH conditions were tested regarding the stability of *hJCV* VP1 VLPs. The results show that the presence of NaCl is a favoring factor, being optimal at 200 mM. At higher concentrations (1 M), a clear interference with the hemagglutination assay was detected (Figure 1b). The dependence in monovalent ions on the assembly of VP1 VLPs has been also demonstrated for SV40 VP1 in insect cell nuclear extracts and it seems to be related to the presence of cellular factors.[27] The influence of pH was also tested showing better hemagglutination performance at pH ranging from 5.8 to 7.5 in the 200 mM NaCl containing buffer (Figure 1c). VP1 VLPs stability was compromised at higher pH values in the same buffer. Consequently, the hemagglutination buffer used in subsequent experiments contained 200 mM NaCl and it was adjusted to pH = 7.5.

hJCV VP1 expression in diverse E. coli DnaK genetic backgrounds

The production of recombinant proteins in heterologous expression systems provokes a stress situation in the cell factory leading, in most cases, to the accumulation of the protein of interest in insoluble structures known as inclusion bodies (IBs). The formation of IBs in expression experiments is enhanced by the limitation of the chaperone/protease system and, therefore, the increase in the amount of chaperones has been used as a strategy to overcome this limitation with the final aim to increase the amount of recombinant protein in the cellular soluble fraction.[28,29] In summary, the results obtained so far by using several chaperone cocktails suggest that, overall; solubility takes priority over conformational quality. Consequently, the more recombinant protein is present in the soluble fraction the less biological activity is retained.[30–32] However, most of the studies performed so far are related to proteins that are not able to form protein complexes and only few of them describe the influence of chaperones in the production of protein complexes as VLPs.[21] Therefore, in order to gather information related to the influence of chaperones on the production of protein complexes, we have studied the effect of the presence of DnaK, a bacterial chaperone of the Hsp70 family on the production of JCV VP1. We transformed wild type *E. coli* MC4100, *E coli* MC4100 DnaK$^-$ (JGT20), and *E. coli* MC4100/pBB535 with an expressing vector coding the VP1 *hJCV* gene and induced its expression under standard conditions. The total amount of recombinant protein was significantly higher in Dnak$^-$ *E. coli* strain, and the presence of DnaK in MC4100 had a profound negative effect on protein yield (Figure 2a). In addition, the fourfold increase in DnaK concentration in MC4100/pBB535 *E. coli* strain[33] did not promote further reduction in protein production. This result might indicate a limited access of overexpressed DnaK to available DnaJ and

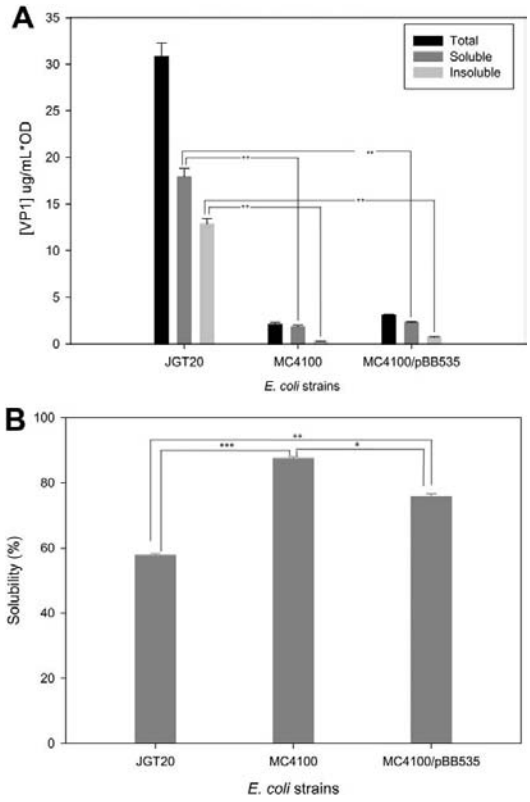


Figure 2. Protein quantification and solubility of hJCV VP1 VLPs.

Protein quantification by Western blot in total, soluble and insoluble cell fractions using commercial VP1 as standrad (A). Relative (*hJCV*) VP1 VLPs presence in soluble cell fractions (B). Asterisks indicate significant differences when comparing with any other data group (*$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$). The amount of protein was normalized according to OD_{550nm}.

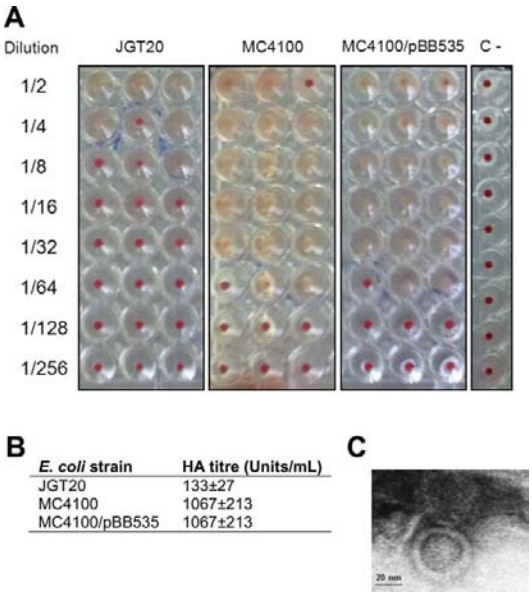


Figure 3. Hemagglutination assay of JCV VP1 containing *E. coli* lysates (A). Negative control wells (C−) correspond to MC4100 *E. coli* strain lysates lacking JCV VP1. Quantification of HA activity of 0.1 μg of *hJCV* VP1 VLPs obtained in different *E. coli* genetic backgrounds (B). Purified JCV VP1 obtained from *E. coli* MC4100/pBB535 induced cell lysates (C).

GrpE cellular levels. However, when soluble and insoluble cellular fractions were individually analyzed, wild type MC4100 and DnaK overexpressing MC4100/pBB535 *E. coli* cells presented higher soluble/insoluble protein ratio than *E. coli* MC4100 DnaK$^-$ (JGT20) (Figure 2b). Under the tested experimental conditions, expression of the chaperone DnaK enhances the solubility of VP1 *hJCV* in the cell although it also induces a reduction in protein yield.

Hemagglutination Assays of Recombinant hJCV VP1 VLPs Produced in E. coli. The effect in biological activity of *hJCV* VP1 VLPs produced in the presence of different amounts of DnaK was determined by analyzing their ability to hemagglutinate RBCs (Figure 3a). In those experiments, the same amount of recombinant *hJCV* VP1 was used and VP1 concentration in cell extracts was calculated by Western blot using standard VP1. The used incubation buffer was the optimized reaction buffer selected in the stabilization experiments described earlier. In those experiments, an *E. coli* cell extract lacking VP1 has been used as negative control. The hemagglutination assay showed that in all three *E. coli* genetic backgrounds, VP1 forms VLPs that can hemagglutinate RBC. In addition, the presence of 42–45 nm particles obtained after ultracentrifugation of *E. coli* MC4100/pBB535 cell extracts confirmed the formation of VLPs under experimental conditions (Figure 3c). However, the presence of DnaK has a positive effect on the ability of VP1 VLPs to hemagglutinate RBC indicating a direct effect of DnaK on the correct assembly of these protein supramolecular assemblies. In addition, it can be seen that when comparing hemagglutination activity on DnaK containing lysates the overexpression of DnaK had no effect (Figure 3b).

Conclusions

Stability of VLPs depends on several factors as oxidative environment and divalent ion concentration among others. It has been demonstrated that pH and monovalent cations are also relevant for the correct assembly of *hJCV* VP1 as they might collaborate in van der Waals interactions.[34]

Cellular protein quality control is accomplished by the coordinated action of chaperones and proteases acting on nascent polypeptides for protein folding and holding, sending folding reluctant intermediates to the proteolysis pathway or accumulating them in dynamic protein-based structures (IBs) for further processing through either folding or proteolysis. The general outcome of this process is improved protein solubility as defined by the portion of protein present in the soluble fraction of the cell. In this work, it has been demonstrated the positive effect on solubility observed in the presence of DnaK chaperone in recombinant *hJCV* VP1 production. In addition, saturation in solubility seems to be reached at the DnaK concentration obtained in wild type *E. coli* genetic background as fourfold higher concentration of DnaK in MC4100/pBB535 is not translated in an increase VP1 solubility (Figure 2a). In addition, the presence of DnaK seems to direct the *hJCV* VP1 to the proteolysis pathway as production of VP1 is dramatically reduced in MC4100 and MC4100/pBB535 *E. coli* strains. Production of VP1 of related polyomavirus simian virus 40 (SV40) in *E. coli* is also dependent on the presence of chaperones although in this case, the presence of DnaK produces an increase in the yield of SV40 VP1.[21] In this study, the different induction conditions used might account for the dissimilar results (2 h at 30°C instead of 4 h at 30°C). Conversely, in eukaryotic expression systems, heat shock cognate protein 70 (hsc70), one of the Hsp70 family member, is involved in the in vivo assembly of polyomavirus VP1 VLPs in the nuclei while in prokaryotes DnaK, one of the Hsp70 family prokaryotic counterparts, recognizes the C-terminus of VP1 and enhances VP1 assembly.[17,21] Therefore, the positive effect on hemagglutination efficiency in DnaK producing *E. coli* strains is in accordance with direct interaction of DnaK with recombinant *hJCV* VP1. This result contrasts with the production of recombinant proteins that do not form supramolecular structures. In this scenario, it has been demonstrated that biological activity is not necessarily favored when gaining solubility. In fact, a wide spectrum of soluble species can be found in the soluble fraction of induced cells.[21,35] This study reveals the positive impact of DnaK on solubility and conformational competence of multiprotein complexes formed by *hJCV* VP1 which give priority to protein yield that is negatively affected in a DnaK-positive genetic background.

Acknowledgments

The authors appreciate the technical support from the UAB Scientific and Technical Services SCAC (Cell Culture Unit: Servei de Cultius Cel·lulars, Producció d'Anticossos i Citometria) and from the Protein Production Platform (PPP, CIBER-BBN and UAB) that belongs to the sePBioEs (Proteomics and Structural Biology: Proteòmica i Biologia Estructural) UAB Scientific and Technical Service. The authors also acknowledge the financial support received for the design and production of VLPs for gene therapy to AV from AGAUR (2009SGR-108) and CIBER de Bioingeniería, Biomateriales y Nanomedicina, 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. PS has received predoctoral fellowship from ISCIII, and AV has been distinguished with an ICREA ACADEMIA award.

Literature Cited

1. Agostini HT, Ryschkewitsch CF, Mory R, Singer EJ, Stoner GL. JC virus (JCV) genotypes in brain tissue from patients with progressive multifocal leukoencephalopathy (PML) and in urine from controls without PML: increased frequency of JCV type 2 in PML. *J Infect Dis*. 1997;176:1–8.
2. Bendiksen S, Rekvig OP, Van GM, Moens U. VP1 DNA sequences of JC and BK viruses detected in urine of systemic lupus erythematosus patients reveal no differences from strains expressed in normal individuals. *J Gen Virol*. 2000;81:2625–2633.
3. Dalianis T, Hirsch HH. Human polyomaviruses in disease and cancer. *Virology*. 2013;437:63–72.
4. Baker TS, Drak J, Bina M. Reconstruction of the 3-dimensional structure of Simian virus-40 and visualization of the chromatin core. *Proc Natl Acad Sci U S A*. 1988;85:422–426.
5. Nilsson J, Miyazaki N, Xing L, Wu BM, Hammar L, Li TC, Takeda N, Miyamura T, Cheng RH. Structure and assembly of a T = 1 virus-like particle in BK polyomavirus. *J Virol*. 2005;79:5337–5345.
6. Liddington RC, Yan Y, Moulai J, Sahli R, Benjamin TL, Harrison SC. Structure of Simian virus-40 at 3.8-A resolution. *Nature* 1991;354:278–284.
7. Rayment I, Baker TS, Caspar DLD, Murakami WT. Polyomavirus capsid structure at 22.5-A resolution. *Nature*. 1982;295:110–115.
8. Chang D, Fung CY, Ou WC, Chao PC, Li SY, Wang M, Huang YL, Tzeng TY, Tsai RT. Self-assembly of the JC virus major capsid protein, VP1, expressed in insect cells. *J Gen Virol*. 1997;78(Pt 6):1435–1439.
9. Goldmann C, Petry H, Frye S, Ast O, Ebitsch S, Jentsch KD, Kaup FJ, Weber F, Trebst C, Nisslein T, Hunsmann G, Weber T, Luke W. Molecular cloning and expression of major structural protein VP1 of the human polyomavirus JC virus: formation of virus-like particles useful for immunological and therapeutic studies. *J Virol*. 1999;73:4465–4469.
10. Goldmann C, Stolte N, Nisslein T, Hunsmann G, Luke W, Petry H. Packaging of small molecules into VP1-virus-like particles of the human polyomavirus JC virus. *J Virol Methods*. 2000;90:85–90.
11. Ou WC, Wang M, Fung CY, Tsai RT, Chao PC, Hseu TH, Chang D. The major capsid protein, VP1, of human JC virus expressed in Escherichia coli is able to self-assemble into a capsid-like particle and deliver exogenous DNA into human kidney cells. *J Gen Virol*. 1999;80(Pt 1):39–46.
12. Sasnauskas K, Bulavaite A, Hale A, Jin L, Knowles WA, Gedvilaite A, Dargeviciute A, Bartkeviciute D, Zvirbliene A, Staniulis J, Brown DW, Ulrich R. Generation of recombinant virus-like particles of human and non-human polyomaviruses in yeast Saccharomyces cerevisiae. *Intervirology*. 2002;45:308–317.
13. Shishido-Hara Y, Hara Y, Larson T, Yasui K, Nagashima K, Stoner GL. Analysis of capsid formation of human polyomavirus JC (Tokyo-1 strain) by a eukaryotic expression system: splicing of late RNAs, translation and nuclear transport of major capsid protein VP1, and capsid assembly. *J Virol*. 2000;74:1840–1853.
14. Shishido Y, Nukuzuma S, Mukaigawa J, Morikawa S, Yasui K, Nagashima K. Assembly of JC virus-like particles in COS7 cells. *J Med Virol*. 1997;51:265–272.
15. Chang CF, Wang M, Ou WC, Chen PL, Shen CH, Lin PY, Fang CY, Chang D. Human JC virus-like particles as a gene delivery vector. *Expert Opin Biol Ther*. 2011;11:1169–1175.
16. Voronkova T, Kazaks A, Ose V, Ozel M, Scherneck S, Pumpens P, Ulrich R. Hamster polyomavirus-derived virus-like particles are able to transfer in vitro encapsidated plasmid DNA to mammalian cells. *Virus Genes*. 2007;34:303–314.
17. Garcea RL, Gissmann L. Virus-like particles as vaccines and vessels for the delivery of small molecules. *Curr Opin Biotechnol*. 2004;15:513–517.
18. Shin YC, Folk WR. Formation of polyomavirus-like particles with different VP1 molecules that bind the urokinase plasminogen activator receptor. *J Virol*. 2003;77:11491–11498.
19. Magaldi TG, Buch MH, Murata H, Erickson KD, Neu U, Garcea RL, Peden K, Stehle T, DiMaio D. Mutations in the GM1 binding site of simian virus 40 VP1 alter receptor usage and cell tropism. *J Virol*. 2012;86:7028–7042.
20. Roy P, Noad R. Virus-like particles as a vaccine delivery system: myths and facts. *Hum Vaccin*. 2008;4:5–12.
21. Wrobel B, Yosef Y, Oppenheim AB, Oppenheim A. Production and purification of SV40 major capsid protein (VP1) in Escherichia coli strains deficient for the GroELS chaperone machine. *J Biotechnol* 2000;84:285–289.
22. Chromy LR, Pipas JM, Garcea RL. Chaperone-mediated in vitro assembly of Polyomavirus capsids. *Proc Natl Acad Sci U S A*. 2003;100:10477–10482.
23. Harrison C. GrpE, a nucleotide exchange factor for DnaK. *Cell Stress Chaperones*. 2003;8:218–224.
24. Schumann W, Ferreira LCS. Production of recombinant proteins in Escherichia coli. *Genet Mol Biol*. 2004;27:442–453.
25. Thomas JG, Baneyx F. Roles of the Escherichia coli small heat shock proteins IbpA and IbpB in thermal stress management: comparison with ClpA, ClpB, and HtpG in vivo. *J Bacteriol*. 1998;180:5165–5172.
26. Tomoyasu T, Mogk A, Langen H, Goloubinoff P, Bukau B. Genetic dissection of the roles of chaperones and proteases in protein folding and degradation in the Escherichia coli cytosol. *Mol Microbiol*. 2001;40:397–413.
27. Mukherjee S, bd-El-Latif M, Bronstein M, Ben-nun-Shaul O, Kler S, Oppenheim A. High cooperativity of the SV40 major capsid protein VP1 in virus assembly. *PLoS One*. 2007;2:e765.
28. Nishihara K, Kanemori M, Yanagi H, Yura T. Overexpression of trigger factor prevents aggregation of recombinant proteins in Escherichia coli. *Appl Environ Microbiol*. 2000;66:884–889.
29. Torres LL, Ferreras ER, Cantero A, Hidalgo A, Berenguer J. Functional expression of a penicillin acylase from the extreme thermophile Thermus thermophilus HB27 in Escherichia coli. *Microb Cell Fact*. 2012;11:105.
30. Garcia-Fruitos E, Martinez-Alonso M, Gonzalez-Montalban N, Valli M, Mattanovich D, Villaverde A. Divergent genetic control of protein solubility and conformational quality in Escherichia coli. *J Mol Biol*. 2007;374:195–205.
31. Martinez-Alonso M, Garcia-Fruitos E, Villaverde A. Yield, solubility and conformational quality of soluble proteins are not simultaneously favored in recombinant Escherichia coli. *Biotechnol Bioeng*. 2008;101:1353–1358.
32. Gonzalez-Montalban N, Garcia-Fruitos E, Villaverde A. Recombinant protein solubility—does more mean better? *Nature Biotechnol*. 2007;25:718–720.
33. Martinez-Alonso M, Vera A, Villaverde A. Role of the chaperone DnaK in protein solubility and conformational quality in inclusion body-forming Escherichia coli cells. *Fems Microbiol Lett*. 2007;273:187–195.
34. Chuan YP, Fan YY, Lua LH, Middelberg AP. Virus assembly occurs following a pH- or Ca2+-triggered switch in the thermodynamic attraction between structural protein capsomeres. *J R Soc Interface*. 2010;7:409–421.
35. 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*. 2008;74:7431–7433.

Manuscript received Feb. 10, 2014, and revision received Feb. 20, 2014.

Article 2

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

Unzueta U*, Saccardo P*, Domingo-Espín J, Cedano J, Conchillo-Solé O, García-Fruitós E, Céspedes MV, Corchero JL, Daura X, Mangues R, Ferrer-Miralles N, Villaverde A, Vázquez E.

Nanomedicine. 2014 Apr;10(3):535-41

*Equal contribution

Gene delivery using self-assembling multidomain protein nanoparticles are still far from the viral efficiency in clinical trial levels. Studying the architectonic properties of spontaneous self-assembling multidomain proteins and their interactions with DNA can furnish precious information for future developing of non-viral gene delivery approaches.

In this work, architectural properties of the paradigm protein R9-GFP-H6 and its interactions with DNA are studied in order to obtain a suitable protein-based artificial virus for gene delivery. It has been observed that in presence of DNA and at slightly acidic pH, R9-GFP-H6 proteins organize in two distinct populations. Microscopy observations showed a supramolecular organization of DNA/nanoparticle complexes, revealing the 9 Arginine and 6 Histidines blocks as promising pleyotropic domains. Moreover, in optimized conditions, R9-GFP-H6 protein has also showed an effective DNA protection against proteases. Finally, we purposed potential structural models of R9-GFP-H6/DNA complexes, based on bioinformatics analysis and experimental data.

BASIC SCIENCE

Nanomedicine: Nanotechnology, Biology, and Medicine
10 (2014) 535–541

Original Article

nanomedjournal.com

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

Ugutz Unzueta, PhD[a,b,c,1], Paolo Saccardo, MS[a,b,c,1], Joan Domingo-Espín, PhD[a,b,c], Juan Cedano, PhD[d], Oscar Conchillo-Solé, MS[a], Elena García-Fruitós, PhD[a,b,c], María Virtudes Céspedes, PhD[c,e], José Luis Corchero, PhD[a,b,c], Xavier Daura, PhD[a,f], Ramón Mangues, PhD[c,e], Neus Ferrer-Miralles, PhD[a,b,c], Antonio Villaverde, PhD[a,b,c,*], Esther Vázquez, PhD[a,b,c,*]

[a]*Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain*
[b]*Department de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain*
[c]*CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, Barcelona, Spain*
[d]*Laboratory of Immunology, Regional Norte, Universidad de la República,, Salto, Uruguay*
[e]*Grup d'Oncogènesi i Antitumorals, Institut de Recerca, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain*
[f]*Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain*

Received 25 September 2013; accepted 11 November 2013

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.

From the Clinical Editor: Finding efficient gene delivery methods still represents a challenge and is one of the bottlenecks to the more widespread application of gene therapy. The findings presented in this paper validate the application of end-terminal cationic peptides as pleiotropic tags in protein building blocks for "viral architecture mimicking" in artificial viruses, representing a promising alternative to the use of viruses and virus-like particles for gene delivery.

Key words: Nanoparticles; Protein building blocks; Self-assembling; Artificial viruses; Gene therapy

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

The authors declare no competing interests.

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.

*Corresponding authors.
E-mail addresses: antoni.villaverde@uab.cat (A. Villaverde), esther.vazquez@uab.cat (E. Vázquez).

[1] Equally contributed.

1549-9634/$ – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.nano.2013.11.006

Please cite this article as: Unzueta U., et al., Sheltering DNA in self-organizing, protein-only nano-shells as artificial viruses for gene delivery. *Nanomedicine: NBM* 2014;10:535-541, http://dx.doi.org/10.1016/j.nano.2013.11.006

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 the 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 were 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 the absence of fixation to prevent internalization artefacts. Nuclei were labeled with 200 ng/ml Hoechst 33342 (Molecular Probes, Eugene, OR, USA) and plasma membranes with 2.5 μg/ml CellMask™ 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 63×/1.4 (oil HC × 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 37 °C, in the 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 70 °C. The remaining DNA was released from protein complexes by adding 10 U of heparin followed by 2-h incubation at 25 °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) analyzer at the wavelength of 633 nm, combined with non-invasive backscatter technology (NIBS) (Zetasizer Nano ZS, Malvern Instruments Limited, Malvern, UK). *Z* potential 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 analyzed by triplicate.

Molecular modeling

To build R9-GFP-H6-based particles, a model of the monomer was first generated using Modeller 9v2[20] and the

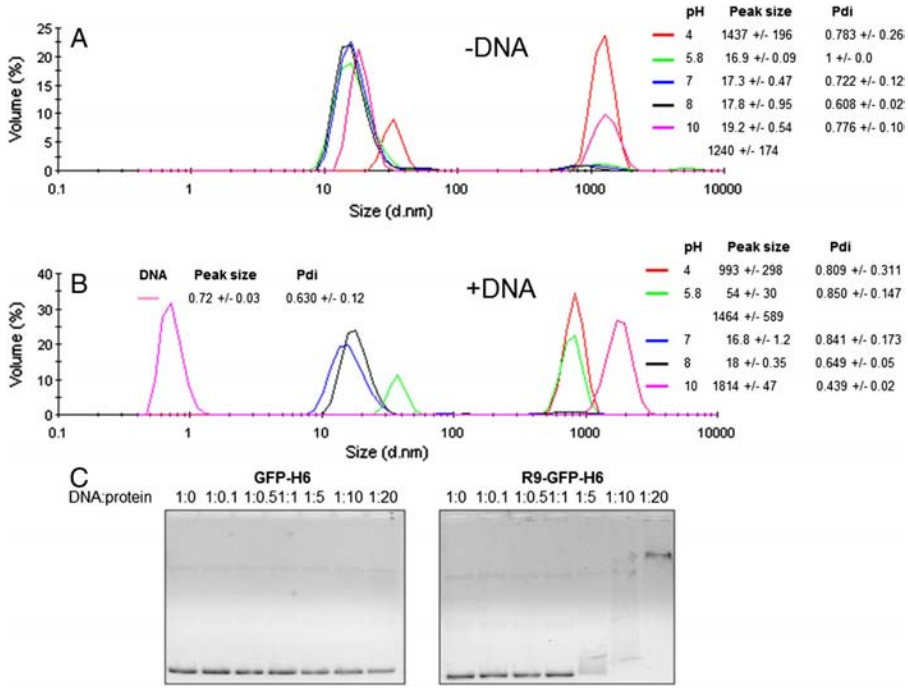


Figure 1. Molecular architecture of R9-GFP-H6-DNA polyplexes. **(A)** Size distribution of R9-GFP-H6 in absence of DNA, at different pH values. Some of the data shown here have been published previously.[15] **(B)** Size distribution of R9-GFP-H6–DNA polyplexes formed at different pH values. DNA alone 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.

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 nine arginines at the N-terminus as active residues and the six 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 were analyzed 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 modeled 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 1, *A*) were incubated with DNA, particle size remained close to 20 nm (Figure 1, *B*), the size previously observed in the 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

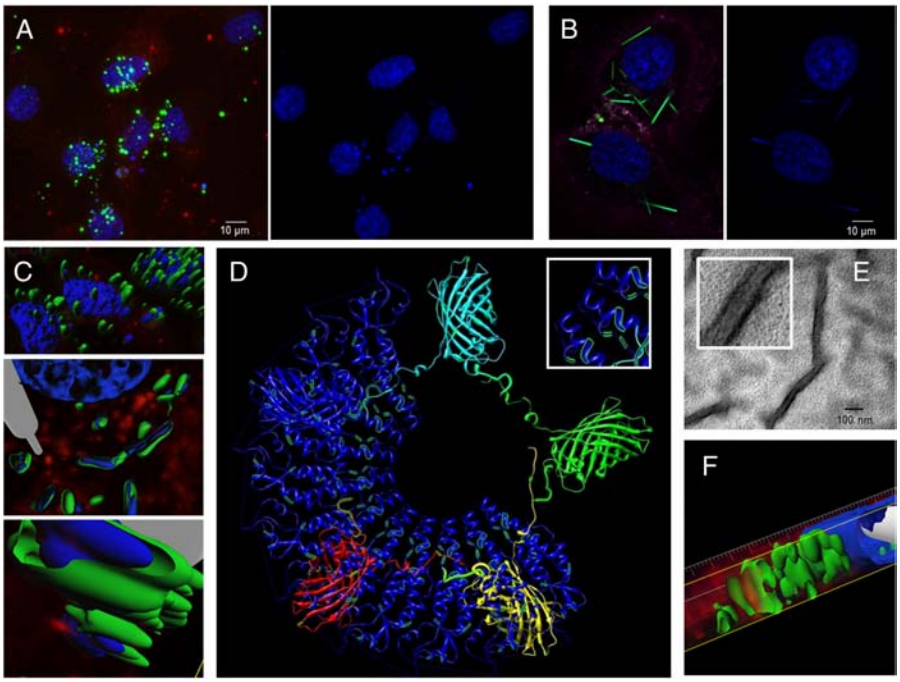


Figure 2. Microscopic analysis of R9-GFP-H6–DNA polyplexes. **(A)** Left: Spherical-shaped green fluorescent signal in HeLa cells exposed for 24 h to R9-GFP-H6–DNA polyplexes. Right: Spherical-shaped blue labels for the same field, corresponding to the embedded DNA. **(B)** Left: Rod-shaped green fluorescent 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.

versions, some slightly twisted or ramified (Figure 2, *B*) were observed, whose size fitted respectively to the two main peaks determined by DLS (Figure 2, *B*). The blue DNA signal appeared coincident with the green label, but its slightly smaller size suggested that DNA occurred in inner cavities of protein entities. Qualitatively, rod-shaped nanoparticles seemed more efficient in embedding DNA than the regular versions, as an important fraction of spheres, but not rods, appeared to be empty (Figure 2, *A* and *B*). Fine confocal sections and 3D isosurface reconstructions strongly suggested that a core DNA was shielded by a solvent-exposed protein layer (Figure 2, *C*), in a virus-like architectonic scheme.

In this regard, rod-shaped forms shown in Figure 2, *A* and *C* strongly evoked the morphologies of capsid proteins observed in plant viruses. Furthermore, a superimposition of the RNA-containing, rod-shaped tobacco mosaic virus (TMV) disk (a structural intermediate in the construction of helical capsids) and an energetically stable, planar, star-shaped molecular model of the self-assembled R9-GFP-H6 at pH 5.8 are presented (Figure 2, *D*), showing coincidence in diameter and in monomer organization. Interestingly, a similar spatial distribution of arginines around the central cavities was found in both viral and non-viral complexes (Figure 2, *D*, inset). TEM images of material deposited on the gird in absence of cells indicated again a prevalence of tubular structures (Figure 2, *E*), with a diameter compatible with the particles observed by confocal analyses (between 20 and 30 nm) and with R9-GFP-H6 disks obtained by molecular modeling (Figure 2, *D*). Importantly, no DNA was found associated to internalized R9-GFP-H6 protein-only nanoparticles (Figure 2, *F*). This indicates that cellular nucleic acids that the protein complexes might eventually find during the intracellular trafficking would result not available for binding, and that the only cargo suitable to form artificial viruses is the nucleic acid loaded in vitro.

Furthermore, DNA embedded in R9-GFP-H6 shells resulted highly protected from DNAse I attack (Figure 3, *A*). This effect was similar to that promoted by the closely related, self-assembling construct T22-GFP-H6. Contrarily, the short modular peptide HNRK,[18,29] which 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

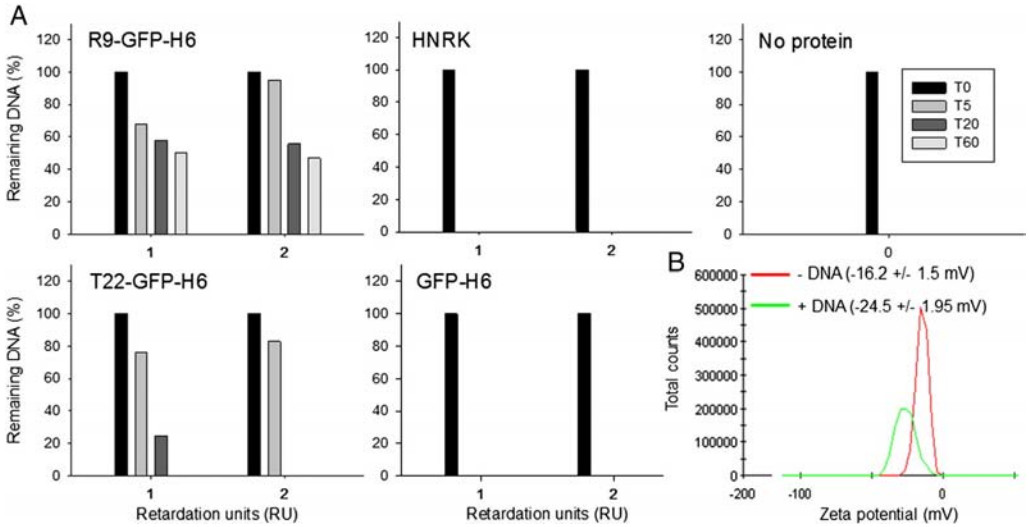


Figure 3. Functional and structural profiling of DNA-loaded nanoparticles. **(A)** Remaining plasmid DNA after treatment with DNAse I, resulting from protection mediated by protein shells at alternative retardation units. Different modular proteins were tested as indicated. At the right, the digestion of protein-free DNA is shown under the same conditions. *T* indicates time of digestion in min. **(B)** Determination of the *Z*-potential of R9-GFP-H6 nanoparticles, with and without DNA.

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 the 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 the 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 intermolecular interactions (Figure 4, *A*). Such pentamers, similarly distributed oligomers (eg hexamers) or 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).

Fig. 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 is 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

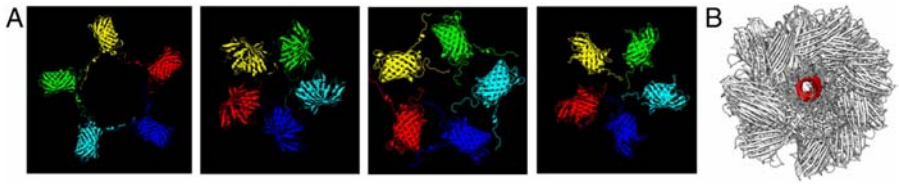


Figure 4. Potential intermolecular contacts in R9-GFP-H6 protein oligomers and in R9-GFP-H6–DNA polyplexes. **(A)** Protein–protein model configurations were obtained by docking simulations using HADDOCK at neutral pH, assuming a pentameric composition that is in agreement with experimental size of protein-only particles. The first model (left) was obtained using R9 residues as active and H6 residues as passive[43] and it was used for the superimposition depicted in Figure 2, *E*. The remaining three models derived from using R9 residues as active and no passive ones. No significant differences in packing were obtained when performing the docking runs at pH 5.8, i.e. with doubly-protonated His (not shown). **(B)** Superposition of the 50 solutions with highest score from a (1:1) DNA–protein docking simulation. The structural fitting is based on the DNA molecule, which is shown in red.

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 the 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 and 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 promote 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, which 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 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 Cellulars 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.
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:W235–9.
24. Martin ACR. ProFit. http://www.bioinf.org.uk/software/profit/index.html.
25. McLachlan AD. Rapid comparison of protein structres. *Acta Crystallogr* 1982;**A38**:871-3.
26. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 1997; **18**:2714-23.
27. Guex N, Diemand A, Peitsch MC. Protein modelling for all. *Trends Biochem Sci* 1999;**24**:364-7.
28. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 2004;**25**:1605-12.
29. Domingo-Espin J, Petegnief V, de VN, et al. RGD-based cell ligands for cell-targeted drug delivery act as potent trophic factors. *Nanomedicine* 2012;**8**:1263-6.
30. Edelstein ML, Abedi MR, Wixon J. Gene therapy clinical trials worldwide to 2007—an update. *J Gene Med* 2007;**9**:833-42.
31. Abbott A. Questions linger about unexplained gene-therapy trial death. *Nat Med* 2006;**12**:597.
32. Williams DA, Medicine Baum C. Gene therapy—new challenges ahead. *Science* 2003;**302**:400-1.
33. Aris A, Villaverde A. Modular protein engineering for non-viral gene therapy. *Trends Biotechnol* 2004;**22**:371-7.
34. Villaverde A. *Nanoparticles in translational science and medicine*. London: 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 Jr S, 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, Díez-Gil C, et al. Protein nanodisk assembling and intracellular trafficking powered by an arginine-rich (R9) peptide. *Nanomedicine (Lond)* 2010;**5**:259-68.

Article 3

Improved performance of protein-based recombinant gene therapy vehicles by tuning downstream procedures.

Unzueta U*, Saccardo P*, Ferrer-Miralles N, García-Fruitós E, Vazquez E, Villaverde A, Cortés F, Mangues R.

Biotechnol Prog. 2013 Nov-Dec;29(6):1458-63.

*Equal contribution

Nucleic acid contaminants in non-viral protein based nanovector production processes can alter the nanocarrier efficiency and cause immunogenic adverse reactions. Nonetheless, a nucleic acid binding activity in the multimodular carrier is required for binding therapeutical genetic material. Enzymatic downstream treatment with nucleases has revealed a good strategy to solve the limitations derived from nucleic acid contamination.

In this work we purpose to study the internalization ability of different T-22 empowered multifunctional proteins and their behavior with nucleic acid contamination after the purification process. DNA binding domain and Nuclear Localization Signal domain were added to original T22-GFP-H6 sequence, to adapt the construct for pDNA targeted delivery. All constructs were described to form nanoparticles and internalize into cell lines by receptor-specific mechanisms. After purification, we observed the presence of nucleic acids in protein only samples, recognized as DNA/RNA mixture contaminants, altering the expected nanoparticle/DNA binding interactions. Finally, we demonstrate that downstream treatment with endonucleases restores the expected nanoparticle's functions.

AIChE

Improved Performance of Protein-Based Recombinant Gene Therapy Vehicles by Tuning Downstream Procedures

Ugutz Unzueta, Paolo Saccardo, Neus Ferrer-Miralles, Elena García-Fruitós, Esther Vazquez, and Antonio Villaverde

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

CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, 08193 Barcelona, Spain

Department de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain

Francisco Cortés

Servei de Cultius Cel·lulars, Producció d'Anticossos i Citometria, (SCAC), Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain

Ramón Mangues

CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, 08193 Barcelona, Spain

Oncogenesis and Antitumor Drug Group, Biomedical Research Institute Sant Pau (IIB-SantPau), Hospital de la Santa Creu i Sant Pau, C/Sant Antoni Maria Claret, 167, 08025 Barcelona, Spain

DOI 10.1002/btpr.1798
Published online September 16, 2013 in Wiley Online Library (wileyonlinelibrary.com)

Protein engineering offers a robust platform for the design and production in cell factories of a plethora of protein-based drugs, including nonviral gene therapy vehicles. We have determined here that a protein nanoparticle, formed by highly cationic protein monomers, fails to bind exogenous DNA and to promote detectable gene expression in target cells despite recruiting all the needed functions. Removal of DNA and RNA with nucleases previous to forming complexes with exogenous DNA dramatically enhances the ability of the protein to bind and transfer DNA to target cell nuclei. These data point out contaminant nucleic acids deriving from the cell factory as a major factor impairing the performance of protein-based artificial viruses and stress the need of a nuclease step in the downstream of proteins whose function is based on cationic domains. © 2013 American Institute of Chemical Engineers *Biotechnol. Prog.*, 29:1458–1463, 2013
Keywords: nanoparticles, nucleic acids, recombinant proteins, gene therapy, downstream

Introduction

Nonviral gene therapy emerges as a safer alternative to virus-based nucleic acid delivery, which despite the recent approval of a few products by different medicine agencies (Oncorine, Gendicine and Glybera) still poses severe biosafety issues.[1,2] The main limitations for nonviral gene therapy are the low transfection efficacy when compared to viral delivery and the transient nature of gene expression. While treating specific conditions might require pulses of gene expression, compatible with the functional profile of nonviral approaches, a consensus exists in that gene transfer and expression levels offered by manmade constructs must be improved in order to raise nonviral gene therapy up to clinical standards.[3–5] Nanotechnologies and material sciences offer principles and tools for the fabrication of tailored vehicles addressed to increase efficacy and to confer specific functions.

In this regard, a spectrum of materials is under exploration for the construction of nano-sized vehicles loadable with nucleic acids. Among them, those based on proteins as building blocks are specially promising, since polypeptides are fully biocompatible and highly versatile.[6] In fact, protein functions can be adjusted by conventional genetic engineering, what offers the possibility to tailor specific activities such as cell-receptor binding and therefore, define biodistribution and establish cell-targeted delivery. Natural protein cages such as virus-like particles (VLPs),[7,8] bacterial microcompartments (BMCs),[9,10] and eukaryotic vaults[11,12] can be produced by recombinant DNA technologies and they have been explored as nanocages to deliver different kind of drugs, including nucleic acids. In addition, multifunctional proteins with modular architecture are especially appealing as diverse functions can be recruited in single polypeptide chains by means of gene fusion.[13–15] This potential allows the resulting construct to mimic the set of biological activities displayed by natural viruses and relevant to gene transfer.[16] Different versions of modular proteins have been proved to be highly promising in the *in vitro* and *in vivo* delivery of therapeutic DNA.[17–19] Also, the fusion of

Ugutz Unzueta and Paolo Saccardo contributed equally to this work.
Correspondence concerning this article should be addressed to: A. Villaverde at antoni.villaverde@uab.cat.

oligomerization domains or shorter architectonic tags permits the self-organization of these hybrid building blocks as nanoparticles of sizes within the viral range,[20,21] altogether permitting the generation of "artificial viruses" that imitate the organization and function of these infectious agents.[22] On the other hand, the cost-effective production of recombinant proteins and the huge spectrum of cell factories available for this purpose offer, in addition, a high versatility regarding biofabrication and downstream processing.[23,24]

Most of the protein constructs intended as components of artificial viruses incorporate cationic stretches as DNA/RNA binding agents.[25] In this study, and by using a family of *de novo* designed, closely related modular building blocks produced in bacteria that self-assemble as nanosized cages, we have determined an unexpected major contamination of bacterial nucleic acids that impairs the gene delivery activities of the resulting artificial viruses. By removing these materials through appropriate nuclease treatments we show dramatic increases in the exogenous DNA binding capacity and in gene expression levels achieved by the nanoparticles upon transfection. Nuclease treatment appears then as a crucial step in the preparation of cationic protein nanoparticles for gene therapy.

Materials and Methods

Protein design, production, and purification

Five chimeric genes encoding different T22-empowered multifunctional constructs were designed in-house and provided by genscript (Piscataway, USA) already subcloned in a pET22b plasmid (Novagen 6744-3) using NdeI/HindIII restriction sites. R9-GFP-H6 protein derivatives (encoded in a pET21b plasmid) containing decreasing number of Arginine residues were also designed and constructed in-house by site directed mutagenesis of parental clone by replacing Arginine residues for Glycines or Alanines to keep the length of the construct constant. All the T22-empowered proteins were produced in *Escherichia coli* Origami B (BL21, OmpT$^-$, Lon-, TrxB$^-$, Gor$^-$, Novagen) overnight at 20°C upon addition of 1 mM IPTG. R9-GFP-H6, R7-GFP-H6, R6-GFP-H6 and R3-GFP-H6 protein constructs were produced in *Escherichia coli* Rosetta BL21 (DE3) overnight at 25°C upon addition of 1 mM IPTG. All the proteins were purified by Histidine tag metal ion affinity chromatography using HiTrap Chelating HP 1 ml columns (GE Healthcare) in an ÄKTA purifier FPLC (GE Healthcare). Cells were disrupted in 20 mM Tris, 500 mM NaCl, 500 mM Imidazole, pH = 8 at 1100 psi by a French press (Thermo FA-078A) and soluble and insoluble fractions separated by centrifugation at 20,000 g for 45 min at 4°C. For comparative purposes, sonication-based cell disruption was also performed with the same buffer at 4°C, through 4x10 min sonication cycles at 40% amplitude each (in a B.Braun Labsonic U). In protein samples intended for nucleic acid removal, an additional step of DNase I and RNase hydrolysis (0.01 μg/μl DNase I, 0.01 μg/μl RNase, 2.5 mM MgCl2, 0.5 mM CaCl2) of soluble extract at 37°C for 1 h was performed. Filtered cell soluble extracts were loaded onto the HiTrap column and then washed with 20 mM Tris, 500 mM NaCl, 10 mM Imidazole, pH = 8 buffer. Proteins were eluted with a linear gradient of Imidazole in the same buffer (20 mM Tris, 500 mM NaCl, 500 mM Imidazole, pH=8) and selected fractions then dialyzed against the buffer at which the proteins are more stable for 2 h at room temperature: 20 mM Tris + 5% Dextrose for T22-KGFP-H6, T22-GFPK-H6, T22-KGFPN-H6, T22-KGFPCmyc-H6, R9-GFP-H6, R3-GFP-H6 and GFP-H6), 20 mM Tris + 500 mM NaCl for T22-GFP-H6, R7-GFP-H6, R6-GFP-H6 and 166 mM NaCO3H + 334 mM NaCl for T22-GFPK-H6 (DNase/RNase), T22-NGFPK-H6 (DNase/RNase). Proteins were then immediately stored at −80°C after 0.22-μm pore membrane filtration. Proteins were characterized by N-terminal sequencing and mass spectrometry (MALDI-TOF) and the amounts determined by Bradford assays.

Dynamic light scattering (DLS)

Volume size distribution of protein nanoparticles were determined by dynamic light scattering at 633 nm (Zetasizer Nano ZS, Malvern Instruments Limited, Malvern, UK).

Cell culture

Sw1417 cells were cultured in DMEM medium (Gibco, Rockville, MD) and HeLa cells in MEM medium (Gibco, Rockville, MD), both supplemented with 10% fetal calf serum (Gibco) and incubated at 37°C in a 5% CO2 humidified atmosphere. Protein nanoparticles were added to cultured cells in presence of Optipro medium (Gibco) 24 h before protein internalization analysis in Sw1417 cells and 48 h before gene expression analysis in HeLa cells. HeLa cell line was obtained from American Type Culture Collection (ATCC, reference CCL-2, Manassas, VA) and Sw1417 cells were a generous gift from Xavier Mayol (Institut Municipal D'Investigacio Médica, Barcelona, Spain).

Protein internalization analysis

Nanoparticle uptake was analyzed by confocal laser scanning microscopy and flow cytometry 24 h after nanoparticles exposure to Sw1417 cells. For confocal analysis cells were grown in MatTek culture dishes (MatTek Corporation, Ashland, MA). The nuclei were labeled with 0.2 μg/ml Hoechst 33342 (Molecular Probes, Eugene, OR) and the plasma membrane with 2.5 μg/ml CellMaskTM Deep Red (Molecular Probes) for 10 min at Room Temperature and then washed in PBS buffer (Sigma-Aldrich Chemie GmbH, Steinheim, Germany). Live cells were recorded by TCS-SP5 confocal laser microscopy (Leica Microsystems, Heidelberg, Germany) using a Plan Apo 63x/1.4 (oil HC x PL APO lambda blue) objective. Hoechst 33342 DNA labels was excited with a blue diode (405 nm) and detected in the 415–460 nm range. GFP-proteins were excited with an Ar laser (488 nm) and detected in the 525–545 nm range. CellMask was excited with a HeNe laser (633 nm) and detected in the 650–775 nm range. For flow cytometry analysis, cell samples were treated with 1 mg/ml Trypsin (Gibco) for 15 min and then analysed on a FACS- Canto system (Becton Dickinson, Franklin Lakes, NJ). Protein fluorescence was excited using a 15 mW air-cooled argon ion laser at 488 nm and detected by a 530/30 nm band pass filter D detector.

Determination of nucleic acids content

Nucleic acid contents within protein samples were determined by Ethidium Bromide staining in agarose gels and by a 200–350 nm absorbance scanning in a UV/visible light spectrophotometer (Genequant 1300, GE Healthcare).

DNA retardation assays

DNA-protein incubation and DNA retardation assays were performed according to previously reported protocols (29).

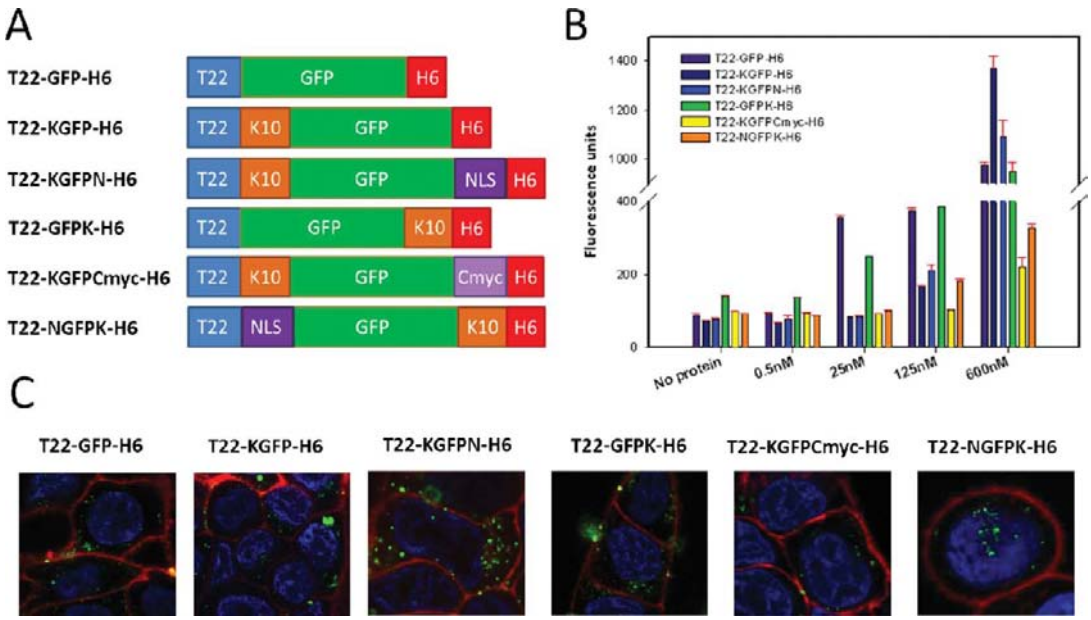


Figure 1. Description of T22-empowered multifunctional modular proteins and their internalization ability in CXCR4$^+$ cells.

A: Schematic representation of T22-empowered constructs. In blue, T22 peptide; in green, GFP; in orange: nucleic acid binding domain; in purple, nuclear localization signal; in red, poly-Histidine tag. NLS indicates the SV40 T antigen nuclear localization peptide. **B:** Dose-response curve of T22-empowered protein constructs internalization in Sw1417 cells. The parental T22-GFP-H6 construct is indicated as a reference. **C:** Confocal images of Sw1417 cells exposed to different T22-empowered multifunctional protein constructs for 24 h. Cell membranes are labeled in red and cell nuclei in blue. Green spots correspond to the fluorescence of internalized nanoparticles.

Cell transfection

For expression experiments, 20 μg of T22-NGFPK-H6 protein (1 retardation unit) and 1 μg of Td Tomato gene containing pCDNA 3.1 plasmid were mixed into a final volume of 60 μl of buffer, and complexes were formed after 1 h at room temperature, after which Optipro was added. The complex was gently added to HeLa cells, followed by incubation for 48 h at 37°C in 5% CO2 atmosphere. TdTomato expression was monitored by flow cytometry and by fluorescence microscopy. As a positive control, 5 μg of polyethyleneimine (PEI) was incubated with 1 μg of Td Tomato gene containing pCDNA 3.1 plasmid (1:5 ratio) in a final volume of 100 μl of Optipro for 15 min at room temperature. Cells without treatment, or just incubated with the expression vector or the protein alone, were used as negative controls. TdTomato and GFP protein fluorescence was detected in no stained cells by fluorescence microscopy (Nikon eclipse TE2000-E) using 465–495 nm laser and 515–555 nm detector for GFP and 528–553 nm laser and 590–650 nm detector for Tdtomato. Red fluorescence in cells was quantified by flow cytometry using a FACS- Canto system (Becton Dickinson, Franklin Lakes, NJ) after detachment with 1 mg/ml Trypsin (Gibco) for 15min. Td tomato protein fluorescence was excited using a 15 mW air-cooled argon ion laser at 488 nm and detected by a 585/42 nm band pass filter.

Data analysis

Mean data, standard deviations, and errors were calculated using Microsoft Office Excel 2003 (Microsoft) and all the graphical representations were done using Sigmaplot 10.0.

Results and Discussion

T22-GFP-H6 is a modular protein monomer that self-assembles spontaneously as nanoparticles of around 13 nm upon purification from producing recombinant bacteria.[20] This protein is stable *in vivo* and targets primary tumor and metastatic foci in colorectal cancer, as the tag T22 promotes internalization into CXCR4+ cells.[26] To adapt this construct to the delivery of therapeutic DNA for cancer therapies we added two additional modules to the polypeptide chain, namely a DNA-binding domain (a decalysine tail, K10) and a nuclear localization signal (NLS, either from SV40 T antigen or from the human C-myc nuclear protein). Different versions of the monomer were constructed that contained one or both additional modules, as summarized in Figure 1A, and produced and purified from *E. coli*. All the proteins were fluorescent and self-assembled as nanoparticles of between 30 and 45 nm (not shown). Internalization analysis of these constructs revealed a slight reduction in the uptake abilities when comparing with the parental construct T22-GFP-H6, which at high doses were not relevant for T22-KGFP-H6, T22-KGFPN-H6, and T22-GFPK-H6 (Figure 1B). All nanoparticles were observed to internalize in CXCR4+ cells, and those containing NLS tags, namely T22-KGFPN-H6, T22-KGFPCmyc-H6, and T22-NGFPK-H6, showed a marked nuclear localization (Figure 1C). When determining the ability of these proteins to bind DNA through gel mobility assays, we surprisingly observed a lack of binding at the tested amounts (Figure 2), which would be not expected for K10-containing polypeptides. However, the high 260/280 absorbance ratio and the staining of protein-only samples in agarose gels (Figure 2, Table 1) were indicative of

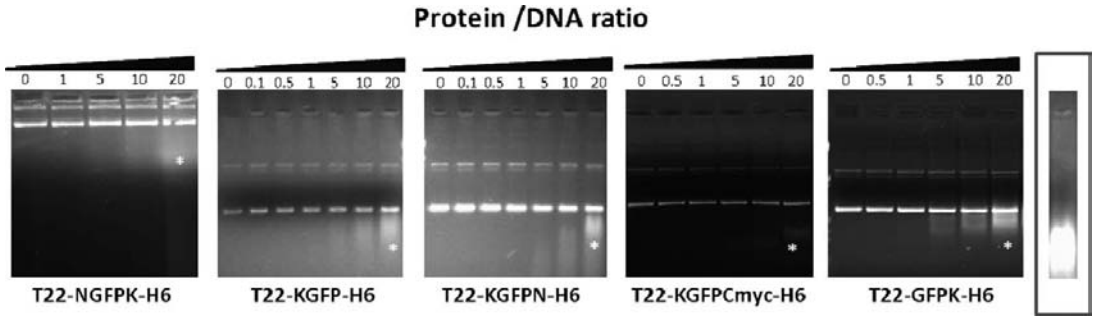


Figure 2. DNA-binding capacity of different protein constructs, monitored by the electrophoretic mobility shift promoted by the proteins on a reporter pDNA (pTurbo FP365).

The nucleic acids detected in a protein-only control (in the rectangle) is also observed in protein/pDNA mixtures at high protein/DNA ratios (*).

Table 1. Ratio Between Absorbance at 260 and 280 nm in Protein Samples, Untreated, and Treated With Nucleases Before Purification

Protein A260/A280	T22-GFP-H6	T22-KGFP-H6	T22-KGFPN-H6	T22-KGFPCmyc-H6	T22-GFPK-H6	T22-NGFPK-H6
No nuclease treatment	0.89	1.54	2.04	1.51	1.38	2.05
Nuclease treatment	nd	nd	nd	nd	0.87	0.67

Nd, not determined.

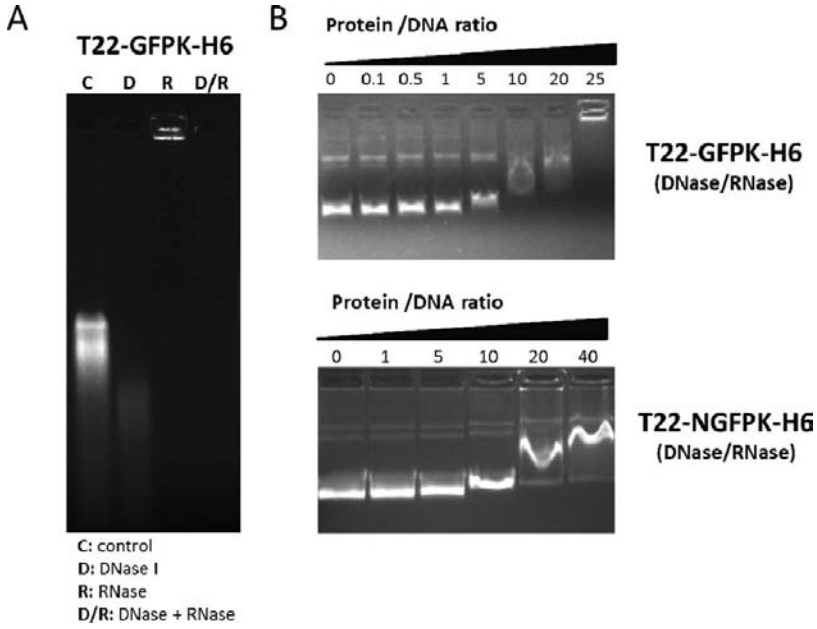


Figure 3. Nucleic acid removal from modular proteins.

A: Nucleic acid removal in T22-GFPK-H6 after DNase or/and RNase digestion. **B:** Evaluation of DNA-binding capacity of nucleic acid free T22-GFPK-H6 and T22-NGFPK-H6 protein constructs monitored by electrophoretic mobility shift assays.

contaminant nucleic acids, probably derived from bacteria, that might interfere in the binding between cationic segments and exogenous DNA. Indeed, treatment with DNAse and RNAse of a model protein indicated the presence of a mixed population of nucleic acids as contaminants of protein samples, among which DNA seemed to be the most prevalent (Figure 3A). A simple combined treatment with both nucleases effectively removed nucleic acids (Table 1) and conferred proteins with the ability to retard exogenous DNA as expected (Figure 3B).

How the nucleic acid removal could enhance the performance of the nanoparticle in transgene delivery was investigated by combining T22-NGFPK-H6 with expressible DNA. When nuclease-treated and nontreated protein versions were compared, no expression of the reporter gene was observed by microscopy neither by flow cytometry, in complexes

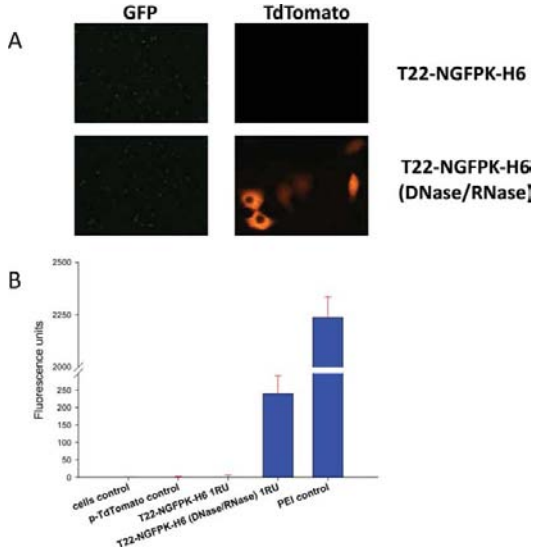


Figure 4. Evaluation of gene transfer properties of nucleic acid free T22-NGFPK-H6 in CXCR4$^+$ cells compared with untreated protein constructs.

A: Fluorescence microscopy images of HeLa cells exposed to T22-NGFPK-H6 / DNA polyplexes for 48 h. Green fluorescence corresponds to GFP and orange fluorescence corresponds to TdTomato protein expressed from the transferred DNA. Fields were selected randomly but images are representative of the whole culture. **B:** TdTomato fluorescence of HeLa cells exposed to T22-NGFPK-H6/DNA polyplexes for 48 h.

formed with nontreated protein samples (Figures 4A and B). However, the nuclease-treated vehicle promoted transgene expression in a significant percentage of cells (Figure 4A), and gene expression levels were clearly over the background provided by nontreated samples (Figure 4B).

Biofabrication of proteins as convenient carrier materials for nonviral gene therapy benefits from the advances of recombinant DNA technologies accumulated in the last 30 years. Many protein products are then used as pharmaceuticals with great success[23] and an important sector of Pharma industries orbits around recombinant protein design and production. Cationic peptides or protein domains are commonly used as functional components of artificial viruses,[25] and in protein only vehicles they are usually incorporated as part of multifunctional proteins.[15] DNA condensation by short multifunctional proteins might have a structural role in the formation of protein/DNA nanoparticles,[19] while the incorporation of cationic end terminal peptides to more complex building blocks drives their self-assembling as stable nanoparticles[20] into which exogenous DNA is smoothly accommodated.[18] We have here constructed a series of modular building blocks in which several cationic peptides are combined to offer both architectonic abilities at the nanoscale and DNA condensing properties (Figure 1). T22-NGFPK-H6, for instance, contains the highly cationic T22, K10, and the protonated form of H6 (at slightly acidic pH). We observed that such a high concentration of cationic elements in the building block eclipsed its expected ability to bind and transfer DNA (Figures 2 and 4), a fact that was unapparent (and possibly milder) in other K10-containing multifunctional proteins,[19,27–29] probably less cationic in global. In agreement, Rn-GFP-H6 protein versions in which the number of N-terminal cationic residues

Table 2. Ratio Between Absorbance at 260 and 280 nm in Samples of Different Versions of R9-GFP-H6, in Which the Number of N-Terminal Arginines Varies

Protein	R9-GFP-H6	R7-GFP-H6	R6-GFP-H6	R3-GFP-H6	GFP-H6
A260/A280	1.60*	1.38	0.65	0.71	0.67

*For comparative purposes, the nucleic acid content of this protein was also determined upon purification from cell samples disrupted by ultrasonication, and the 260/280 ratio was found to be 1.61 (not shown).

(Arginines) was engineered showed correlative amounts of attached nucleic acids (Table 2). Interestingly, although it could not be discarded that the chosen downstream procedure might influence the nucleic acid content of the samples, and it even could potentially minimize the need for nucleases, the use of an alternative cell disruption method (ultrasonication) does not show any observable effect on the 260/280 ratio (Table 2). Also, the fact that nucleic acid contamination is persistent even in the high salt buffers used in our protocols (usually 500 mM), is in agreement with the previous suggestion that particularly high cationic constructs such as those generated in this study show an intrinsic functional limitation associate to contaminating nucleic acids. The expected functions of the resulting nanoparticles were, however, restored by a simple nuclease digestion step previous to purification from bacteria (Figure 3). These data indicates the convenience of including this step in the bioproduction of proteins as building blocks of artificial viruses, when their function is at least partially based on cationic, DNA-binding domains. Although regarding biosafety, contamination with nucleic acids is a particular issue in protein drugs produced in mammalian cells,[30] the particular use of cationic proteins as DNA condensing agents stresses the need of surveillance and efficient removal treatment, for functional reasons, in any type of cell factory.

Acknowledgments

The authors appreciate the support of the Cell Culture Unit of Servei de Cultius Cel·lulars Producció d'Anticossos i Citometria (SCAC, UAB) and of the Protein Production Platform (http://www.bbn.ciber-bbn.es/programas/plataformas/equipamiento). We are also grateful to Maria Carme Fàbrega from IRB for her helpful comments. The authors also acknowledge the financial support from FIS (PI12/00327 to E.V. and PI12/01861 to R.M.) and from Agència de Gestió d'Ajuts Universitaris i de Recerca (2009SGR-108 to AV and 2009-SGR-1437 to R.M), 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 a PhD fellowship from ISCIII. A.V. has been distinguished with an ICREA ACADEMIA Award. The authors declare no conflict of interest.

Literature Cited

1. Guo J, Xin H. Chinese gene therapy. Splicing out the West? *Science* 2006;314:1232–1235.
2. Edelstein ML, Abedi MR, Wixon J. Gene therapy clinical trials worldwide to 2007—An update. *J Gene Med* 2007;9:833–842.
3. Li S, Huang L. Nonviral gene therapy: Promises and challenges. *Gene Ther* 2000;7:31–34.
4. Wiethoff CM, Middaugh CR. Barriers to nonviral gene delivery. *J Pharm Sci* 2003;92:203–217.

5. Jafari M, Soltani M, Naahidi S, Karunaratne DN, Chen P. Non-viral approach for targeted nucleic acid delivery. *Curr Med Chem* 2012;19:197–208.
6. Vazquez E, Villaverde A. Engineering building blocks for self-assembling protein nanoparticles. *Microb Cell Fact* 2010;9:101.
7. Petry H, Goldmann C, Ast O, Luke W. The use of virus-like particles for gene transfer. *Curr Opin Mol Ther* 2003;5:524–528.
8. Georgens C, Weyermann J, Zimmer A. Recombinant virus like particles as drug delivery system. *Curr Pharm Biotechnol* 2005; 6:49–55.
9. Cheng S, Liu Y, Crowley CS, Yeates TO, Bobik TA. Bacterial microcompartments: Their properties and paradoxes. *Bioessays* 2008;30:1084–1095.
10. Fan C, Cheng S, Liu Y, Escobar CM, Crowley CS, Jefferson RE, Yeates TO, Bobik TA. Short N-terminal sequences package proteins into bacterial microcompartments. *Proc Natl Acad Sci USA* 2010;107:7509–7514.
11. Han M, Kickhoefer VA, Nemerow GR, Rome LH. Targeted vault nanoparticles engineered with an endosomolytic peptide deliver biomolecules to the cytoplasm. *ACS Nano* 2011;5:6128–6137.
12. Rome LH, Kickhoefer VA. Development of the vault particle as a platform technology. *ACS Nano* 2012;7:889–902.
13. Vazquez E, Ferrer-Miralles N, Villaverde A. Peptide-assisted traffic engineering for nonviral gene therapy. *Drug Discov Today* 2008;13:1067–1074.
14. Ferrer-Miralles N, Vazquez E, Villaverde A. Membrane-active peptides for non-viral gene therapy: making the safest easier. *Trends Biotechnol* 2008;26:267–275.
15. 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.
16. Aris A, Villaverde A. Modular protein engineering for non-viral gene therapy. *Trends Biotechnol* 2004;22:371–377.
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* 2006;7:35.
18. Vazquez E, Roldán M, Diez-Gil C, Unzueta U, Domingo-Espín J, Cedano J, Conchillo O, Ratera I, Veciana J, Daura X, Ferrer-Miralles N, Villaverde A. Protein nanodisk assembling and intracellular trafficking powered by an arginine-rich (R9) peptide. *Nanomedicine (Lond)* 2010;5:259–268.
19. Domingo-Espín J, Vazquez E, Ganz J, Conchillo O, García-Fruitós E, Cedano J, Unzueta U, Petegnief V, Gonzalez-Montalbán N, Planas AM, Daura X, Peluffo H, Ferrer-Miralles N, Villaverde A. Nanoparticulate architecture of protein-based artificial viruses is supported by protein-DNA interactions. *Nanomedicine (Lond)* 2011;6:1047–1061.
20. Unzueta U, Ferrer-Miralles N, Cedano J, Zikung X, Pesarrodona M, Saccardo P, García-Fruitós E, Domingo-Espín J, Kumar P, Gupta KC, Mangues R, Villaverde A, Vazquez E. Non-amyloidogenic peptide tags for the regulatable self-assembling of protein-only nanoparticles. *Biomaterials* 2012;33:8714–8722.
21. Yang Y, Burkhard P. Encapsulation of gold nanoparticles into self-assembling protein nanoparticles. *J Nanobiotechnology* 2012;10:42.
22. Mastrobattista E, van der Aa MA, Hennink WE, Crommelin DJ. Artificial viruses: A nanotechnological approach to gene delivery. *Nat Rev Drug Discov* 2006;5:115–121.
23. Ferrer-Miralles N, Domingo-Espin J, Corchero JL, Vazquez E, Villaverde A. Microbial factories for recombinant pharmaceuticals. *Microb Cell Fact* 2009;8:17.
24. Corchero JL, Gasser B, Resina D, Smith W, Parrilli E, Vázquez F, Abasolo I, Giuliani M, Jäntti J, Ferrer P, Saloheimo M, Mattanovich D, Schwartz S Jr, Tutino ML, Villaverde A. Unconventional microbial systems for the cost-efficient production of high-quality protein therapeutics. *Biotechnol Adv* 2013; 31:140–153.
25. Saccardo P, Villaverde A, Gonzalez-Montalban N. Peptide-mediated DNA condensation for non-viral gene therapy. *Biotechnol Adv* 2009;27:432–438.
26. Unzueta U, Cespedes MV, Ferrer-Miralles N, et al. Intracellular $CXCR4^+$ cell targeting with T22-empowered protein-only nanoparticles. *Int J Nanomedicine* 2012;7:4533–4544.
27. 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* 2000;68:689–696.
28. Aris A, Villaverde A. Molecular organization of protein-DNA complexes for cell-targeted DNA delivery. *Biochem Biophys Res Commun* 2000;278:455–461.
29. Aris A, Villaverde A. Engineering nuclear localization signals in modular protein vehicles for gene therapy. *Biochem Biophys Res Commun* 2003;304:625–631.
30. Demain AL, Vaishnav P. Production of recombinant proteins by microbes and higher organisms. *Biotechnol Adv* 2009;27:297–306.

Manuscript received May 10, 2013, and revision received Jul. 16, 2013.

Discussion

One of the greatest technical challenges of modern medicine is to introduce new genes into human cells aiming to cure or ameliorate genetic disorders. Unfortunately, despite decades of best efforts spent by researchers worldwide, gene therapy has showed only limited success in treating diseases[253]. Since efficient and targeted nucleic acid delivery is crucial for the therapy success, gene delivery vectors, then, should be developed to overcome extracellular and intracellular barriers, and, above all, to avoid adverse reactions that can be fatal for the patient[254]. Moreover, economic efforts spent in research and development should be reflected in the commercialization of an appealing and convenient product. By now, several nuclei acid carriers based on viral approaches have been developed but, despite their high efficiency, adverse immunological reactions and biosafety issues in viral production, handling and administration are slowing down the development of safe and commercially attractive therapies[255]. As an alternative to the viral approaches, engineered protein-based artificial viruses or viral mimetics are being explored as gene delivery vehicles. Versatility in protein structure and function makes them fully tunable and biocompatible vectors. It has been observed that cooperative processes involving different functional building blocks can lead modular proteins to their assembly in nanoscale ordered artificial viruses able to accommodate nucleic acid and protect them from nucleases (**Article 2**). Interestingly, recombinant proteins show some other great advantages over viruses, allowing recovery of gene therapy vectors from a wide range of cell factories and permitting an easier scalability of processes and avoiding all the biosafety concerns derived from contaminants of viral origin[256]. Production process then, became a key issue in which a great number of variables can be modified. At the cellular level, protein-folding modulators play an important role in recombinant protein production. In fact, these molecular tools help nascent polypeptides to fold in an active three-dimensional structure or send them to the proteasome complex[257]. Tuning the cellular protein quality control system, then, could have strong effects on protein production yield and conformational quality and, moreover, adjusting downstream procedures can also improve vehicle performance, showing the high level of versatility and customization of the production system.

Exploring molecular folding modulators in Virus-Like Particles production

E.coli

Virus-like particles (VLPs) are one of the most studied protein-only artificial viruses in non-viral gene therapy. It has been described that *human JC polyomavirus* VP1 major capsid protein is able to self-assemble in VLPs when expressed in several protein factories such as *E. coli*[258], yeast[229] and insect cells-baculovirus[259]. Interestingly, VP1 VLPs retain the typical icosaedrical (T7) ordered structure of wild type virus but lack immunogenicity[259]. Moreover *JC virus* VP1 VLPs, devoid of the viral genome, do not cause the demyelinating disease of central nervous system known as Progressive Multifocal Leuconcefalopaty in immunodepressed human patients[260] as the *JC virus* causes, and are able to encapsidate and deliver nucleic acids to target cells[261]. VP1 VLP formation is favored by divalent ions as Ca^{2+} or Mg^{2+} that stabilize Van der Waals interactions within the GH exposed loop and by disulfide bonds formed between structural proteins[262]. In order to establish which are the effects of different buffers in VP1 VLP produced in *E. coli* we tested a wide set of buffers, starting from a base formulation of Tris 1 mM $CaCl_2$ and $MgCl_2$ pH 7.5 and changing the molar concentration of NaCl and pH values. VLPs structures were detected by red blood cell hemagglutination assay (HA) in progressive dilution in order to reveal the best combination of salt concentration and pH for VLPs assembly. HA assay reveals the best hemagglutination activity at 200 mM (**Article 1 Fig. 1a**). Within different pH values at 200 mM NaCl, pH 7.5 showed the best hemagglutination activity, setting the optimal buffer composition in Tris 1 mM $CaCl_2$, 1 mM $MgCl_2$, 200 mM Nacl, pH 7.5. The presence of monovalent ions have already being described as stabilizing factors for other polyoma viruses VLPs like VP1 SV40 VLPs and their assembling favoring action seems to be necessary for VLP formation since no hemagglutination activity was observed at low salt concentration[263] (**Article 1 Fig. 1a**). Once established the optimal buffer composition for VLP formation, three different *E. coli* DnaK genetic backgrounds were used as VP1 protein factories. The member of Hsp70 family DnaK protein is a key molecular chaperone in the protein quality control system in Prokaryotes. It takes part in the first step of protein synthesis, helping newly synthesized proteins folding and preventing the formation of inclusion bodies[264]. In order to study the effects of DnaK in the formation of protein complexes, we analyzed VP1 protein production in wild type MC4100 *E. coli* strain, in the DnaK defective *E. coli* strain JGT20 and in the over expressing Dnak *E. coli* strain MC4100/pBB535. Analysis on soluble/insoluble/total fractions of VP1 produced in the three different strains revealed that the total amount of recombinant protein produced

was notably higher in DnaK defective strain JGT20. On the other hand, the effects of DnaK in wild type MC4100 strain and in DnaK overexpressing strain revealed an equal profound negative effect on VP1 yield. Interestingly, separate analysis of soluble and insoluble fractions revealed a higher soluble/insoluble protein ratio levels in Dnak wild type and DnaK overexpressing strains (**Article 1 Fig. 2b**). These data showed that the over-expression of DnaK chaperone does not alleviate the negative effects in total protein yield, probably due to a limited access to DnaJ and GrpE chaperone proteins, nonetheless the expression of DnaK chaperone in MC4100 and MC4100/pBB535 has significant positive effects in VP1 solubility.

Since we were interested in exploring the conformational quality of VLPs in different chaperone genetic backgrounds, we analyzed the three different soluble fraction samples with a hemagglutination assay (**Article 1 Fig. 3a**). Using the same amount of VP1 protein in the optimized incubation buffer described above, HA assay reveals that VP1 can form VLPs structures when produced in all of the three different *E. coli* strains. However samples produced in DnaK positive strains showed a higher hemagglutination activity, indicating a positive effect of DnaK chaperone in the formation of VLPs supramolecular organization. Nonetheless, over-expression of DnaK chaperone does not increase the hemagglutination activity of the samples more than the VP1 VLPs produced in wild type conditions (**Article 1 Fig. 3b**). VLPs structures were visualized by transmission electron microscopy (TEM) after a 40 % sucrose cushion ultracentrifugation showing the typical 42-45 nm icosahedral rounded shape conformation nanoparticles (**Article 1 Fig. 3c**).

Taking all these observations, we demonstrate the positive effect of DnaK Hsp70 chaperone protein in the production of VP1 *hJC virus* VP1 in the *E. coli* cell factory. Moreover, the four-time over-expression of DnaK in MC4100/pBB535 strain does not reveal any increasing in VP1 solubility if compared with wild type conditions, suggesting a top limit in VP1 solubility at DnaK wild type expression levels.

In addition, it seems that Dnak drives the recombinant expressed VP1 proteins through the proteolysis pathway, since the total yield is strongly compromised in DnaK$^+$ strains in comparison with JGT20 non-expressing DnaK strain. We also proved that VP1 VLPs quality structural conformation is positively influenced by the presence of Dnak chaperone in the cell production environment with a ceiling at wild type level conditions.

Insect cells

As mentioned above, *hJC virus* VP1 VLPs can be successfully produced also in Insect cells through the baculovirus-insect cells expression system (BES)[259], providing an optimized molecular environment for eukaryotic recombinant protein production. After exploring the VP1 VLPs conformational quality in *E. coli*, and being chaperones responsible of protein folding and regulated degradation pathway, we wanted to investigate the re-hosting of bacterial Hsp70 DnaK and Hsp40 DnaJ chaperones in *Spodoptera frugiperda* (*Sf9*) insect cell factory, and the effects on VP1 VLPs production yield and conformational quality. As first step, a successful recovery and amplification of engineered Baculovirus vector harboring *hJC virus* VP1 protein under polyhedrin promoter, was achieved. Time course assay allowed establishing a multiplicity of infection (M.O.I) of 1 and a time of infection (T.O.I.) of 96 h as the best condition for VP1 production in the soluble cell fraction (**Annex 1 Fig. 1**). In order to re-host the Dnak and DnaJ chaperones, we first infect *Sf9* cells with the engineered baculovirus expressing both chaperones under the control of the polyhedrin and p10 promoters, respectively. Western blot analysis revealed a correct production of both proteins (**Annex 1 Fig. 2a**). Subsequently, *Sf9* co-infection with optimal M.O.I. of 0.1 for DnaK/DnaJ baculoviruses and M.O.I. 1 for VP1 expressing baculoviruses was performed, as well as a simple Sf9 infection with M.O.I 1 for VP1 expressing baculoviruses. The total amount of VP1 protein obtained in the different cell culture supernatants was nearly two fold higher for VP1 produced in absence of DnaK/DnaJ folding modulators (**Annex 1 Fig. 2b**). Beyond the effects of chaperones' activity, the negative effects of protein folding modulators on VP1 co-expression could be due to different phenomena: to the negative interference of p10 promoter over polyhedrin promoter activity in case of multiple expression genes or maybe due to the co-infection approach, which can cause a competition between baculovirus vectors for the infection process. Moreover, in co-infection processes, optimal M.O.I. and T.O.I. of different baculovirus constructs can produce variation on protein productions. In this work a higher M.O.I. for VP1 baculovirus revealed a decrease in protein production yield[84].

Both VP1 production supernatant samples were purified and concentrated by 40 % sucrose cushion ultracentrifugation (**Annex 1 Fig. 3**). Transmission electron microscopy observation revealed the presence of VP1 VLPs structures in both samples (**Annex 1 Fig. 4**). Conformational quality analysis of VP1 VLPs produced with or without Dnak/DnaJ support was performed once again by hemagglutination assay. In addition, we wanted to test the control of VLPs assembling/disassembling by changing the buffer composition. The HA assay revealed the presence of VP1 VLPs at 72 h and 96 h post infection in Sf9 supernatant samples showing a higher HA activity at 96 h p.i. (**Annex 1**

Fig. 5a). Purified and concentrated VP1 VLP samples showed higher HA activity, as expected (**Annex 1 Fig. 5b**). Hemagglutination activity of VLPs produced with or without Dnak/DnaJ co-expression was subsequently tested under three different conformational states: originally assembled, opened and re-assembled. Control on VLPs open/close state is achieved removing divalent ions from buffer with chelating molecules as EDTA and reducing agents as DTT. Restoring the original optimal buffer composition is observed to favoring the re-assembling of VLPs[247]. Hemagglutination assay revealed a higher hemagglutination activity of VP1 VLPs produced with DnaK/DnaJ, showing a higher conformational quality even if the VP1 yield is compromised (**Annex 1 Fig. 6**). The achieved control on open/close state of VLPs was revealed not only by the HA profile but also with observations through TEM (**Annex 1 Fig. 7**). Being the low fraction of correctly assembled protein one of the main problems in multiprotein complex production, we demonstrated the positive effects on biological activity and structural conformation in *hJCV* VP1 VLPs, co-produced with bacterial DnaK/DnaJ rehosted in Sf9 insect cell expression system. On the contrary, negative effects are reported for recombinant protein production yield.

Interestingly, the same approach applied by Martínez-Alonso and collaborators for mGFP, foot-and-mouth disease virus (FMDV) VP1 and VP2 capsid proteins and human alpha-galactosidase gave positive effects not only in protein quality but also in production yield confirming that the effects on protein solubility or structural conformation may vary depending on cell factory or recombinant protein itself[265-267]. When comparing results, a factor to take into account is the different approach adopted in Martínez-Alonso and collaborators work, in which a single baculovirus clone produced all recombinant proteins. In our case, co-infecting in the same cells batch with multiple viruses at different M.O.I. could lead to different subpopulations[84], affecting the recombinant protein expression, even thought the co-infection approach for recombinant self-assembling protein production has already been reported [268].

Exploring DNA sheltering and its architectonic role in modular protein nanoparticles

An alternative to VLPs for protein only vectors generation is the *de novo* design of multifunctional proteins, able to show viral properties. The main drawback of this strategy is due to the failure of engineered multifunctional proteins to promote nanoscale organization. If in VLPs the self-assembling capacity is intimately related to their nature, only few cases of successful production of self-assembling multidomain protein vectors have been described[269]. Generally, in non-assembling proteins, it has been observed that interactions between cationic peptide residues and DNA can strongly affect supramolecular structure and size, leading to unordered protein-protein interactions and generating polydisperse soluble aggregates[208,219]. Nonetheless, intermolecular protein-protein and protein-DNA interactions and their relation with the supramolecular organization of multifunctional nanoparticles it's so far to being elucidated. However, the formation of some monodisperse protein nanoparticles, stabilized by DNA molecules has been described[270] (**Article 2**). In a previous work[271] we recently reported the self-assembling of pleiotropic R9 poly-arginine peptide, fused with a hexa-histidine (H6) tagged Green Fluorescent Protein (GFP) domain. Interestingly these R9-GFP-H6 nanoparticles of 20 nm in size were able to condense DNA and showed high cell transfection efficiency at pH 5.8. Those studies permitted to define R9, or more generally a cationic tag, and the H6 tag as structural architectonics domains for self-assembling protein platform. Subsequently, R9 peptide, which is a Cell Penetrating Peptide (CPP), was substituted by an engineered segment derivative of polyphemusin II from the horseshoe crab, named T22 peptide with a high affinity for CXCR4 receptor. Purified T22-GFP-H6 protein was proven to self-assemble in nanoparticles and was efficiently internalized in $CXCR4^+$ cell line[217].

In this context, we wanted to explore the nanoparticle-DNA interactions in R9-GFP-H6 multidomain model self-assembling protein, especially focusing on how functionality is affected by supramolecular organization. With this aim, R9-GFP-H6 protein supermolecular structures were analyzed by dynamic light scattering (DLS) after protein incubation with an external cargo DNA at different pH conditions. Probably due to the strong denaturing conditions, protein-DNA complexes generated at pH 4 and pH 10 showed a severe tendency to aggregate. At pH 7 and pH 8, samples do not showed any changes from the complexes formed by protein only incubation. Interestingly, at pH 5.8, proteins complexes split into two different populations of 38 nm and 700-800 nm (**Article 2 Fig. 1a**). Being the pH 5.8 also the optimal condition for best internalization of R9-GFP-H6, these complexes were exposed to cultured cells and observed by confocal microscopy. As detected by DLS measurement, we found two different populations of

particles: the first one, more abundant, with small spherical shape (**Article 2 Fig. 1b-c**) and the second one with larger rod-shaped morphology (**Article 2 Fig. 1d-e**) strongly resembling viral structures. Both populations were found within the cells, but DNA staining revealed nucleic acid presence principally in co-localization with rod-shaped protein structures, suggesting a less efficiency of small spherical complexes to embed DNA. Interestingly, *in silico* representation of R9-GFP-H6 nanoparticles generated at pH 5.8 superimposed to the one of RNA-containing rod-shaped tobacco mosaic virus (TMV) structure, showed a diameter, distribution and organization of arginine residues, matching with the inner part of the central cavity of the viral structure (**Article 2 Fig. 1f**). These observations were further integrated with DNAse I assay in order to describe the R9-GFP-H6 DNA complex protection against nucleases. Results showed high protection efficiency up to two hours post reaction. T22-GFP-H6 nanoparticles were also able to protect DNA, even with slight decrease of efficiency. The lower DNA binding affinity showed by T22-GFP-H6, can be an explanation of the smaller protection observed. On the other hand, polyplexes of the modular protein HNRK and DNA (21651444), challenged with nucleases, do not show any DNA protection activity (**Article 2 Fig. 2**). This observation is in complete agreement with previous HNRK-DNA supramolecular structure characterization, which reported the HNRK lacking of architectonic properties and, moreover, the DNA overhanging from cargo polyplexes[270] (**Fig. 28**).

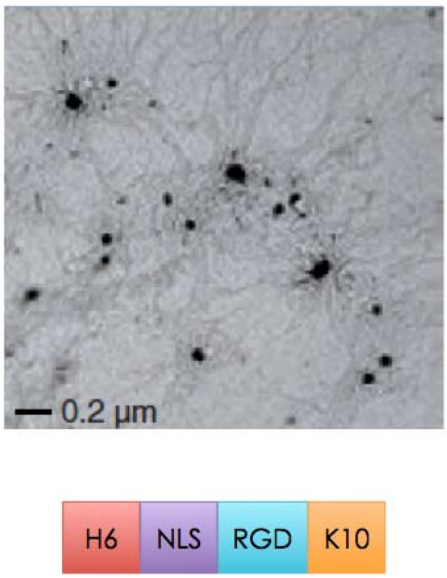


Fig. 28 TEM images of HNRK as complexes with plasmid pcDNA3.1 and schematic representation of HNRK domains order and composition. Adapted from [270].

Even if it's not completely clear why R9-GFP-H6 proteins incubated with DNA at pH 5.8 organize in spherical shaped and rod-shaped structures, we can reason about the protonation stage of pleiotropic domains. Having the histidine imidazole group a

pKa=6, we can assume that at the optimal internalization value of pH 5.8 they are in a protonation equilibrium, which could confer to the nanoparticles a certain grade of elasticity and dynamism in their organization, enabling the formation of both morphological distinct structures. For the generation of stable protein-protein and protein-DNA interactions and the subsequently formation of ordered arrangements, the most feasible morphology configuration is the one corresponding to an inner part with orientated positively charged arginine residues. In this way, DNA accommodation would be favored by surrounding protein complexes, as supported by the Z-potential analysis that reported a higher negative charge in the outside surface of protein-DNA polyplexes than in protein nanoparticles alone (**Article 2 Fig. 2b**).

Starting from the observation reported above and taking into account the whole combination of electrostatic interactions[272], Van der Waals forces and hydrogen bonds, a model based on intermolecular interactions for both spherical-shaped and rod-shaped stables structures was generated *in silico* (**Article 2 Fig. 3**).

With all these data, we demonstrated that functionalized self-assembling protein nanoparticles, combined with external cargo DNA, show a clear tendency to assume an ordered supramolecular organization able to protect exogenous DNA from protease challenge. Moreover, being nanoparticle morphology resembling viral structures, we propose a reasonable structural model following charge distribution generated by protein-protein and protein-DNA interactions. These characteristics make the described polyplexes a suitable biological material for gene delivery.

Exploring downstream process for improving modular protein nanoparticles efficiency

As described in a previous work T22-GFP-H6 is described to spontaneously self-assemble in protein nanoparticles of around 13 nm upon recombinant production and purification from bacteria[217] (22923991). In this study we adapted T22-GFP-H6 with additional modules to determine if this artificial virus with specific tropism to CXCR4+ cells of primary tumor and metastatic foci in colorectal cancer is a suitable candidate for a targeted therapy delivery. In this way, five different T22-empowered multifunctional proteins were engineered with a decalysine tail (K10) and nuclear localization signal domain (NLS), either from SV40 T antigen or from the human C-myc nuclear protein (**Fig. 29**).

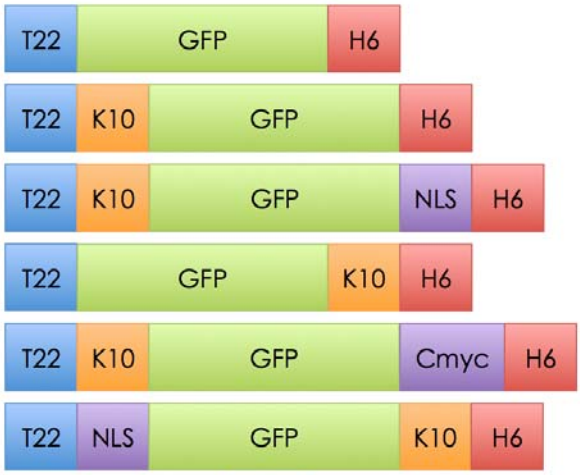


*Fig. 29 From the top to the bottom: T22-GFP-H6, T22-KGFP-H6, T22KGFPN-H6, T22GFPK-H6, T22-KGFPCmyc-H6, T22-NGFPK-H6, multidomain proteins engineered. Adapted from **Article 3**.*

All purified proteins showed a self-assembled nanoparticle organization and were tested for a CXCR4+ cells internalization experiment showing only a slightly reduction in their uptake abilities (**Article 3 Fig. 1b-c**). To determine their DNA condensing competence, electrophoretic mobility shift assays (EMSA) were performed showing a full lacking of DNA binding capacity for all the engineered proteins. On the contrary, 260/280 absorbance ratio (**Article 2 Table 1**) and the intense fluorescence signal detected in controls (**Article 3 Fig. 2**), suggest the presence of nucleic acids also in non-challenged proteins. These observations suggest us to consider nucleic acid contamination deriving from bacterial expression and purification, which might impede efficient bind of added DNA condensation ability.

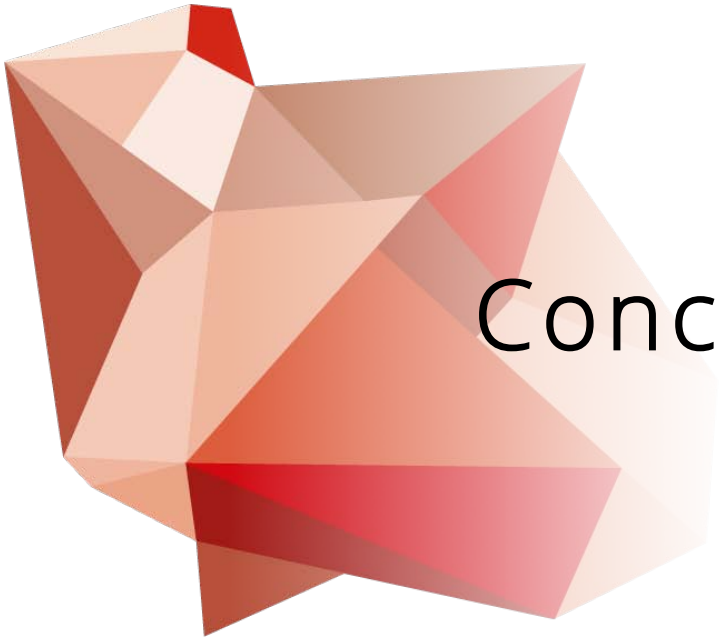
In fact, the presence of nucleic acids contaminants, with a prevalence of DNA over RNA, was determined after challenging the model protein T22-GFP-H6 to nucleases

(**Article 3 Fig. 3a**). Interestingly, a RNAse and DNase combined treatment effectively remove nucleic acid contaminations and restore the ability to incorporate exogenous DNA (**Article 3 Fig. 3b**).

If compared with R9-GFP-H6, the T22-empowered proteins contain a higher number of positive charged residues, showing a higher bacterial nucleic acid contamination tendency. R9-GFP-H6 has not shown to be so susceptible of bacterial nucleic acid contamination but still retain an exogenous DNA binding capacity after purification[271]. In this regard, R9-empowered engineered protein with decreasing positive charge residues showed a decreasing binding activity of bacterial nucleic acid (**Article 3 Table 2**).

All these observations lead us to conclude that, in order to allow of T22-empowered proteins to deliver a cargo DNA in CXCR4+ cells, the simple nanoparticles production and purification was not enough to obtain suitable nanocarriers. We moreover assumed that the weak point in the procedure is the failure of modular proteins to bind and condense exogenous DNA after purification due to the variable tendency of proteins to bind cellular nucleic acids during the purification process.

Therefore we investigated the effects of DNase and RNase treatment in pre-purified proteins, in order to establish if a contaminant nucleic acid hydrolysis in downstream process could enhance T22-NGFP-H6 nanoparticles-exogenous DNA delivery efficiency. When nuclease treated and non-treated protein samples were compared, we observed that both protein versions could efficiently internalize CXCR4+ cells but exogenous gene reporter expression was detected only in cells challenged with nuclease-treated samples (**Article 3 Fig. 4a-b**). In agreement with experimental data, we have reported the design of a T22-empowered modular protein, described penetrate CXCR4+ cells. Moreover, we observed a decrease in exogenous DNA binding ability associated to an increase of positive charges in building blocks. This fact was proved to be due to a nucleic acid contamination of bacterial origin. The loss of functionality was then restored improving a downstream process, adding a treatment with nucleases before the purification step. This is particularly important since we already observed that protein DNA interactions are intimately correlated to nanoparticle formation by self-assembling and also to the nanoparticle morphology (**Artice 2**). This modification to the protocol permits then to obtain fully functional artificial virus and moreover to give a higher level of vector safety, being bacterial genetic material removed. As final observation, variation on bacterial cell disruption methods does not show any observable lowering effect on the bacterial nucleic acid contamination.



Conclusions

1. VP1 *hJC virus* VLPs can be produced in the *E. coli* cell factory and their conformational stability is influenced by salt concentration and pH of storage buffer compositions.

2. When VP1 recombinant protein is produced in *E. coli* upon three different DnaK molecular chaperone environment, namely DnaK wild type expression, DnaK over-expression and DnaK lacking of expression, it forms VLPs structures in all conditions.

3. Yield is favored when VP1 is produced in the absence of DnaK chaperone meanwhile conformational quality is favored when expressed in $DnaK^+$ strains.

4. Overexpression of DnaK does not favour formation of VP1 VLPs more than the wild type DnaK at physiological expression level.

5. VP1 *hJC virus* VLPs can be produced in Sf9 insect cell factory and the production can be optimized depending on the multiplicity of infection and time of infection of baculoviral vector.

6. The duo of DnaK and DnaJ bacterial chaperones can be rehosted in insect cells protein factory through co-expression baculoviral vector infection.

7. VP1 protein and DnaK/DnaJ chaperones can be co-expressed in insect cells showing a decrease in the yield of VP1 protein.

8. VP1 VLPs produced upon chaperone modulation show a higher conformational quality if compared with VP1 VLPs expressed alone in insect cells.

9. VP1 VLPs structures produced both with or without chaperones can be disassembled/assembled by changing storage buffer composition.

10. R9-GFP-H6 self-assembling protein nanoparticles, containing nucleic acid binding domains, forms stable artificial virus-like structure when combined with exogenous DNA in specific environmental conditions.

11. R9-GFP-H6 structures can be divided into two different particle populations: one with small spherical shape and the other with larger rod-shaped morphology.

12. The cargo DNA is found to be completely shielded in the inner part of R9-GFP-H6/nucleic acid supramolecular structures. These protein structures efficiently protect DNA form DNase I mediated hydrolysis.

13. Starting from experimental observation it is possible to suggest a bioinformatic protein-protein model interaction, based on the DNA fitting in the inner part of the nanostructures.

14. Strong detriment of artificial viruses functionality is detected when T22 empowered self-assembling proteins containing nucleic acid binding domains, are expressed in bacterial hosts due to bacterial nucleic acid contamination.

15. Bacterial nucleic acid contamination is principally due to DNA and his magnitude depends on the number of positive charges in the nucleic acid binding domain.

16. Treatment of purified protein nanoparticles formed by R9GFPH6 and T22GFPH6 with DNase/RNase allows obtaining fully functional DNA/RNA contaminant-free artificial viruses suitable for their formulation as gene delivery vehicles.

Annexes

Annex 1

Improving the conformational quality of *hJCV* VP1 virus-like particles in recombinant baculoviruses by overexpression of DnaK/J *E. coli* chaperones

Saccardo P, Ferrer-Miralles N, Unzueta U, Villaverde A, Rodríguez-Carmona E.

MANUSCRIPT
Submitted to Applied Microbiology and Biotechnology

Rehosting bacterial chaperone DnaK/DnaJ in eukaryotic insect cells protein production system can be useful to explore recombinant protein production optimization. In this way, DnaK/DnaJ folding task can be uncoupled from prokaryotic protease activity. Moreover, multiple expression feature of Baculovirus expression system in insect cells, offer an interesting possibility to investigate *human JC* VP1 virus-ike particles produced upon different molecular chaperones environment.

In this work we explored the effects in yield and conformational quality of VP1 *hJCV* virus-like particles upon re-hosted DnaK/DnaJ expression in *Spodoptera Furgiperda* insect cells. Virus-like particles' conformation was analyzed through transmission electron microscopy and hemagglutination assay. Being VP1 VLPs open/close state controlled *in vitro* upon divalent ions and reducing agent concentration in buffer composition, we tested the VLPs produced in different chaperone environment for the open/close conformation state flexibility.

Title: "Improving the conformational quality of *hJCV* VP1 virus-like particles in recombinant baculoviruses by overexpression of DnaK/J *E. coli* chaperones"

Paolo Saccardo[1,2,3], Neus Ferrer-Miralles[1,2,3], Ugutz Unzueta[1,2,3], Antonio Villaverde[1,2,3], Escarlata Rodríguez-Carmona[1,2,3*]

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

*Corresponding author
Escarlata Rodríguez-Carmona
Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona and CIBER en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, 08193 Barcelona, Spain
e-mail: escarlata.rodriguez@uab.cat

Abstract

Virus-like particles (VLPs) are self-assembling nanostructures formed by viral capsid proteins that are able to spontaneously self-assemble, when produced as recombinant proteins in cell factories. These protein structures are considered promising nanovehicles for targeted drug delivery since they retain original tridimensional conformation, cellular tropism and the ability to deliver internalized molecules, meanwhile lacking of infective viral genome. VP1 major capsid protein of human JC virus (*hJCV*) forms VLPs structures when expressed in *Spodoptera frugiperda* insect cells (Sf9) and elicit red blood cells (RBC) hemagglutination. In this work we re-hosted prokaryotic DnaK and DnaJ chaperone pair in insect cells and we have explored the effects on VP1 *hJCV* VLPs behavior's. Co-infection of prokaryotic DnaK/J chaperone pair and VP1 protein, cloned in 2 different baculoviruses, showed a decrease in VLPs production. However, hemagglutination assay revealed that even in minor concentration, purified VLPs obtained with chaperones showed high titer than VLPs obtained without DnaK/J co-expression. Buffer compositions can also module hemagglutination properties of VP1 VLPs, being reduction agents and chelating molecules responsible of nanoparticles disassembling, in a reversible process.

Finally, the modulation of hemagglutination properties of samples suggests an influence of re-hosted chaperone proteins in VP1 *hJCV* VLPs yield and quality conformation, being a lower protein production favoring a higher quality structure of self-assembling proteins.

Keywords Baculovirus; bacterial chaperones; hemagglutination; human JC virus VP1 virus-like particles; ultracentrifugation; protein nanoparticle.

Introduction

The nonenveloped JC polyomavirus (*hJCV*), is a double strand DNA virus which infects more than 80% of the human population, showing no adverse effects in normal healthy individuals. However, in immunocompromised patients the virus spreads to the brain causing a fatal demyelination disease, known as progressive multifocal leukoencephalopaty (PML) and, moreover, it has been related to tumors in central nervous system (Dalianis and Hirsch 2013; Tan and Koralnik 2010). *hJCV* capsids are composed by three viral proteins, VP1, VP2 and VP3. The major viral capsid protein VP1 assemble in 72 pentamers forming a T=7 icosahedrical structure of about 50 nm in diameter meanwhile minor capsid proteins VP2 or either VP3 bind the central cavity of pentamers (Rayment et al. 1982). It has been widely described that VP1 protein is able to self-assemble into virus-like particles (VLPs) structures in absence of VP2 and VP3 proteins. These nanoparticles retain the structural conformation of the virus as well as cell specificity targeting, while lacking the viral DNA and minor proteins. The dynamic assembly/disassembly of these nanostructures can also be controlled by presence or absence of reductive compounds and divalent ions, respectively. Interestingly, in the reassociation process, reassembled VLPs showed the ability to encapsulate foreign nucleic acids or drugs (Chang et al. 1997; Goldmann et al. 2000; Chang et al. 2011). Moreover, while most of VLPs derived from *Polyomaviridae* family have shown to elicit immune response when administrated in an *in vivo* system, *hJCV* VP1 virus-like particles show different behavior, lacking the activation of immune response. Due to this feature, these nanostructures are good candidates for nanomedical applications as nucleic acids/drug delivery systems (Gedvilaite et al. 2006).

Human JCV VP1 VLPs can be produced in either eukaryotic (Chang et al. 1997; Hale et al. 2002) or prokaryotic cells (Ou et al. 1999). Limitations of prokaryotic expression system for posttranslational modifications including the formation of disulfide bonds are added to the strict regulation imposed by the protein quality control system for expression of recombinant proteins. In this context, in a previous work, we described the influence of the bacterial heat shock protein DnaK, a major cytosolic chaperone, on the production, solubility and supramolecular organization of VP1 *hJCV* VLPs in an *E. coli* prokaryotic expression system (Saccardo et al. 2014). Besides avoiding aggregation and promoting correct folding of misfolded proteins in cellular stress condition, Hsp70 DnaK chaperone mainly assist the newly translated proteins, taking part in the guided translocation system and in the degradation pathway of unstable proteins (Bukau and Horwich 1998; Gross et al. 1996). As the newly synthesized protein emerges from the ribosomal complex, molecular chaperones bind it, stabilizing the polypeptide chain. This is possible because of Hsp70-ATP interaction. These interactions cause a chaperone

conformational change exposing hydrophobic residues, which bind the hydrophobic domains of unfolded nascent protein. ATP hydrolysis to ADP leads Hsp70 to the release of the polypeptide chain, acting as a driving force for the substrate folding. The ATP hydrolysis is also mediated by other chaperone proteins like DnaJ, which take part to the folding cycle stimulating the ATPase activity of Hsp70 (Wall et al. 1994; Karzai and McMacken 1996; Szabo et al. 1996).

The continuous binding/release cycles not only drive the folding reaction, but also permit other chaperones such as Hsp90 complex system or proteases as Lon and ClpP to take part of a dynamic multidirectional folding process (Szabo et al. 1994; Buchberger et al. 1996).

In these work we have explored the co-production of DnaK and DnaJ folding modulators and VP1 *hJCV* VLPs in insect cells/baculovirus eukaryotic expression system. Since insect cells are devoid of ClpP/Lon proteases orthologs we expect that protease target sites exposed by DnaK/DnaJ activity would not be recognized. Moreover, insect cells are described to be able to produce VLPs structures up to several mg of proteins per liter of culture (Roldao et al. 2010).

With this aim we re-hosted the DnaK/DnaJ *E. coli* chaperones in *Spodoptera frugiperda* (Sf9) cells by co-infecting baculoviruses that express the chaperone pair and *hJCV* VP1 protein, exploring the effect of the bacterial DnaK/DnaJ chaperones on the protein yield, hemagglutination activity and structural conformation of VP1 *hJCV* VLPs.

Material and methods

Recombinant baculovirus generation

Recombinant baculoviruses expressing VP1 *hJCV* protein were obtained using the Bac-to-Bac Baculovirus Expression System by Invitrogen (#10712-024). Briefly, VP1 *hJCV* coding sequence was obtained by PCR from genomic DNA of the viral strain MAD-1 (ATCC # 45027) using primers JCVP1_Fw: 5' CGGGTATACGGATCCGCCACCATGGCCCCAACAAAAAGAAAAG 3' and JCVP1_Rv: 5' CGGGTATACGGATCCTTACAGCATTTTTGTCTGCAAC 3'. Underlined nucleotides representing the *BamH*I restriction site and italic nucleotides the Kozak sequence added prior to the start codon. The amplified DNA fragment was subcloned in *BamH*I restriction site (10220612001, Roche) into pFastBac™ Dual expression vector (10712-024, Invitrogen, USA) under Polyhedrin promoter (P pol) control. The protein sequence corresponds then to UniPROT P03089-1. The new vector, named pFastBacDual_VP1 was later transformed in DH10Bac *E. coli* strain (Invitrogen, F-mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ-rpsL nupG/pMON14272/pMON7124) harboring a baculovirus shuttle vector (bMON14272) and a helper plasmid (pMON7142). Positive recombinant baculoviruses were selected by blue/white colony colorimetric assay and subsequently recovered with GeneJET Plasmid Miniprep Kit (K0502, Thermo Scientific). Recombinant baculoviruses were then transfected in *Spodoptera frugiperda* insect cells (Sf9) with FuGene HD transfection reaction agent (#04709691001, Roche).

The baculovirus expressing both DnaK and DnaJ *E. coli* proteins was described previously (Martinez-Alonso et al. 2009).

Time course experiment

Sf9 cells were grown in Insect-XPRESS™ Protein-free Insect Cell Medium (Lonza) suspension at 27°C under constant shaking at 110 rpm. In order to establish the optimal condition for VP1 protein expression, 4 Sf9 cultures in exponential growth phase at cellular density of $1x10^6$ cell/ml were infected by VP1 baculovirus at M.O.I of 1-2-5-10 respectively. For VP1-DnaK/DnaJ time course experiments, 4 Sf9 cultures in exponential growth phase at cellular density of $1x10^6$ cell/ml were infected by VP1 baculovirus at M.O.I of 1-2-5-10 respectively and by DnaK/DnaJ baculovirus at M.O.I. 0,1. Every 24 hours, until 4 days post infection, 1 ml of sample was taken from Sf9 cultures. Each pellet/supernatant fraction from different M.O.I. culture was obtained by 4 min centrifugation at 250g. Pellet was dissolved in an equal volume of PBS and every fraction was analyzed by western blotting.

In all experiments, VP1 protein and DnaK/DnaJ were revealed by SDS-PAGE transferred onto nitrocellulose membrane and blocked with 5% milk O/N. Membrane was subsequently incubated with 1:1,500 PBS dilution of mouse monoclonal primary antibody to *hJCV* Polyomavirus capsid protein VP1 (Abcam, ab34756) and a 1:2000 diluted goat anti-mouse IgG (H+L)-HRP conjugate antibody (Bio-Rad) was used for detection. DnaK and DnaJ were detected using an hyperimmune rabbit serum and commercial rabbit polyclonal antibody (Stressgen, #SPA-410), respectively. In both cases, 1:2000 diluted goat anti-rabbit IgG (H+L)-HRP conjugate antibody (Bio-Rad) was used for detection.

For the quantification experiments, the Major Capsid VP1 Protein (Abcam, ab74569) was used as protein standard for 240 ng to 30 ng standard curves. Based on linear regression analysis we determine the *hJCV* VP1 samples concentrations.

VLPs production and purification

According to time course experiments, VP1 VLPs productions were done infecting 200 ml of Sf9 cell culture at M.O.I. 1 and collecting cells suspension at 96 hours post infection. Pellet and supernatant were separated centrifuging cell culture 15 min at 10,000g. DnaK/DnaJ recombinant proteins and VP1 VLPs were produced co-infecting 200 ml of Sf9 cell culture at M.O.I. of 0.1 and M.O.I. of 1, respectively. After 96 hours, supernatant and cellular pellet were separated as described before. Supernatant was then applied to a 40% sucrose PBS cushion and centrifuged at 4°C during 4 hours at 100,000g with a SW 27 swinging bucket Beckman rotor. After centrifugation, pellets were dissolved in 400 µl of PBS. Samples were then dialyzed O/N in Tris 20 mM, 150 mM NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, pH 7.5 buffer (Goldmann et al. 2000).

Hemagglutination (HA) assay

Red blood cells (RBC) hemagglutination assay for detection of *hJCV* VP1 protein was made mixing 0.15 µg of protein fit to 50 µl of volume, with an equal volume of 0.5% of RBC diluted in PBS. The 96 well plates with two fold serial dilution of protein and RBC were incubated for 2 h at 4°C and then observed. HA assays were performed by duplicate for the time course analysis and in single replica for disassembling/reassembling assay.

Disassembling/reassembling assay

0.75 µg of purified VP1 VLPs were disassociated in 250 µl dissociation buffer (20 mM Tris–HCl, pH 7.5, 150 mM NaCl, 10 mM EGTA, 5 mM DTT) for 1h at room temperature. For reassociation assay, samples were dialyzed O/N against Tris 20 mM, 150 mM NaCl, 1mM

$MgCl_2$, 1mM $CaCl_2$, pH 7.5 buffer. After purification, disassociation and reassociation steps, samples of 50 µl were taken in order to perform HA analysis. The HA Units/ml value was obtained correcting the highest dilution factor that produce a positive reading to a factor of 20.

Transmission electron microscopy

For the Transmission Electron Microscopy (TEM) analysis, 10 µl of VP1 VLPs samples were loaded on a carbon coated grids for 2 min and then dried. Subsequently, 10 µl of uranyl acetate 2 % [w/v] aqueous solution were applied and dried after 2 min. Photographs were taken with JEM-1400 transmission electron microscope.

Results

Production of hJCV major capsid VP1 in insect cells

The major capsid structural protein VP1 of *hJCV* was successfully produced by recombinant baculovirus containing *hJCV* DNA fragment encoding the entire VP1 region under the control of baculovirus polyhedrin promoter in sf9 insect cells according to the western blot (Fig 1a). One major protein with molecular mass of 39.6 kDa was detected both in cell lysates and in the culture medium supernatants. VP1 protein was observed in cell lysates at 48 hours achieving a maximum yield at 72 hours post-infection (p.i.) independently of the MOI assayed. At this time, VP1 protein appeared in culture medium supernatants, achieving a peak at 96 hours p.i. at MOI 1. As the yield of protein in supernatants increased, the protein expression in cell lysates decreased gradually until it disappeared completely at 4 days p.i. According to these results, 4 days p.i. at MOI 1 was chosen as the harvesting culture medium conditions for VP1 *hJCV* VLPs production.

Co-expression of hJCV major capsid VP1 with E. coli DnaK/DnaJ folding modulators

To explore the co-production of *E. coli* DnaK and DnaJ folding modulators and VP1 *hJCV* VLPs in insect cells-baculovirus eukaryotic expression system, the DnaK/DnaJ *E. coli* chaperone pair were first co-produced in Sf9 cells at MOI 0.1 and detected by western blot analysis using commercial rabbit polyclonal antibody against DnaK/DnaJ, after 4 days post infection. Two major proteins with molecular mass of 69.1 and 41.1 kDa, respectively, were detected in the culture medium supernatants (Fig 2a). A similar DnaK/DnaJ expression pattern was detected when Sf9 cells were co-infected with baculoviruses producing the *hJCV* VP1 protein and DnaK/DnaJ *E. coli* chaperone pair. The produced proteins had also a molecular weight of 69.1 and 41.1 kDa.

On the other hand, the major capsid structural protein VP1 of *hJCV* was successfully produced when Sf9 cells were co-infected with baculoviruses producing the DnaK/DnaJ *E. coli* chaperone pair. VP1 protein was detected by western blot analysis with specific mouse monoclonal primary antibody to *hJCV* Polyomavirus capsid protein VP1, as one major protein with molecular mass of 39.6 kDa in the culture medium supernatants (Fig 2b). Interestingly, the *hJCV* VP1 protein production was negatively affected by DnaK/DnaJ production if compared with *hJCV* VP1produced alone. Therefore, in order to collect information related to the influence of DnaK/DnaJ chaperone pair folding modulators on the production of *hJCV* VP1 VLPs, we quantified the total amount of recombinant VP1 obtained in both types of experiments. The total amount of VP1 was nearly twofold higher in *hJCV* VP1 VLPs produced alone and the

presence of DnaK/DnaJ had a negative effect on VP1 production (Fig 2b) as also observed in VP1-DnaK/DnaJ time course expression assay (Fig 1b).

Concentration and characterization of purified VP1 and DnakJ-VP1 VLPs

The results of concentrated sample western blot proved that VLPs were formed by *hJCV* VP1 protein as expected (Fig 3). The main component of the VP1 *hJCV* VLP fraction was a 39.6 kDa protein which corresponds to the molecular weight of monomeric VP1. In addition, trimeric forms of the recombinant VP1 were detected when overexpressed as previously described even under denaturing conditions (Chen et al. 2001). As observed by TEM, the purified and concentrated *hJCV* VP1 VLP (Fig 4a) and DnaK/DnaJ-*hJCV* VP1 VLP (Fig 4b) fractions contained both, particles with regular icosahedral symmetry and a diameter of ≈50 nm, which were identified as VP1 VLPs exhibiting the usual capsid morphology.

HA activity of *hJCV* VP1 and Dnak/DnaJ-VP1 VLPs

The effect in biological activity of *hJCV* VP1 VLPs produced in supernantants of Sf9 cells at MOI 1, 72 and 96 hours p.i. and the concentrated and purified VLPs fractions obtained by ultracentrifugation at MOI 1 and 96 hours post infection, was determined by analyzing their ability to hemagglutinate red blood cells (Figure 5). In those experiments, the same sample volume of recombinant *hJCV* VP1 was used. PBS was used as negative control. As shown in Figure 5, both the supernatants and the purified and concentrated VLP fractions can hemagglutinate RBC. HA activity with higher titre was observed at 96 hours p.i (160 HA units/mL) compare to that obtained at 72 hours p.i (40 HA units/mL) (Fig 5a) corresponding to the different VP1 concentrations observed at each specific harvesting time post-infection as seen in western blot of Figure 1. The concentrated fractions obtained by 40% sucrose cushion ultracentrifugation (Fig 5b and 5c) had HA activities with higher titres (2560 HA units/mL) than that observed in the supernatant fractions that agglutinated until 1:8 per 50 µl volume sample at 96 h p.i. These data suggest that *hJCV* VP1 VLPs have correctly assembled and retain hemagglutination activity.

Taking into account the results shown above and in order to estimate the influence of DnaK/DnaJ on the biological activity VLPs, we determined the ability to hemagglutinate red blood cells of the purified and concentrated *hJCV* VP1 and DnaK/DnaJ-*hJCV* VP1 VLPs once associated, dissociated and reassociated (Fig 6a and 6b). The hemagglutination assay results showed that the presence of DnaK/DnaJ had a significant positive effect on the ability of VP1 VLPs to hemagglutinate RBC (8 fold HA titres higher than *hJCV* VP1 VLP HA titers) indicating a direct effect of DnaK/DnaJ on the correct assembly of these protein supramolecular assemblies. In addition, it has to

be taken into account that in the same sample volume, the total amount of VP1 is significantly lower in DnaK/DnaJ-*hJCV* VP1 VLPs compared to *hJCV* VP1 VLPs. About 50 % and 25 % of initially used *hJCV* VP1 VLPs and DnaK/DnaJ- *hJCV* VP1 VLPs, respectively, could be reassociated from VP1-pentamers to complete VP1 VLPs. Moreover, once dissociated and reassociated, the DnaK/DnaJ pair still continues to have a positive effect with higher HA titres (5120 HA units/mL) than that obtained with *hJCV* VP1 VLPs (1280 HA units/mL).

The efficacy of the dissociation and reassociation process was also determined by TEM (Fig 7). As mentioned before, it could be observed a higher amount of *hJCV* VP1 VLPs compare to VP1 VLPs produced in the presence of DnaK/DnaJ (Fig 7a). After the dissociation and reassociation process, the same pattern is observed with a higher amount of reassociated VP1 VLPs.

Discussion

The baculovirus-insect cell expression system (BES) has been successfully used as eukaryotic system for the production and functional self-assembly of recombinant multiprotein complexes such as virus-like particles (Liu et al. 2013; Yamaji 2014). BES, in contrast to prokaryotic expression systems, exhibits some properties including prostranslational modifications and presence of appropriate chaperones (Roy and Noad 2012) that support proper assembly. Production and self-assembly of recombinant *hJCV* VP1 VLPs has been described using BES and their biological activity and correct supramolecular assembly has been proved by hemagglutination assay and transmission electron microscopy (TEM) (Chang et al. 1997; Goldmann et al. 1999; Kobayashi et al. 2013). However, one of the main problems for heterologous multiprotein complex production is the low fraction of correctly assembled protein. In this scenario, one of the approaches to overcome this limitation has been the co-expression of foreign chaperones in order to improve protein yield and solubility and structural and functional conformation (Sokolenko et al. 2012; Martinez-Alonso et al. 2010).

As mentioned before, chaperone proteins play a key role in the folding process. The heat shock proteins machinery can lead to a correct tridimensional structure conformation of newly synthesized protein or send it to proteases complexes. This balance can be specifically modulated by switching off the chaperone expression or, by the other side, over expressing heat shock proteins. The effects on protein solubility or structural conformation may vary depending on cell factory or recombinant proteins itself (Martinez-Alonso et al. 2009; Saibil 2013; Kolaj et al. 2009). In this study we wanted to explore the effect of the coproduction of DnaK and DnaJ *E. coli* folding modulators, prokaryotic homologues of eukarytoic Hsp70 and Hsp40 chaperones, using the BES on the production of recombinant *hJCV* VP1 VLPs. The protein yield, biological activity and structural conformation were determined by hemagglutination activity and TEM *in vivo*. For that, we have used the co-infection approach by infecting the same batch of Sf9 cells with two recombinant baculoviruses, one of them expressing the chaperone pair and the other one expressing the *hJCV* VP1 protein. In this study, it has been observed that the production of VLPs is negatively affected in the presence of DnaK/J. This was in contrast to what was observed by Martínez-Alonso and collaborators in a previous work. Shortly, they used a single baculovirus infection to co-produce the same chaperones and recombinant mGFP, foot-and-mouth disease virus (FMDV) VP1 and VP2 capsid proteins and human alpha-galactosidase A in insect cells, showing a positive effect on protein yield, proteolytic stability, protein solubility and global biological activity (Martinez-Alonso et al. 2009). Although a positive effect on *hJCV* VP1

protein yield was observed by co-producing the chaperones machinery, the two baculoviral co-infection approach used in our work could have led to different protein yields. Moreover, it has been described that p10 promoter can affect the expression of the genes that are under the transcriptional control of the polyhedrin promoter (Sokolenko et al. 2012).

In addition, the decrease in the amount of *hJCV* VP1 VLPs in the presence of DnaK/J could be due to the effect of competition between the two baculoviruses used during the co-infection process, being the virus encoding the *hJCV*-VP1 less efficient that the virus encoding the DnaK/J folding modulators, but also depending on the ratios of the viruses chosen. M.O.I. and T.O.I. in fact, are variables that can affect the recombinant protein expression, especially in co-infection with two or more baculovirus (Sokolenko et al. 2012). In our case, the ratio of the two baculoviruses used was 1:10 (M.O.I. of 0.1 for DnaK/J and M.O.I. of 1 for *hJCV* VP1) and even that, the expression of the gene that was negatively affected corresponds to the more abundant recombinant baculovirus used in the infection.

Although the co-infection approach has been used successfully for the production of many self-assembly multi-protein complexes (Roy and Noad 2012), it could have a great impact on the production of optimal yield of recombinant proteins because the ratio of expression of the different subunits in the complex (expressed with different viruses in the same batch of infected cells with the subsequent generation of different subpopulations) could vary (Roy and Noad 2012). However, in this work, it has been demonstrated the positive effect on biological activity and structural conformation observed in the presence of DnaK/J chaperone in recombinant *hJCV* VP1 production. Similar effects were found in previous works in different cell factories like Insect cells and *E.coli,* verifying the role of chaperones in regulation of protein folding. Interestingly, recombinant protein production yield not always seems to be positively affected.

In conclusion, the yield of *hJCV* VP1 protein, when co-produced with DnaK/DnaJ, is negatively affected since a remarkable reduction in protein detection is observed in insect cell culture supernatant. Nevertheless the biological activity of VP1 VLPs is positively affected by the presence of chaperones in the expression system, reflected by a higher titer in Red Blood Cells hemagglutination assay. Moreover, the hJVP1 VLPs produced in a DnaK/DnaJ background retain the improved conformational state when subjected to association/disassociation procedures proving the importance of chaperone co-production when aiming to obtain high conformational quality recombinant proteins. In this respect, it has been shown that protein quality and solubility are not coincident, as reported also in previous work (Garcia-Fruitos et al. 2007).

Acknowledgments

The authors acknowledge the financial support received for the design and production of VLPs for gene therapy to AV from AGAUR (2009SGR-108) and CIBER de Bioingeniería, Biomateriales y Nanomedicina, 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. The authors gratefully acknowledge Polly Roy and Mónica Martínez-Alonso for their support in DnaK/J recombinant baculovirus construction. We would also thank the technical support received from the Scientific and Technical Services SCAC (Cell Culture Unit: Servei de Cultius Cel•lulars, Producció d'Anticossos i Citometria) of the Universitat Autònoma de Barcelona and from the Protein Production Platform (PPP, CIBER-BBN andUniversitat Autònoma de Barcelona) that belongs to the sePBioEs (Proteomics and Structural Biology: Proteómica i Biologia Estructural) Scientific and Technical Service of the Universitat Autònoma de Barcelona. TEM images were obtained at the Microscopy Scientific and technical Service of the Universitat Autònoma de Barcelona. Red blood cells were kindly provided by Olivia Cano Garrido, IBB, UAB. PS and UU have received predoctoral fellowship from ISCIII, and AV has been distinguished with an ICREA ACADEMIA award.

References

Buchberger A, Schroder H, Hesterkamp T, Schonfeld HJ, Bukau B (1996) Substrate shuttling between the DnaK and GroEL systems indicates a chaperone network promoting protein folding. JMolBiol 261 (3):328-333

Bukau B, Horwich AL (1998) The Hsp70 and Hsp60 chaperone machines. Cell 92 (3):351-366

Chang CF, Wang M, Ou WC, Chen PL, Shen CH, Lin PY, Fang CY, Chang D (2011) Human JC virus-like particles as a gene delivery vector. ExpertOpinBiolTher 11 (9):1169-1175

Chang D, Fung CY, Ou WC, Chao PC, Li SY, Wang M, Huang YL, Tzeng TY, Tsai RT (1997) Self-assembly of the JC virus major capsid protein, VP1, expressed in insect cells. JGenVirol 78 (Pt 6):1435-1439

Chen PL, Wang M, Ou WC, Lii CK, Chen LS, Chang D (2001) Disulfide bonds stabilize JC virus capsid-like structure by protecting calcium ions from chelation. FEBS Lett 500 (3):109-113

Dalianis T, Hirsch HH (2013) Human polyomaviruses in disease and cancer. Virology 437 (2):63-72

Garcia-Fruitos E, Martinez-Alonso M, Gonzalez-Montalban N, Valli M, Mattanovich D, Villaverde A (2007) Divergent genetic control of protein solubility and conformational quality in Escherichia coli. Journal of molecular biology 374 (1):195-205. doi:10.1016/j.jmb.2007.09.004

Gedvilaite A, Dorn DC, Sasnauskas K, Pecher G, Bulavaite A, Lawatscheck R, Staniulis J, Dalianis T, Ramqvist T, Schonrich G, Raftery MJ, Ulrich R (2006) Virus-like particles derived from major capsid protein VP1 of different polyomaviruses differ in their ability to induce maturation in human dendritic cells. Virology 354 (2):252-260

Goldmann C, Petry H, Frye S, Ast O, Ebitsch S, Jentsch KD, Kaup FJ, Weber F, Trebst C, Nisslein T, Hunsmann G, Weber T, Luke W (1999) Molecular cloning and expression of major structural protein VP1 of the human polyomavirus JC virus: formation of virus-like particles useful for immunological and therapeutic studies. J Virol 73 (5):4465-4469

Goldmann C, Stolte N, Nisslein T, Hunsmann G, Luke W, Petry H (2000) Packaging of small molecules into VP1-virus-like particles of the human polyomavirus JC virus. JVirolMethods 90 (1):85-90

Gross M, Robinson CV, Mayhew M, Hartl FU, Radford SE (1996) Significant hydrogen exchange protection in GroEL-bound DHFR is maintained during iterative rounds of substrate cycling. Protein Sci 5 (12):2506-2513

Hale AD, Bartkeviciute D, Dargeviciute A, Jin L, Knowles W, Staniulis J, Brown DW, Sasnauskas K (2002) Expression and antigenic characterization of the major capsid proteins of human polyomaviruses BK and JC in Saccharomyces cerevisiae. JVirolMethods 104 (1):93-98

Karzai AW, McMacken R (1996) A bipartite signaling mechanism involved in DnaJ-mediated activation of the Escherichia coli DnaK protein. JBiolChem 271 (19):11236-11246

Kobayashi S, Suzuki T, Igarashi M, Orba Y, Ohtake N, Nagakawa K, Niikura K, Kimura T, Kasamatsu H, Sawa H (2013) Cysteine residues in the major capsid protein, Vp1, of the JC virus are important for protein stability and oligomer formation. PLoS One 8 (10):e76668. doi:10.1371/journal.pone.0076668

Kolaj O, Spada S, Robin S, Wall JG (2009) Use of folding modulators to improve heterologous protein production in Escherichia coli. MicrobCell Fact 8:9

Liu F, Wu X, Li L, Liu Z, Wang Z (2013) Use of baculovirus expression system for generation of virus-like particles: successes and challenges. Protein ExprPurif 90 (2):104-116

Martinez-Alonso M, Gomez-Sebastian S, Escribano JM, Saiz JC, Ferrer-Miralles N, Villaverde A (2010) DnaK/DnaJ-assisted recombinant protein production in Trichoplusia ni larvae. ApplMicrobiolBiotechnol 86 (2):633-639

Martinez-Alonso M, Toledo-Rubio V, Noad R, Unzueta U, Ferrer-Miralles N, Roy P, Villaverde A (2009) Rehosting of bacterial chaperones for high-quality protein production. ApplEnvironMicrobiol 75 (24):7850-7854

Ou WC, Wang M, Fung CY, Tsai RT, Chao PC, Hseu TH, Chang D (1999) The major capsid protein, VP1, of human JC virus expressed in Escherichia coli is able to self-assemble into a capsid-like particle and deliver exogenous DNA into human kidney cells. JGenVirol 80 (Pt 1):39-46

Rayment I, Baker TS, Caspar DL, Murakami WT (1982) Polyoma virus capsid structure at 22.5 A resolution. Nature 295 (5845):110-115

Roldao A, Mellado MC, Castilho LR, Carrondo MJ, Alves PM (2010) Virus-like particles in vaccine development. Expert review of vaccines 9 (10):1149-1176. doi:10.1586/erv.10.115

Roy P, Noad R (2012) Use of bacterial artificial chromosomes in baculovirus research and recombinant protein expression: current trends and future perspectives. ISRNMicrobiol 2012:628797

Saccardo P, Rodriguez-Carmona E, Villaverde A, Ferrer-Miralles N (2014) Effect of the DnaK chaperone on the conformational quality of JCV VP1 virus-like particles produced in Escherichia coli. BiotechnolProg 30 (3):744-748

Saibil H (2013) Chaperone machines for protein folding, unfolding and disaggregation. NatRevMolCell Biol 14 (10):630-642

Sokolenko S, George S, Wagner A, Tuladhar A, Andrich JM, Aucoin MG (2012) Co-expression vs. co-infection using baculovirus expression vectors in insect cell culture: Benefits and drawbacks. BiotechnolAdv 30 (3):766-781

Szabo A, Korszun R, Hartl FU, Flanagan J (1996) A zinc finger-like domain of the molecular chaperone DnaJ is involved in binding to denatured protein substrates. EMBO J 15 (2):408-417

Szabo A, Langer T, Schroder H, Flanagan J, Bukau B, Hartl FU (1994) The ATP hydrolysis-dependent reaction cycle of the Escherichia coli Hsp70 system DnaK, DnaJ, and GrpE. ProcNatlAcadSciUSA 91 (22):10345-10349

Tan CS, Koralnik IJ (2010) Progressive multifocal leukoencephalopathy and other disorders caused by JC virus: clinical features and pathogenesis. Lancet Neurol 9 (4):425-437

Wall D, Zylicz M, Georgopoulos C (1994) The NH2-terminal 108 amino acids of the Escherichia coli DnaJ protein stimulate the ATPase activity of DnaK and are sufficient for lambda replication. JBiolChem 269 (7):5446-5451

Yamaji H (2014) Suitability and perspectives on using recombinant insect cells for the production of virus-like particles. ApplMicrobiolBiotechnol 98 (5):1963-1970

Fig 1

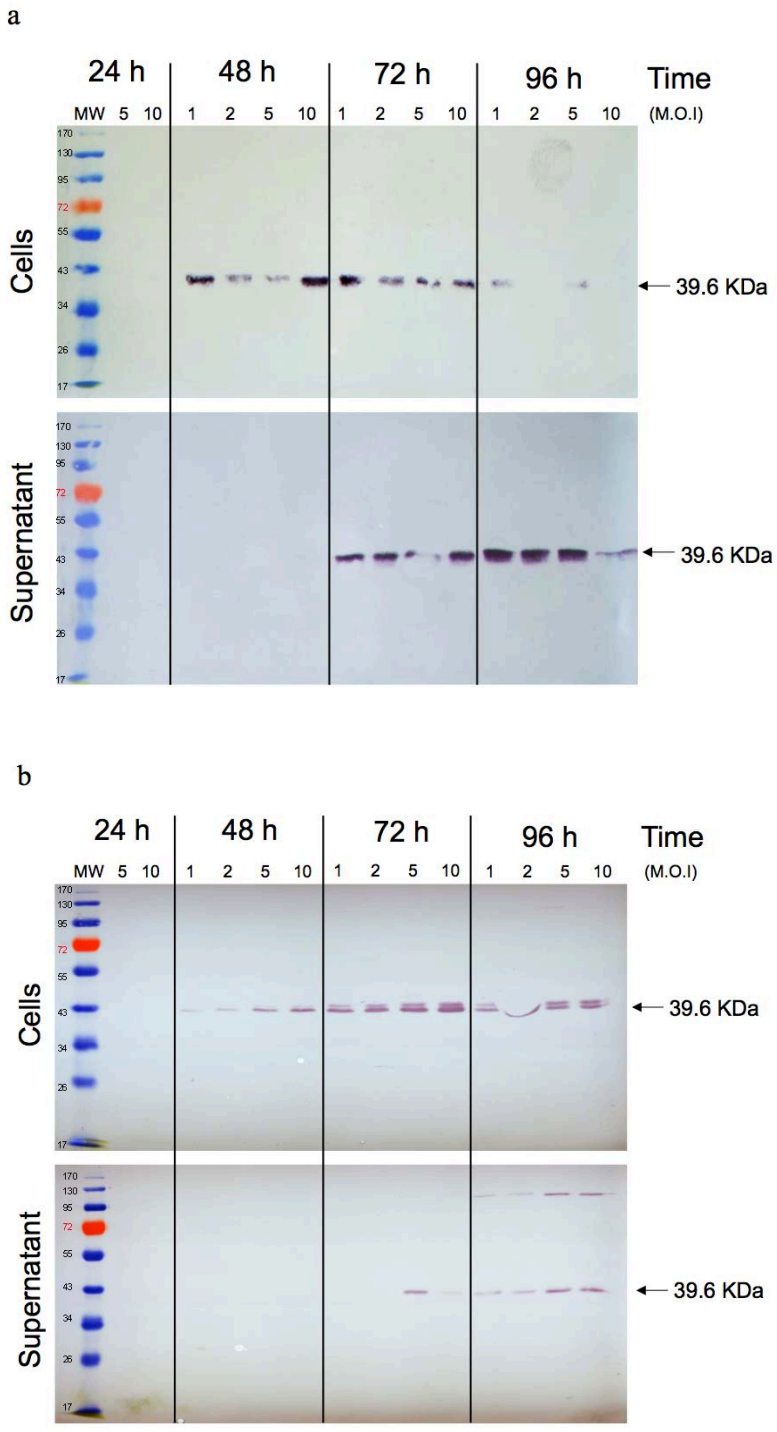


Figure 1 a) Time course of *hJCV* VP1 protein production. Different multiplicity of infection (M.O.I.) of 1, 2, 5, 10 were used to infect Sf9 cells cultures. VP1 protein yield was monitored in supernatant and cells fraction every 24 hours during 4 days. Results show that M.O.I. of 1 and 96 h of infection provide the best combination of protein expression in supernatant fraction. b) Time course of *hJCV* VP1 protein co-produced wit DnaK/DnaJ chaperones. Different VP1 baculovirus multiplicity of infection (M.O.I.) of 1,

2, 5, 10 and fixed DnaK/DnaJ baculovirus M.O.I of 0.1 were used to infect Sf9 cells cultures. VP1 protein yield was monitored in supernatant and cells fraction every 24 hours during 4 days.

Fig 2

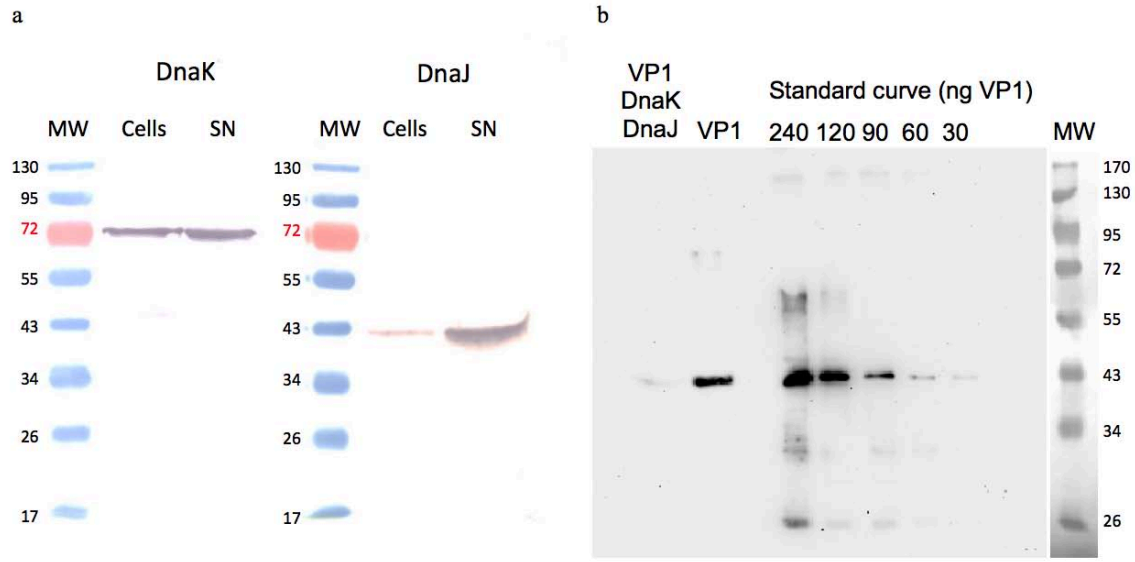


Figure 2 a) Western blot analysis of DnaK and DnaJ expression in cellular (Cells) and supernatant (SN) fraction after Sf9 insect cells infection at M.O.I. 0,1 and 96 h p.i. b) Western blot analysis and quantification of VP1 protein in supernatant fractions of Sf9 insect cells culture. From the left: VP1 sample obtained with DnaK/DnaJ co-expression, VP1 sample obtained without chaperone co-expression and 240 ng to 30 ng concentration commercial VP1 standard curve.

Fig 3

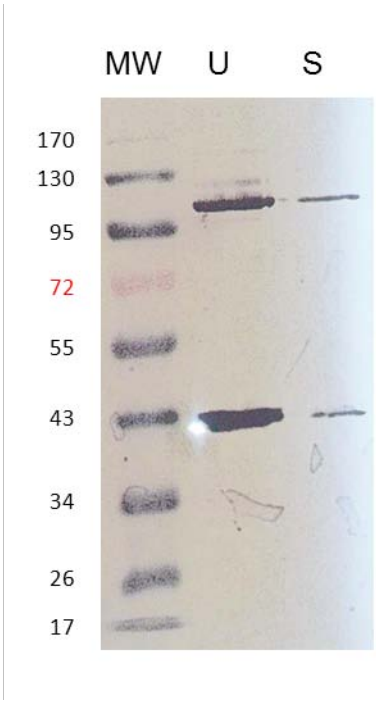


Figure 3 Western blot analysis of 500-fold volume concentrated VP1 VLPs pellet, obtained by 40 % sucrose cushion ultracentrifugation (U), and the original supernatant fraction (S). Analysis was performed loading the same volume of samples.

Fig 4

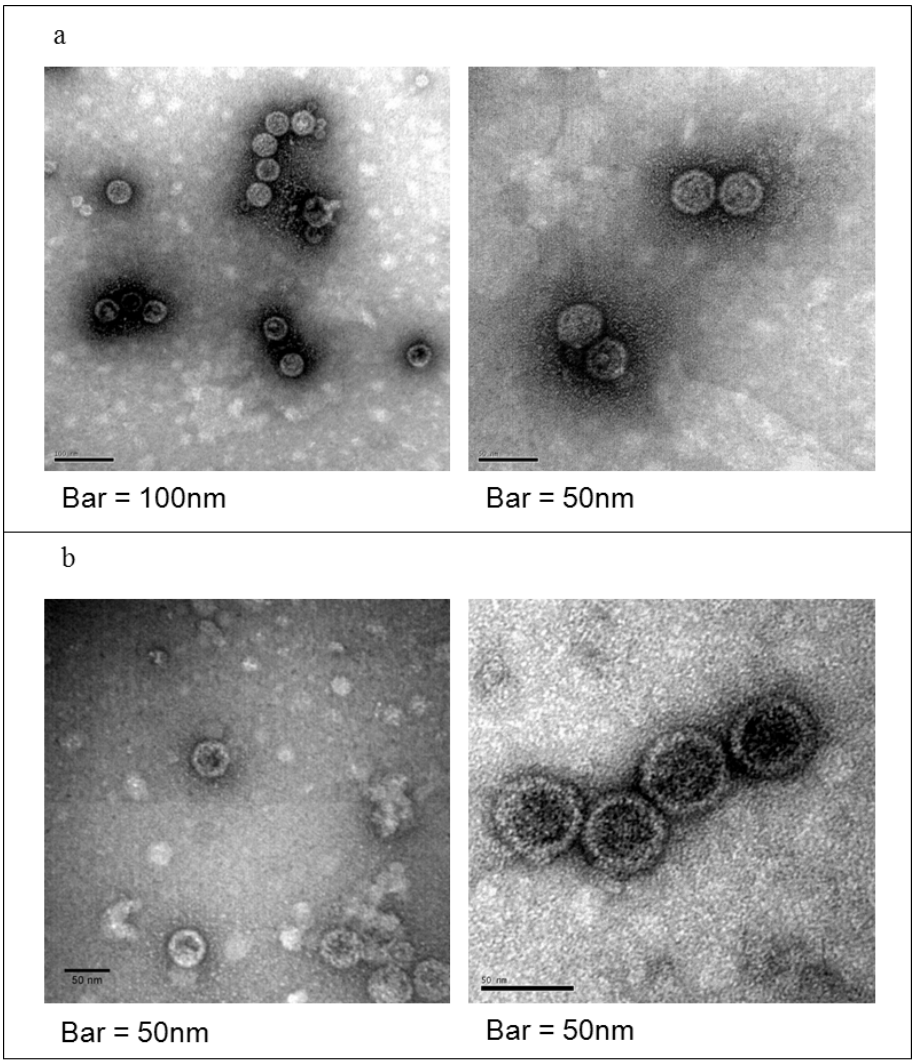
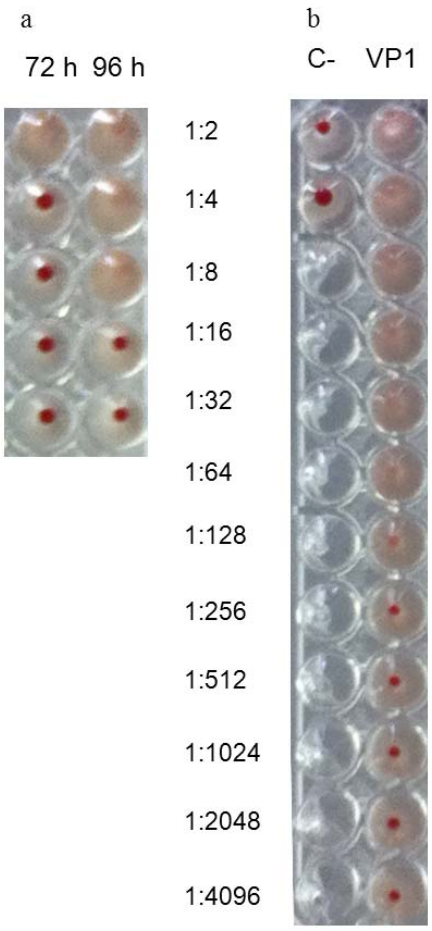


Figure 4 Transmission electron microscopy (TEM) images of a) VP1 VLPs and b) VP1 VLPs co-expressed with DnaK and DnaJ chaperones. Samples were obtained by ultracentrifugation with 40 % sucrose cushion.

Fig 5



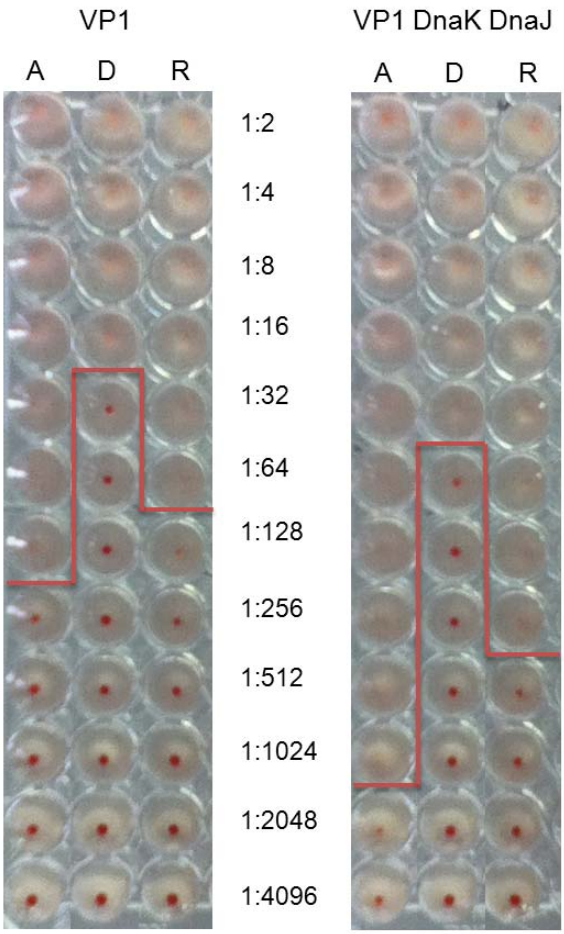
c

VLPs	HA titre (units/mL)
Supernantant hJCV VP1 VLP 72 h p.i.	40
Supernantant hJCV VP1 VLP 96 h p.i.	160
Ultracentrifugated 500foldX VLP hJCV VP1	2560

Figure 5 Chicken red blood cells hemagglutination assay (HA). a) 50 µl samples of infected sf9 culture supernatant were recollected at 72 and 96 hours post infection. In according with Figure 1, higher titre of VP1 VLPs is found at 96 hours post infection (160 HA units ml) if compared with 72 hours p.i. (40 HA units/ml). b) 50 µl of VP1 VLPs concentrated fraction sample obtained by 40 % sucrose cushion ultracentrifugation show higher titres (2560 HA units/mL) than that observed in the supernatant fractions. c) HA values (units/ml) of VP1 VLPs samples in HA assay reported in figure 5.

Fig 6

a



b

VLPs	HA titre (units/mL)
hJCV VP1- association buffer	2560
hJCV VP1- dissociation buffer	320
hJCV VP1- reassociation buffer	1280
DnaK/DnaJ- hJCV VP1- association buffer	20480
DnaK/DnaJ- hJCV VP1- dissociation buffer	640
DnaK/DnaJ- hJCV VP1- reassociation buffer	5120

Figure 6 a) Chicken red blood cells hemagglutination assay (HA) of VP1 and VP1 DnaK DnaJ samples in assembled initial conformation (A), disassembled conformation (D) and reassembled conformation (R). Assembled/disassembled/reassembled states are controlled by changes in buffer composition. b) HA units/ml values of VP1 VLPs samples in HA assay reported in figure 6.

Fig 7

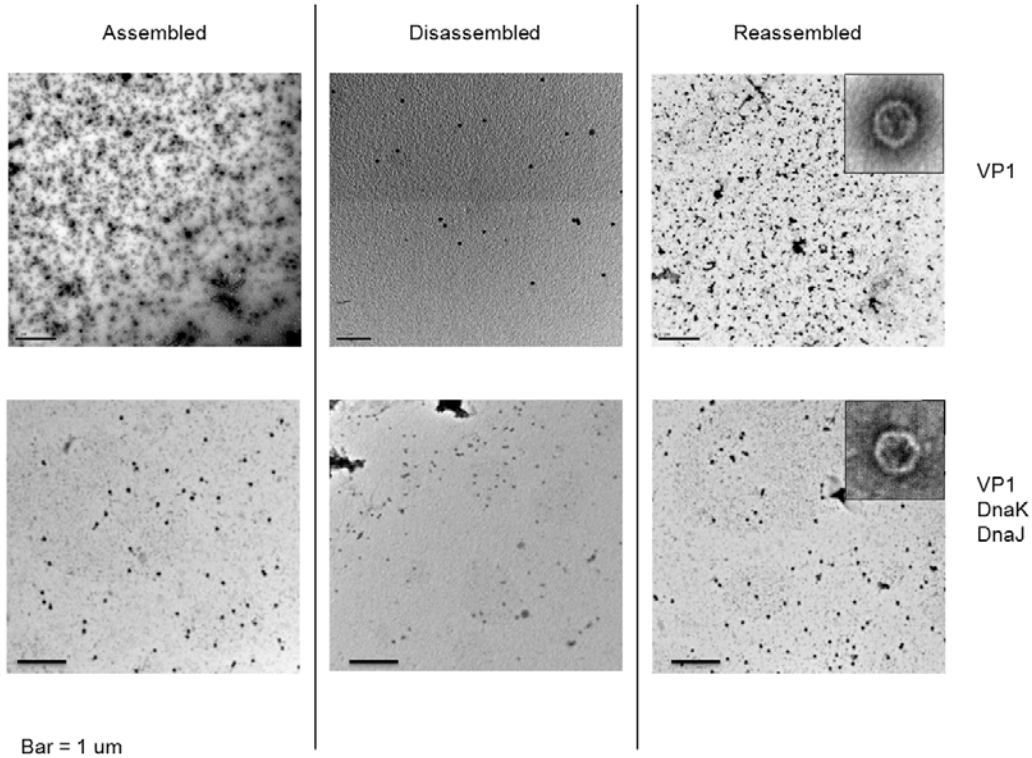


Figure 7 Transmission electron microscopy images of VP1 VLPs and VP1 expressed with DnaK DnaJ VLPs samples in the 3 different assembled/disassembles/reassembled states. According to Figure 2 and Figure 6, different concentration of VLPs can be found between the open/close states and between the two VP1 samples. In detail box: magnification of VLPs structures.

Annex

Annex 2

Biotechnology Advances 27 (2009) 432–438

Contents lists available at ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

Research review paper

Peptide-mediated DNA condensation for non-viral gene therapy

Paolo Saccardo [a], Antonio Villaverde [a,b,c,*], Nuria González-Montalbán [a,b,c]

[a] Institute for Biotechnology and Biomedicine, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain
[b] Department of Genetics and Microbiology, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain
[c] CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, 08193 Barcelona, Spain

ARTICLE INFO

Article history:
Received 26 September 2008
Received in revised form 19 March 2009
Accepted 20 March 2009
Available online 31 March 2009

Keywords:
Recombinant drugs
Gene therapy
Protein engineering
Multifunctional proteins
Cationic peptides
DNA binding
Nanoparticles
Artificial viruses

ABSTRACT

The construction of non-viral, virus-like vehicles for gene therapy involves the functionalization of multipartite constructs with nucleic acid-binding, cationic agents. Short basic peptides, alone or as fusion proteins, are appropriate DNA binding and condensing elements, whose incorporation into gene delivery vehicles results in the formation of protein–DNA complexes of appropriate size for cell internalization and intracellular trafficking. We review here the most used cationic peptides for artificial virus construction as well as the recently implemented strategies to control the architecture and biological activities of the resulting nanosized particles.

Contents

1. Introduction

Gene therapy requires safe and efficient vehicles to transfer and deliver expressible genetic material or silencing nucleic acids to target tissues. The most extensively used delivery tools are viral-based vectors since the properties of the viral cycle permit receptor-mediated recognition and cell internalization, endosomal escape, nuclear transport and DNA integration (Aris and Villaverde 2004). In the last decade, engineered adenoviruses and retroviruses have been largely explored as transfer vehicles with an acceptable degree of success (Edelstein et al., 2007). Nevertheless, the use of viruses as gene delivery systems requires the modification of the viral genome in order to prevent replication and suppress undesirable side effects while keeping the required properties. Reaching a compromise between sufficient efficiency and acceptable safety is an extremely complex issue that has generated intense scientific debates regarding the future development of viral gene therapy (Marshall 2002, 2003; Abbott 2006; Edelstein et al., 2007). In fact, the occurrence of clinically important side effects (ranging from inflammation to death of the patients) has severely delayed the incorporation of new viral vehicles into clinical trials and the whole progress of gene medicine (Edelstein et al., 2007).

* Corresponding author. Institute for Biotechnology and Biomedicine, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain. Tel.: +34 935813086; fax: +34 935812011.
E-mail address: antoni.villaverde@uab.es (A. Villaverde).

0734-9750/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.biotechadv.2009.03.004

Alternatively, non-viral vehicles, being safer than the viral counterparts are under continuous development and optimization, essentially focused to increase the efficiency of DNA or RNA delivery. Such vehicles, that include polymeric constructs, virus-like particles (VLPs) and protein-only shells (Ferrer-Miralles et al., 2008), can successfully mimic viral properties regarding both size and biological properties (Fig. 1). Therefore, they are usually referred as 'artificial viruses' (Mastrobattista et al., 2006). Artificial viruses have to fulfill some functional requirements such as highly stable transport of DNA, protection against cell nucleases, membrane crossing abilities, high and steady expression of the therapeutic DNA, low systemic and cellular toxicity and low cell complement activation (Plank et al., 1999; Kim et al., 2003). Such properties must be accompanied by the collapse of extended DNA molecule into compact, orderly nanoparticles. DNA condensation has been deeply studied in the context of packaging into viruses and virus-like particles and the revealed concepts can be applied to the functional improvement of artificial viruses. Among the available DNA-condensing tools, cationic peptides are specially appropriate for those vehicles based solely on protein elements (Table 1).

Cationic peptides and other basic polymers are positively charged and interact with the negatively charged phosphate backbone of DNA through electrostatic interactions (Bloomfield 1996). Cationic and basic peptides as well as polymer-based vehicles are known to enhance the condensation of DNA to small particles with variable shapes, namely rods, toroids and spheroids, and finite size conformation (Bloomfield 1997), mainly depending on DNA size (Vijayanathan et al., 2002; Rimann et al., 2008). Moreover, the net positive charge exhibited by cationic-related polyplexes and peptides complexes permit them to interact with cell membranes and internalize into the cell, both *in vivo* and *in vitro* systems (Wadhwa et al., 1997; Wolfert and Seymour 1998; McKenzie et al., 2000a; Tolmachov and Coutelle 2007; Kumar et al., 2007; Henke et al., 2008), overcoming membrane barriers and allowing nuclear gene delivery and expression (He et al., 2000; Martin and Rice 2007). DNA-condensing peptides also prevent DNA from being degraded by cytosolic nucleases (Wolfert and Seymour 1998) and prolong the half-life of the targeted nucleic acid, which is known to be of 50–90 min for naked plasmid DNA (Lechardeur et al., 1999). Nevertheless, the cationic carrier must retain condensation abilities in terms of controlled reversibility, for instance, by adding active groups sensitive to cellular redox-potential gradients to the peptide-based vector (Manickam et al., 2005). In lipid-derived vehicles an irreversible association that prevents DNA from being expressed is often observed (Zabner et al., 1995; Brewer et al., 1999), whereas early dissociation is the result of inadequately condensation (Keller et al., 2003).

2. Poly-L-lysines and polylysine-containing peptides as DNA-binding domains

The most used DNA-condensing cationic peptides in gene delivery systems are poly-L-lysines. Poly-L-lysines (PLL) and related peptides bind the negatively charged backbone of DNA chain, not only promoting

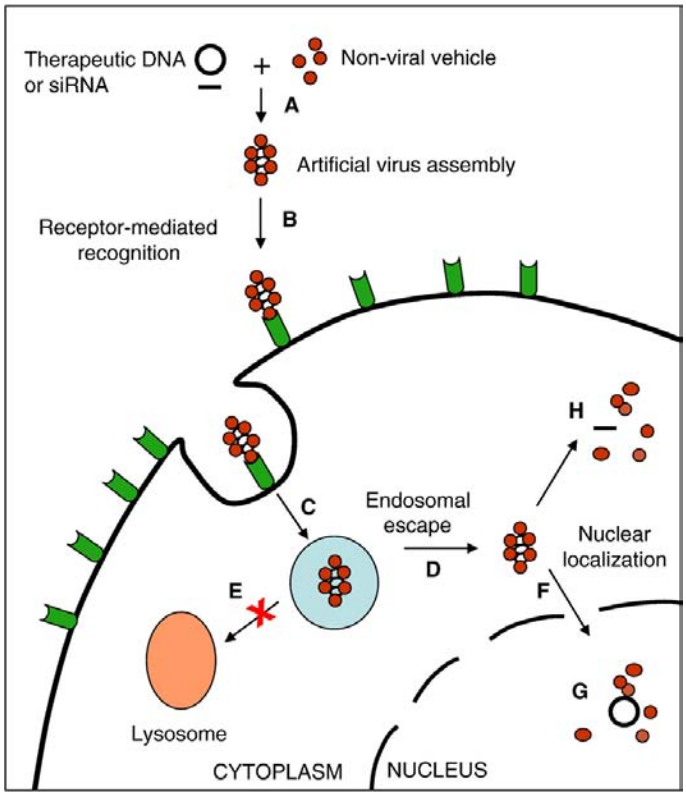


Fig. 1. Targeted cell penetration and trafficking properties of the viral cycle mimicked by an artificial virus. The coating material condenses tightly while binds reversibly the target nucleic acid, usually *in vitro* (A). The final nanocomplex recognizes specific receptors at the cell membrane and is internalized via endocytic pathways (B). Once located within early endosomes (C), the artificial virus escapes from the endosomal route (D) and thus avoids lysosomal targeting (E). Finally, the nanocomplex enters into the cell nucleus (F) and permits transgene expression (G). For the delivery of siRNA, nuclear localization signals are absent and the nucleic acid is released in the cytoplasm (H).

Table 1
Peptides used as DNA-binding agents for non-viral gene therapy.

Peptide	Amino acid sequence	Functional features	Ref.
RVG-9R	YTIWMPENPRPGTPCDIFTNS RGKRASNGGGGRRRRRRRR	Blood brain barrier crossing siRNA binding	Kumar et al. (2007)
R15	RRRRRRRRRRRRRRR	Higher DNA binding than cationic liposomes Caveolae-dependent internalization Higher transfection efficiency than commercial agents Transfection efficiency not affected by serum	Kim et al. (2003, 2007)
μ	MRRAHHRRRRASHRRMRGG	Multiple binding sites per peptide (±1300 per molec.) DNA collapse is a cooperatively peptide-mediated process	Tecle et al. (2003)
Pep V	RPRRRATTRRRTTTGTRRRRRRR	Multiple binding sites per peptide (±800 per molec.) DNA collapse is a cooperatively peptide-mediated process	Tecle et al. (2003)
Tat_{49-57} derivatives	RKKRRQRRR	High cellular uptake tendency High protection against serum nucleases	Wender et al. (2000)
POLYTAT	$(CGRKKRRQRRRGC)_n$	Higher ability to condense DNA than other Tat peptides Endosomal escape High gene expression level	Manickam et al. (2005)
L-Arg	L-R9	Higher effectiveness in cellular uptake than Lys peptides	Wender et al. (2000)
D-Arg	D-R9	Higher effectiveness in cellular uptake than Lys peptides More stable than L-arg in the presence of serum	Wender et al. (2000)
Poly-lysines	$(L)_n$	Modest transfection efficiencies High-molecular weight PLL are able to bind DNA tighter and form more stable complexes than low molecular weight	Mannisto et al. 2002; Martin and Rice 2007)
RPC	$(CL_{10}C)_n$	High endosomal escape capability when combined with the cationic lipid DOTAP Higher gene expression level than PLLs by an improved DNA release efficiency High gene expression level in the presence of serum in combination with cationic lipids	Read et al. (2003)

its condensation but also favouring the interaction of the nanoparticle with the cell membrane and the consequent internalization (Vijayanathan et al., 2002; Martin and Rice 2007). Up to now, lysines have been used in non-viral vectors as single chain polylysine peptides but also as oligolysine-containing proteins (Plank et al., 1999), thus allowing modifications of peptide length, presentation and sequence to increase transfection efficiency.

2.1. Poly-L-lysines

PLL peptides were the first non-viral system used at the end of 80s as gene therapy vectors and are the better characterized DNA-binding domains. The degree of polymerization of Lys, as a synthetic repeat amino acid chain, can range between 90 and 450 (Wadhwa et al., 1997), and leads to the formation of a polypeptide chain with an acceptable degree of biodegradability, important in terms of cell physiology and controlled release of the DNA into cell nuclei. However, the degree of polymerization has shown to be directly related to toxic effects, the longer the Lys chain the more cytotoxic PLL is (Plank et al., 1999; Lee et al., 2002a; Martin and Rice 2007). In this regard, PLLs of high-molecular weight are degraded slowly and thus are toxic to cultured cells (Putnam et al., 2001; Li and Huang 2004; Symonds et al., 2005). To reduce the associated cytotoxicity, PLLs can be conjugated with coating elements such as poly(ethylene)glycol (PEG) (Choi et al., 1998; Lee et al., 2002b; Rimann et al., 2008) or imidaloze groups (Putnam et al., 2001), balancing by these means the cationic charge density and the capability of the PLLs to bind and condense DNA. Furthermore, PLL polypeptides can be also conjugated with other functional elements such as cell ligands to enhance receptor-mediated uptake (Wagner et al., 1998; Vijayanathan et al., 2002).

In 1987, Wu and co-workers showed that PLL coupled with asialoorosomucoid (asOR) formed soluble polyplexes, enabling gene expression targeted to those cells exposing the receptor for asOR on their surface (Wu and Wu, 1987). Furthermore, they described the use of low molecular weight PLL for intravenous gene delivery to hepatocytes, both *in vivo* and *in vitro* (Wu and Wu, 1988a,b). Since PLLs alone show modest to low transfection efficiencies they have been conjugated with agents driving endosomal escape, such as chloroquine (Wolfert and Seymour, 1998; Martin and Rice, 2007). Also, to prevent the attachment of low molecular weight PLL to the plasma membrane, to increase the circulation half-life of the polyplex (Lee et al., 2002b) and to protect condensed DNA against serum components and DNase I-mediated degradation (Rimann et al., 2008), PLL has been usually combined with poly(ethylene)glycol (PEG), a non-toxic polymer suitable as coating element for *in vivo* drug delivery purposes. PEG, in addition, minimizes the cytotoxicity of PLL (Rimann et al., 2008), assumed to be associated with the length of the peptide chain (Plank et al., 1999).

Read et al. (2003) designed a linear reducible polycation (RPC) based on the oxidative polycondensation of Cys-Lys_{10}-Cys peptide. RPC was shown not only to reversibly condense DNA but also to release DNA into the intracellular environment by polyplex-controlled cleavage.

The high-molecular weight and the polydispersity of these PLLs drive to a disturbed chemical production control, which leads to the formation of heterogeneous polyplexes (Wadhwa et al., 1997) and offers an explanation for the recurrent irreproducible efficiencies of many PLL-based vectors for gene therapy. Tecle et al. (2003) partially solved the heterogeneity of nanosized DNA complexes by concatenating several cationic peptide chains, to create a polymer-like scaffold acting as a more controlled condensation domain.

2.2. Polylysine-containing peptides

PLLs can also be genetically fused to other peptides or full proteins, offering a fine control of synthesis and reducing toxicity; in addition, cell-attachment domains or additional cationic amino acids can be incorporated in this way (Gottschalk et al., 1996; McKenzie et al., 2000b; Haines et al., 2001). Furthermore, PLLs can be fused to enzymes, and the enzymatic activity is usually conserved allowing monitoring of the whole protein–DNA complex.

In this context, protein 249AL is an engineered *Escherichia coli* β-galactosidase containing a viral RGD cell-binding region and a poly (10)-lysine peptide. 249AL is capable of accommodating and condensing plasmid DNA without significant alteration on the whole construct structure and thus maintaining its enzymatic activity (Aris and Villaverde, 2000). Interestingly, 249AL has shown to efficiently deliver and promote the expression of plasmid DNA *in vitro* (Aris and Villaverde, 2000; Aris et al., 2000) and *in vivo* in a model of excitotoxic damage in postnatal rat brains (Peluffo et al., 2003). Furthermore, the addition of the SV40 virus NLS (nuclear localization signal) to the 249AL vector, resulting in the multifunctional protein NLSCt, enhances the gene

expression levels up to 30 times *in vitro* (Aris and Villaverde, 2003) and promotes the reduction of the lesioned area in model rats for brain ischemia (Peluffo et al., 2006).

Wadhwa et al. (1997) synthesized a series of CWK_n ($n=3$–36) peptides to test the influence of Lys content in both DNA condensation and transfection efficiency. They demonstrated that oligopeptides containing more than 13 Lys residues led to the formation of tight, small polyplexes sizing between 50 and 200 nm, whereas shorter peptides containing 8 or less Lys residues formed large condensates ranging from 0.7 to 3 μm that weakly bound to DNA. McKenzie et al. (1999) compared a panel of lysine-containing peptides, namely $AlkCWK_{18}$, $AlkCYK_{18}$ and K_{20}. The aim of the study was to test the possible influence of the insertion of the aromatic amino acid tyrosine and tryptophan regarding DNA-binding affinity. The authors concluded that aromatic amino acid substitution did not significantly influenced DNA binding, condensation and gene transfer efficiency. Furthermore, this group designed four variants of CWK_{20} peptide differing in the number of cysteine residues (1 to 4) inserted into the peptide. After DNA binding, cysteine residues spontaneously oxidized, and interpeptide disulfide bonds prevented DNA from dissociating leading to a smaller nanoparticle below 40 and 50 nm. Once internalized, the reducing cell cytoplasm environment enhanced the relaxation of the polyplex allowing the readily release of DNA. The gene expression level obtained with $CWK_{17}C$ was 60-fold higher than reached by $AlkCWK_{18}$ (McKenzie et al., 1999). Surprisingly, the CWK_4C permitted a gene expression level similar to that promoted by its 17-Lysines counterpart peptide, thus similar transfection efficiency was achieved with a peptide containing a shorter Lys moiety. Other variants in lysine-based peptides were those containing the sequence $YKAK_nWK$ ($n=4$–12) designed by Gottschalk et al. (1996), which were used to establish the minimum number of lysines to efficiently condensate DNA and achieve high levels of gene expression. $YKAK_8K$, which contained 10 lysines and a tryptophan, resulted to be the best variant in a variety of cell lines when assembled with a membrane-destabilizing peptide. Plank et al. (1999) proposed a model based in branched cationic peptides where the length and the type of residues varied. A lysine residue was used as the branching point to insert lysine, arginine or ornithine residues that acted as DNA-binding domains. They concluded that these branched peptides allowed the establishment of 6 lysine residues as the minimal chain length for the formation of DNA complexes able to act as receptor-mediated gene delivery vehicles. The most interesting feature of these branched peptides was that they contained a terminal glycine acting as an attachment point for effectors, ligands and stabilizers like PEG at the C-terminus.

3. Basic and arginine-rich peptides

The cationic amino acid arginine has been used in the recent years as an alternative to lysine in non-viral gene delivery systems (Tecle et al., 2003). In fact, the most used natural arginine-containing peptides as artificial virus components are the HIV Tat protein and *Drosophila* homeodomain-derived Antennapedia (Astriab-Fisher et al., 2002; Moschos et al., 2007). Antennapedia (or penetratin) and antennapedia-like amphiphilic peptides contain several lysines and arginines in their sequence, acting as both DNA-condensing agents and potent membrane active peptides that facilitate the intracellular trafficking of protein–DNA complexes, through promoting endosomal escape (Ferrer-Miralles et al., 2008). This seems to occur by the induction of a negative membrane curvature through interactions with the fluid membrane domains (Lamaziere et al., 2008). Tat contains a short arginine-rich sequence that electrostatically binds plasmid DNA and internalizes into mammalian cells through endocytosis-mediated mechanism (Ignatovich et al., 2003). POLYTAT, a biodegradable high-molecular weight form of Tat, was shown to exhibit a reversible affinity with DNA under cellular redox-potential gradients allowing controlled DNA release, endosomal escape and improving gene expression. Moreover, the high-molecular weight of the POLYTAT improved the transfection efficiencies compared to that exhibited by the control Tat peptides by an increase in binding to cell membrane, and reduced the size particle due to its improved ability to condense DNA (Manickam et al., 2005).

Kim et al. (2003) constructed a panel of arginine peptides ranging from 9 to 15 residues to identify the Tat regions useful to optimize DNA binding and gene transfer. All arginine-rich peptides showed complete DNA retardation and fully protected nucleic acid against degradation. However, transfection efficiencies positively varied depending on the number of residues, while all the candidates resulted to be non-cytotoxic and obviated serum sensitivity. Based on this data, further studies were performed using R15 peptide as a gene transfer vehicle which was able to bind to and condense DNA, forming stable nanocomplexes. R15 was able to carry fluorescently labelled DNA into the cell nuclei of mammalian cells by a caveolae-dependent pathway (Kim et al., 2007). Kumar et al. (2007) constructed a chimeric peptide by fusing R9 peptide to the carboxy terminus of rabies virus glycoprotein (RVG). This RVG-R9 vector was able to bind and transduce siRNA to neuronal cells, condensing RNA in a manner that allowed the exposure of positive charges to permit the crossing of the blood brain barrier.

4. Covalently-linked and specific-binding domains

Cationic peptides, as mentioned above, interact in a charge-depending manner with DNA. However, an antisense DNA covalently linked to a peptide vehicle can be also used as a way to mediate DNA condensation. The peptide F3, that targets specifically tumour endothelial cells, was covalently bound to an antisense sequence of Id-1, a transcriptional regulator required for tumour metastasis. The final complexes, namely Id1–PCAO, were constructed by cross-linking a C6-amino linker specially modified at the 5′ end of siRNA to F3-peptide N-Terminal cysteine through a hetero-bifunctional linker (GMBS) (Henke et al., 2008). The conjugation, and thus the presence of the anionic antisense moiety, led to higher specificity of uptake since the Id1–PCAO complexes only bound the endothelial cells with the highest concentration of receptors exposed in the cell surface (Henke et al., 2008). Antibodies offer the possibility to generate DNA–peptide complexes through specific binding. A collagen scaffold linked to an anti-DNA antibody by means of the cross-linking reagent N-succinimidyl 3-(2-pyridyldithio) propionate (SPDP) resulted to specifically bind DNA in a higher extent than the unlinked collagen scaffold. Moreover, the retention of the target DNA into the collagen scaffold, mediated by the anti-DNA antibody interaction, was sustained for 13 days and the expression of transgene turned out to be highly localized. Furthermore, efficient gene delivery by the protein matrix was observed with no detrimental effects on cultured cell growth (Jin et al., 2008).

Nevertheless, specific binding can also be achieved by precise interactions between peptides and DNA sequences. It is known that some polypeptides bind covalently to inverted terminal DNA repeats (ITRs) sequences at 5′ end of DNA *in vivo*, as it occurs in some adenoviruses, bacteriophages, bacteria and eukaryotic organisms. Tolmachov and Coutelle (2007) suggested adding ITR sequences at the ends of the therapeutic DNA, and consecutively coupling the modified gene through ITRs to a peptide carrier. Among others, the main advantages of this approach are the high specificity in DNA binding and the protection against endonucleases of the 5′ end of the therapeutic gene. In this context, the TetR protein binds specifically to the short DNA sequence *tetO*. Vaysse et al. (2004) took advantage of this specific-binding feature to construct a chimeric protein by fusing the TetR protein and the SV40-derived NLS peptide. This construct, named TetR-NLS, bound specifically to a therapeutic gene previously fused to the *tetO* sequence. The fusion of TetR to NLS sequence did not alter the interaction between the carrier and the transgene. Furthermore, TetR-

NLS was shown to bind and efficiently transfer all *tetO*-containing DNAs, and the expression levels of the therapeutic genes were high even in non dividing cells.Facchini and Lingwood (2001) used the same strategy and fused the phage lambda Cro monomer, the smallest known DNA-binding protein, with the Verotoxine I B subunit (VTB), which naturally binds to the cell surface glycolipid, globotriaosylceramide (Gb_3). VTB-Cro carrier efficiently bound and condensed the 25-bp DNA fragment containing the consensus Cro operator, reaching the Gb_3 positive cell nucleus by retrograde transit via the Golgi and thus protecting condensed DNA against cytosolic nucleases. The yeast transcription factor Gal4 has been also widely used as a specific-binding domain to condense DNA. For instance, GD5 is a designed peptide composed by the DNA-binding domain of Gal4 connected to an antibody fragment specific for the tumour-associated ErbB2 antigen via an internal diphtheria toxin (DT) translocation domain. GD5 allowed specific binding and condensation of plasmid DNA containing the Gal4 recognition motif and these complexes bound specifically ErbB2-expressing cells. Furthermore, the internal DT translocation domain acted as an endosomal escape motif (Uherek et al., 1998). A variant of Gal4 containing residues comprised between 1 and 147, namely Gal4(1–147), was able to specifically bind to plasmid DNA containing the Gal4 recognition domain and to enhance its transfection into cells. Interestingly, the DNA-binding domain and nuclear localization signal of the Gal4 are located in the same region of the sequence, and they have been shown to be activities mutually exclusive. Therefore, the enhancement of the transfection efficiency was shown not to be mediated through NLS nuclear targeting but through DNA compaction or DNA protection against cellular degradation (Chan et al., 1998).

5. Protein scaffolds

The development of temporary protein platforms, mainly collagen scaffolds that lodge therapeutic DNA, could be useful in tissue engineering, especially in those cases in which the transgenic expression must be prolonged and localized (Fig. 2). Scaffolds construction and DNA cross-linking methods can influence the efficiency of plasmid attachment, release and protection.

Cohen-Sacks et al. (2004) developed a facile, non-viral and biocompatible method to embed cationic pDNA carriers in collagen matrices. The authors used two PLLs at dp of 9 and 99 residues as DNA-condensing agents, which were cross-linked to the collagen scaffold prior to pDNA binding. PLL-pDNA matrices presented high DNA-binding efficiencies; in fact, increasing the molecular weight of PLL, a higher binding to collagen scaffold efficiency was achieved, and thus also a higher pDNA embedding efficiency. On the contrary, Capito and Spector (2007) developed two methods for DNA binding to collagen scaffolds without the mediation of any condensing cationic domain. DNA could be embedded through soaking–freeze/drying process, or directly linked to the protein scaffold. When using the first approach, a high initial incorporation of DNA was achieved although after 2 weeks of drying step almost the 80% of the plasmid was released and only about 20% was retained by the scaffold. Furthermore, 2 weeks after the implantation, only about 12% of the retained DNA was released in the tissue, mainly during the first days. When taking the second approach, incorporation efficiencies were lower than those observed by the first method. However the DNA release was slower and gradual. In both cases the plasmid integrity was demonstrated by restriction enzyme protocols.

A similar approach in scaffold-based gene delivery systems was developed using a derivative of TraI protein (*E. coli* helicase). It consisted on the insertion in a permissive site of TraI of a cysteine-constrained 12-residue Cu_2O binding sequence, namely CN225. The chimeric protein, TraIi1753::CN225, kept its Cu_2O binding ability. Therefore, in the presence of the electrolyte and DNA, the protein formed a shell surrounding a Cu_2O core with DNA-binding properties, that self-assembled in nanoparticles onto circular dsDNA (Dai et al., 2005).

6. Future prospects

DNA–protein interactions result from several forces including hydrogen bond formation, hydrophobic or electrostatic interactions

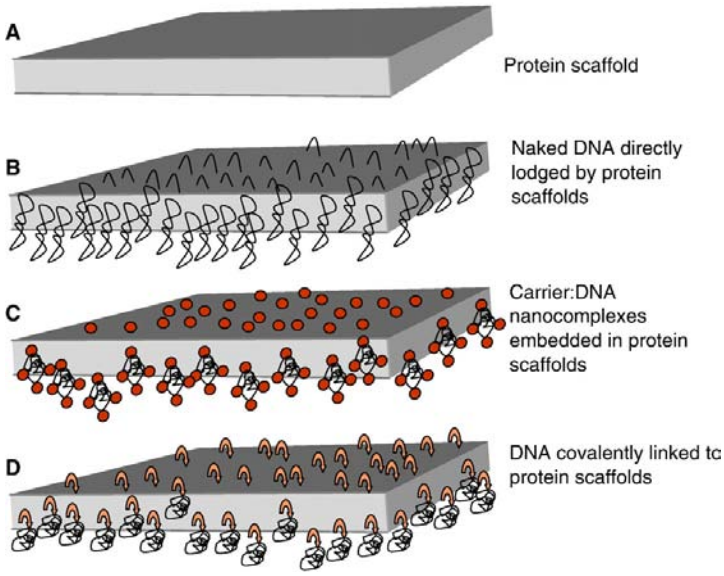


Fig. 2. Gene delivery systems based in protein scaffolds. A. A protein matrix acts as a scaffold to lodge the therapeutic nucleic acid. B. Naked DNA is directly embedded into the protein scaffold through soaking–freeze/drying processes. C. DNA is lodged by the protein scaffold by means of embedding a DNA:condensing carrier. This procedure can be achieved by two strategies: (i) cross-linking the condensing carrier to the scaffold prior to DNA condensation, (ii) embedding condensing carrier:DNA into protein scaffold once the nanocomplex is formed. D. Naked DNA is covalently linked to the protein scaffold. These systems allow prolonged and localized target gene expression in cell culture and tissues.

and water extrusion effects. Correlative and structural information as well as conservative comparisons obtained by X-ray crystallography, NMR and genome sequences, suggests that there is a statistical correlation between base pairs interaction and their evolutionary conservation. Changes in the nature and number of interactions are generally due to alterations of DNA sequences (Mirny and Gelfand, 2002). Other studies reveal that the interactions between peptides and nucleic acids do not follow any specific dynamics. Actually, peptide–DNA interactions and condensation resemble to that found in protein folding processes, in which there are multiple ways for reaching the native, correct conformation of the protein. Thus, peptide–DNA interaction and DNA condensation, in terms of folding, is a rather unpredictable process (Tecle et al., 2003). These data prompt to design novel peptides in which conformational dynamics and structural features would permit a more standardizable linkage process and a higher efficiency in gene delivery. This would allow a faster development of artificial viruses for gene therapy and of other nanosized manmade drugs, by means of a more rational design of their specific components that are hold the cargo nucleic acid.

Acknowledgements

The authors are in debt with Neus Ferrer-Miralles and Esther Vazquez for carefully reading the manuscript and helpful comments. The authors appreciate the financial support to their research on gene delivery through grants BIO2007-61194 (MEC) and 2005SGR-00956 (AGAUR), and from the CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Spain. NGM is recipient of a fellowship from MEC, Spain. AV has been supported by ICREA (Generalitat de Catalunya) through an ICREA ACADEMIA award.

References

Abbott A. Questions linger about unexplained gene-therapy trial death. Nat Med 2006;12:597.
Aris A, Villaverde A. Molecular organization of protein–DNA complexes for cell-targeted DNA delivery. Biochem Biophys Res Commun 2000;278:455–61.
Aris A, Villaverde A. Engineering nuclear localization signals in modular protein vehicles for gene therapy. Biochem Biophys Res Commun 2003;304:625–31.
Aris A, Villaverde A. Modular protein engineering for non-viral gene therapy. Trends Biotechnol 2004;22:371–7.
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 2000;68:689–96.
Astriab-Fisher A, Sergueev D, Fisher M, Shaw BR, Juliano RL. Conjugates of antisense oligonucleotides with the Tat and antennapedia cell-penetrating peptides: effects on cellular uptake, binding to target sequences, and biologic actions. Pharm Res 2002;19:744–54.
Bloomfield VA. DNA condensation. Curr Opin Struct Biol 1996;6:334–41.
Bloomfield VA. DNA condensation by multivalent cations. Biopolymers 1997;44:269–82.
Brewer LR, Corzett M, Balhorn R. Protamine-induced condensation and decondensation of the same DNA molecule. Science 1999;286:120–3.
Capito RM, Spector M. Collagen scaffolds for nonviral IGF-1 gene delivery in articular cartilage tissue engineering. Gene Ther 2007;14:721–32.
Chan CK, Hubner S, Hu W, Jans DA. Mutual exclusivity of DNA binding and nuclear localization signal recognition by the yeast transcription factor GAL4: implications for nonviral DNA delivery. Gene Ther 1998;5:1204–12.
Choi YH, Liu F, Kim JS, Choi YK, Park JS, Kim SW. Polyethylene glycol-grafted poly-L-lysine as polymeric gene carrier. J Control Release 1998;54:39–48.
Cohen-Sacks H, Elazar V, Gao J, Golomb A, Adwan H, Korchov N, et al. Delivery and expression of pDNA embedded in collagen matrices. J Control Release 2004;95:309–20.
Dai H, Choe WS, Thai CK, Sarikaya M, Traxler BA, Baneyx F, et al. Nonequilibrium synthesis and assembly of hybrid inorganic-protein nanostructures using an engineered DNA binding protein. J Am Chem Soc 2005;127:15637–43.
Edelstein ML, Abedi MR, Wixon J. Gene therapy clinical trials worldwide to 2007—an update. J Gene Med 2007;9:833–42.
Facchini LM, Lingwood CA. A verotoxin 1 B subunit-lambda CRO chimeric protein specifically binds both DNA and globotriaosylceramide (Gb(3)) to effect nuclear targeting of exogenous DNA in Gb(3) positive cells. Exp Cell Res 2001;269:117–29.
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.
Gottschalk S, Sparrow JT, Hauer J, Mims MP, Leland FE, Woo SL, et al. A novel DNA-peptide complex for efficient gene transfer and expression in mammalian cells. Gene Ther 1996;3:448–57.
Haines AM, Irvine AS, Mountain A, Charlesworth J, Farrow NA, Husain RD, et al. CL22 — a novel cationic peptide for efficient transfection of mammalian cells. Gene Ther 2001;8:99–110.
He S, Arscott PG, Bloomfield VA. Condensation of DNA by multivalent cations: experimental studies of condensation kinetics. Biopolymers 2000;53:329–41.
Henke E, Perk J, Vider J, de CP, Chin Y, Solit DB, et al. Peptide–conjugated antisense oligonucleotides for targeted inhibition of a transcriptional regulator in vivo. Nat Biotechnol 2008;26:91–100.
Ignatovich IA, Dizhe EB, Pavlotskaya AV, Akifiev BN, Burov SV, Orlov SV, et al. Complexes of plasmid DNA with basic domain 47–57 of the HIV-1 Tat protein are transferred to mammalian cells by endocytosis-mediated pathways. J Biol Chem 2003;278:42625–36.
Jin X, Mei L, Song C, Liu L, Leng X, Sun H, et al. Immobilization of plasmid DNA on an anti-DNA antibody modified coronary stent for intravascular site-specific gene therapy. J Gene Med 2008;10:421–9.
Keller M, Harbottle RP, Perouzel E, Colin M, Shah I, Rahim A, et al. Nuclear localisation sequence templated nonviral gene delivery vectors: investigation of intracellular trafficking events of LMD and LD vector systems. Chembiochem 2003;4:286–98.
Kim HH, Lee WS, Yang JM, Shin S. Basic peptide system for efficient delivery of foreign genes. Biochim Biophys Acta 2003;1640:129–36.
Kim HH, Choi HS, Yang JM, Shin S. Characterization of gene delivery in vitro and in vivo by the arginine peptide system. Int J Pharm 2007;335:70–8.
Kumar P, Wu H, McBride JL, Jung KE, Kim MH, Davidson BL, et al. Transvascular delivery of small interfering RNA to the central nervous system. Nature 2007;448:39–43.
Lamaziere A, Wolf C, Lambert O, Chassaing G, Trugnan G, yala-Sanmartin J. The homeodomain derived peptide Penetratin induces curvature of fluid membrane domains. PLoS ONE 2008;3:e1938.
Lechardeur D, Sohn KJ, Haardt M, Joshi PB, Monck M, Graham RW, et al. Metabolic instability of plasmid DNA in the cytosol: a potential barrier to gene transfer. Gene Ther 1999;6:482–97.
Lee CJ, Paik SH, Ko KH, Kim KC. Effects of polycationic peptides on mucin release from airway goblet cells: relationship between polymer size and activity. Inflamm Res 2002a;51:490–4.
Lee H, Jeong JH, Park TG. PEG grafted polylysine with fusogenic peptide for gene delivery: high transfection efficiency with low cytotoxicity. J Control Release 2002b;79:283–91.
Li Z, Huang L. Sustained delivery and expression of plasmid DNA based on biodegradable polyester, poly(D,L-lactide-co-4-hydroxy-L-proline). J Control Release 2004;98:437–46.
Manickam D, Bisht H, Wan L, Mao G, Oupicky D. Influence of TAT-peptide polymerization on properties and transfection activity of TAT/DNA polyplexes. J Control Release 2005;102:293–306.
Mannisto M, Vanderkerken S, Toncheva V, Elomaa M, Ruponen M, Schacht E, et al. Structure–activity relationships of poly(L-lysines): effects of pegylation and molecular shape on physicochemical and biological properties in gene delivery. J Control Release 2002;83:169–82.
Marshall E. Gene therapy. What to do when clear success comes with an unclear risk? Science 2002;298:510–1.
Marshall E. Gene therapy. Second child in French trial is found to have leukemia. Science 2003;299:320.
Martin ME, Rice KG. Peptide-guided gene delivery. AAPS J 2007;9:E18–29.
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.
McKenzie DL, Collard WT, Rice KG. Comparative gene transfer efficiency of low molecular weight polylysine DNA-condensing peptides. J Pept Res 1999;54:311–8.
McKenzie DL, Kwok KY, Rice KG. A potent new class of reductively activated peptide gene delivery agents. J Biol Chem 2000a;275:9970–7.
McKenzie DL, Smiley E, Kwok KY, Rice KG. Low molecular weight disulfide cross-linking peptides as nonviral gene delivery carriers. Bioconjug Chem 2000b;11:901–9.
Mirny LA, Gelfand MS. Structural analysis of conserved base pairs in protein–DNA complexes. Nucleic Acids Res 2002;30:1704–11.
Moschos SA, Jones SW, Perry MM, Williams AE, Erjefalt JS, Turner JJ, et al. Lung delivery studies using siRNA conjugated to TAT(48–60) and penetratin reveal peptide induced reduction in gene expression and induction of innate immunity. Bioconjug Chem 2007;18:1450–9.
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 2003;14:1215–23.
Peluffo H, Acarin L, Aris A, Gonzalez P, Villaverde A, Castellano B, 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 2006;7:35.
Plank C, Tang MX, Wolfe AR, Szoka Jr FC. Branched cationic peptides for gene delivery: role of type and number of cationic residues in formation and in vitro activity of DNA polyplexes. Hum Gene Ther 1999;10:319–32.
Putnam D, Gentry CA, Pack DW, Langer R. Polymer-based gene delivery with low cytotoxicity by a unique balance of side-chain termini. Proc Natl Acad Sci U S A 2001;98:1200–5.
Read ML, Bremner KH, Oupicky D, Green NK, Searle PF, Seymour LW. Vectors based on reducible polycations facilitate intracellular release of nucleic acids. J Gene Med 2003;5:232–45.
Rimann M, Luhmann T, Textor M, Guerino B, Ogier J, Hall H. Characterization of PLL-g-PEG-DNA nanoparticles for the delivery of therapeutic DNA. Bioconjug Chem 2008;19:548–57.
Symonds P, Murray JC, Hunter AC, Debska G, Szewczyk A, Moghimi SM. Low and high molecular weight poly(L-lysine)S/poly(L-lysine)-DNA complexes initiate mitochondrial-mediated apoptosis differently. FEBS Lett 2005;579:6191–8.

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 2003;42:10343–7.
Tolmachov O, Coutelle C. Covalent attachment of multifunctional chimeric terminal proteins to 5′ DNA ends: a potential new strategy for assembly of synthetic therapeutic gene vectors. Med Hypotheses 2007;68:328–31.
Uherek C, Fominaya J, Wels W. A modular DNA carrier protein based on the structure of diphtheria toxin mediates target cell-specific gene delivery. J Biol Chem 1998;273:8835–41.
Vaysse L, Harbottle R, Bigger B, Bergau A, Tolmachov O, Coutelle C. Development of a self-assembling nuclear targeting vector system based on the tetracycline repressor protein. J Biol Chem 2004;279:5555–64.
Vijayanathan V, Thomas T, Thomas TJ. DNA nanoparticles and development of DNA delivery vehicles for gene therapy. Biochemistry 2002;41:14085–94.
Wadhwa MS, Collard WT, Adami RC, McKenzie DL, Rice KG. Peptide-mediated gene delivery: influence of peptide structure on gene expression. Bioconjug Chem 1997;8:81–8.
Wagner E, Ogris M, Zauner W. Polylysine-based transfection systems utilizing receptor-mediated delivery. Adv Drug Deliv Rev 1998;30:97–113.
Wender PA, Mitchell DJ, Pattabiraman K, Pelkey ET, Steinman L, Rothbard JB. The design, synthesis, and evaluation of molecules that enable or enhance cellular uptake: peptoid molecular transporters. Proc Natl Acad Sci U S A 2000;97:13003–8.
Wolfert MA, Seymour LW. Chloroquine and amphipathic peptide helices show synergistic transfection in vitro. Gene Ther 1998;5:409–14.
Wu GY, Wu CH. Receptor-mediated in vitro gene transformation by a soluble DNA carrier system. J Biol Chem 1987;262:4429–32.
Wu GY, Wu CH. Evidence for targeted gene delivery to Hep G2 hepatoma cells in vitro. Biochemistry 1988a;27:887–92.
Wu GY, Wu CH. Receptor-mediated gene delivery and expression in vivo. J Biol Chem 1988b;263:14621–4.
Zabner J, Fasbender AJ, Moninger T, Poellinger KA, Welsh MJ. Cellular and molecular barriers to gene transfer by a cationic lipid. J Biol Chem 1995;270:18997–9007.

Annex 3

Summary of the most relevant protein modifications.

Chemical groups	
Phosphorylation	Phosporylation is a reversible PTM. It occurs in serine, threonine and tyrosine amino acids and causes a conformational change in proteins. Because of that it's involved in protein function regulation of catalytic activity and cell signaling. The reaction is mediated by Kinase and phosphatase enzymes, which add or remove phosphate groups, respectively[273].
Methylation	Methyltransferases are the enzymes responsible for the transfer of one carbon-methyl group from S-adenosyl methionine (SAM) into amino acidic nitrogen or oxygen. These reactions permit to increase hydrophobicity of the peptide chain and it's known to be one of the most important regulatory epigenetic mechanisms. Histonemethylation, in fact, is the responsible of DNA availability for transcription. Single or multiple methyl groups can be conjugated to a single amino acid. The reaction is irreversible, except in rare cases of O-methylation[274].
N-Acetylation	N-terminal acetylation of proteins is catalyzed by methionine aminopeptidase (MAP) that cuts the initial methionine, and by N-acetyl transferase (NAT) enzyme, which replace the methionine with an acetyl group from Acetyl-CoA. The reaction is sometimes reversible and it occurs when the peptide chain is still growing and attached to ribosome. Like methylation, these reactions are involved in chromosomal condensation, being histone conformation sensible to acetylation[275, 276].

Complex molecules	
Glycosylation	Glycosylation consists in the attachment of sugar moieties to proteins. About half of produced proteins undergo these PTMs, and it has a massive effect on folding, activity and stability. The reactions take part in the Endoplasmatic Reticulum and in Golgi and involve a N- O- and C- linked glycosylation, Glypiation (GPI) and phosphoglycosytation. These reactions are involved in the protein quality control mechanism and in subcellular distribution targeting. Moreover, sugars can act as ligands in cell surface in case of surface exposed proteins[277].

Isoprenylation	It refers to a lipid modification, which is involved in protein-membrane and protein-protein interaction. There are different kinds of lipids that can be linked to the C terminus cysteine: 15- carbonyl farnesyl and 20-carbon geranylgeranyl are the most common. Isoprenyl transferase enzymes are the responsible of the covalent attachment of isoprenyl-derived lipids through a thioether linkage to the cysteine[278]. Other kind of lipid modifications are involved in protein targeting to membranes in organelles (i.e. GPI anchors, N-terminal myristoylation, S-palmtoylation)[279-281].

Peptides/Proteins	
Ubiquitination	Ubiquitin is a 76 amino acid polypeptide, which is added to targeted proteins by ubiquitin activating/conjugating/ligases enzymes. Single or multiple residues of ubiquitin can be attached to the epsilon amino group of protein's lysine. This PTM is involved in the degradation of proteins being ubiquitin chain recognized by proteasomes, which catalyze protein degradation. Ubiquitination is also involved in protein-protein interaction, sub-cellular protein localization and activity[282].
Sumoylation	Small ubiquitin-like modifier (SUMO) proteins are involved in several mechanism like nuclear-cytosolic transport, apoptosis, protein stability, transcriptional regulation, cell cycle progression but they are not involved in protein degradation[283]. As Ubiquitne, SUMO proteins are generally small (12 KDa in mass) and they are linked to the targeted protein in lysines[284].

Cleavage	
Proteolysis	Proteolysis is an irreversible and thermodynamically stable reaction. There are 11.000 different enzymes involved in protein peptide bonds cleavage. This process is involved in two fundamental steps: the first one for removing unassembled or misfolded protein subunits, and the second one in activation of Zymogens, which require specific cleavage for acquire enzyme function. Proteolysis activity is tightly regulated by a temporal/spatial control mechanism in order to avoid uncontrolled degradation reactions[285].

Annex 4

Other publications:

María Luciana Negro-Demontel, **Paolo Saccardo**, Cecilia Giacomini, Rafael Joaquín Yáñez-Muñoz, Neus Ferrer-Miralles, Esther Vazquez, Antonio Villaverde, Hugo Peluffo. **Comparative analysis of lentiviral vectors and modular protein nanovectors for traumatic brain injury gene therapy.** Molecular Therapy - Methods & Clinical Development 1, Article number: 14047, Published online 15 October 2014.

Unzueta U, Ferrer-Miralles N, Cedano J, Zikung X, Pesarrodona M, **Saccardo P**, García-Fruitós E, Domingo-Espín J, Kumar P, Gupta KC, Mangues R, Villaverde A, Vazquez E**. Non-amyloidogenic peptide tags for the regulatable self-assembling of protein-only nanoparticles.** Biomaterials. 2012 Nov;33(33):8714-22.

Domingo-Espín J, Petegnief V, de Vera N, Conchillo-Solé O, **Saccardo P**, Unzueta U, Vazquez E, Cedano J, Negro L, Daura X, Peluffo H, Planas AM, Villaverde A, Ferrer-Miralles N. **RGD-based cell ligands for cell-targeted drug delivery act as potent trophic factors.** Nanomedicine. 2012 Nov;8(8):1263-6.

Domingo-Espín J, Unzueta U, **Saccardo P**, Rodríguez-Carmona E, Corchero JL, Vázquez E, Ferrer-Miralles N.
Engineered biological entities for drug delivery and gene therapy protein nanoparticles. Prog Mol Biol Transl Sci.

References

1. Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171, 737-738.
2. Lehman, I.R., Bessman, M.J., Simms, E.S., and Kornberg, A. (1958). Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from Escherichia coli. The Journal of biological chemistry 233, 163-170.
3. Matthaei, J.H., Jones, O.W., Martin, R.G., and Nirenberg, M.W. (1962). Characteristics and composition of RNA coding units. Proceedings of the National Academy of Sciences of the United States of America 48, 666-677.
4. Jackson, D.A., Symons, R.H., and Berg, P. (1972). Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America 69, 2904-2909.
5. Sanger, F., and Coulson, A.R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal of molecular biology 94, 441-448.
6. Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harbor symposia on quantitative biology 51 Pt 1, 263-273.
7. Johnson, I.S. (1983). Human insulin from recombinant DNA technology. Science 219, 632-637.
8. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature 409, 860-921.
9. Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. (2001). The sequence of the human genome. Science 291, 1304-1351.
10. Elzoghby, A.O., Samy, W.M., and Elgindy, N.A. (2012). Protein-based nanocarriers as promising drug and gene delivery systems. Journal of controlled release : official journal of the Controlled Release Society 161, 38-49.
11. Brondyk, W.H. (2009). Selecting an appropriate method for expressing a recombinant protein. Methods in enzymology 463, 131-147.
12. Berg, J., Tymoczko, J., and Stryer, L. (2002). Eukaryotic Protein Synthesis Differs from Prokaryotic Protein Synthesis Primarily in Translation Initiation. Biochemistry 5th edition Section 29.5.
13. Scientific, T. In. (
14. Sahdev, S., Khattar, S.K., and Saini, K.S. (2008). Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. Molecular and cellular biochemistry 307, 249-264.
15. Kalyanpur, M. (2002). Downstream processing in the biotechnology industry. Molecular biotechnology 22, 87-98.
16. Gao, B., and Tsan, M.F. (2003). Endotoxin contamination in recombinant human heat shock protein 70 (Hsp70) preparation is responsible for the induction of tumor necrosis factor alpha release by murine macrophages. The Journal of biological chemistry 278, 174-179.
17. Erickson, P.R., and Herzberg, M.C. (1993). Evidence for the covalent linkage of carbohydrate polymers to a glycoprotein from Streptococcus sanguis. The Journal of biological chemistry 268, 23780-23783.
18. Sorensen, H.P., and Mortensen, K.K. (2005). Advanced genetic strategies for recombinant protein expression in Escherichia coli. Journal of biotechnology 115, 113-128.
19. Huang, C.J., Lin, H., and Yang, X. (2012). Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements. Journal of industrial microbiology & biotechnology 39, 383-399.

20. Ferrer-Miralles, N., Domingo-Espin, J., Corchero, J.L., Vazquez, E., and Villaverde, A. (2009). Microbial factories for recombinant pharmaceuticals. Microbial cell factories 8, 17.
21. Sezonov, G., Joseleau-Petit, D., and D'Ari, R. (2007). Escherichia coli physiology in Luria-Bertani broth. Journal of bacteriology 189, 8746-8749.
22. Shiloach, J., and Fass, R. (2005). Growing E. coli to high cell density--a historical perspective on method development. Biotechnology advances 23, 345-357.
23. Yang, L.C., Zhu, Z., and Yang, X.G. (2003). Purification and immunity analysis of recombinant 6His-HPT protein expressed in E. coli. Biomedical and environmental sciences : BES 16, 149-156.
24. Srichaisupakit, A., Ohashi, T., and Fujiyama, K. (2014). Identification of a protein glycosylation operon from Campylobacter jejuni JCM 2013 and its heterologous expression in Escherichia coli. Journal of bioscience and bioengineering 118, 256-262.
25. Marisch, K., Bayer, K., Cserjan-Puschmann, M., Luchner, M., and Striedner, G. (2013). Evaluation of three industrial Escherichia coli strains in fed-batch cultivations during high-level SOD protein production. Microbial cell factories 12, 58.
26. Terpe, K. (2006). Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Applied microbiology and biotechnology 72, 211-222.
27. Baneyx, F. (1999). Recombinant protein expression in Escherichia coli. Current opinion in biotechnology 10, 411-421.
28. Kurland, C., and Gallant, J. (1996). Errors of heterologous protein expression. Current opinion in biotechnology 7, 489-493.
29. McNulty, D.E., Claffee, B.A., Huddleston, M.J., Porter, M.L., Cavnar, K.M., and Kane, J.F. (2003). Mistranslational errors associated with the rare arginine codon CGG in Escherichia coli. Protein expression and purification 27, 365-374.
30. Deutscher, M.P. (2006). Degradation of RNA in bacteria: comparison of mRNA and stable RNA. Nucleic acids research 34, 659-666.
31. Rauhut, R., and Klug, G. (1999). mRNA degradation in bacteria. FEMS microbiology reviews 23, 353-370.
32. O'Hara, E.B., Chekanova, J.A., Ingle, C.A., Kushner, Z.R., Peters, E., and Kushner, S.R. (1995). Polyadenylylation helps regulate mRNA decay in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America 92, 1807-1811.
33. Cao, G.J., Kalapos, M.P., and Sarkar, N. (1997). Polyadenylated mRNA in Escherichia coli: modulation of poly(A) RNA levels by polynucleotide phosphorylase and ribonuclease II. Biochimie 79, 211-220.
34. Brawner, M.E. (1994). Advances in heterologous gene expression by Streptomyces. Current opinion in biotechnology 5, 475-481.
35. Gutierrez, J., Bourque, D., Criado, R., Choi, Y.J., Cintas, L.M., Hernandez, P.E., and Miguez, C.B. (2005). Heterologous extracellular production of enterocin P from Enterococcus faecium P13 in the methylotrophic bacterium Methylobacterium extorquens. FEMS microbiology letters 248, 125-131.
36. Desplancq, D., Bernard, C., Sibler, A.P., Kieffer, B., Miguet, L., Potier, N., Van Dorsselaer, A., and Weiss, E. (2005). Combining inducible protein overexpression with NMR-grade triple isotope labeling in the cyanobacterium Anabaena sp. PCC 7120. BioTechniques 39, 405-411.
37. Hansson, M., Samuelson, P., Nguyen, T.N., and Stahl, S. (2002). General expression vectors for Staphylococcus carnosus enabled efficient production of the outer membrane protein A of Klebsiella pneumoniae. FEMS microbiology letters 210, 263-270.
38. Schneider, J.C., Jenings, A.F., Mun, D.M., McGovern, P.M., and Chew, L.C. (2005). Auxotrophic markers pyrF and proC can replace antibiotic markers on protein production plasmids in high-cell-density Pseudomonas fluorescens fermentation. Biotechnology progress 21, 343-348.

39. Wang, Y.C., Peterson, S.E., and Loring, J.F. (2014). Protein post-translational modifications and regulation of pluripotency in human stem cells. Cell research 24, 143-160.
40. Smith, G.E., Summers, M.D., and Fraser, M.J. (1983). Production of human beta interferon in insect cells infected with a baculovirus expression vector. Molecular and cellular biology 3, 2156-2165.
41. Kost, T.A., Condreay, J.P., and Jarvis, D.L. (2005). Baculovirus as versatile vectors for protein expression in insect and mammalian cells. Nature biotechnology 23, 567-575.
42. Monteiro, F., Carinhas, N., Carrondo, M.J., Bernal, V., and Alves, P.M. (2012). Toward system-level understanding of baculovirus-host cell interactions: from molecular fundamental studies to large-scale proteomics approaches. Frontiers in microbiology 3, 391.
43. Smith, G.E., Fraser, M.J., and Summers, M.D. (1983). Molecular Engineering of the Autographa californica Nuclear Polyhedrosis Virus Genome: Deletion Mutations Within the Polyhedrin Gene. Journal of virology 46, 584-593.
44. Carpentier, D.J., and King, L. (2009). The long road to understanding the baculovirus P10 protein. Virologica Sinica 24, 227-242.
45. Taylor, A.L., Haze-Filderman, A., Blumenfeld, A., Shay, B., Dafni, L., Rosenfeld, E., Leiser, Y., Fermon, E., Gruenbaum-Cohen, Y., and Deutsch, D. (2006). High yield of biologically active recombinant human amelogenin using the baculovirus expression system. Protein expression and purification 45, 43-53.
46. Moscardi, F., de Souza, M., de Castro, M., Lara Moscardi, M., and Szewczyk, B. (2011). Baculovirus Pesticides: Present State and Future Perspectives. In Microbes and Microbial Technology, I. Ahmad, F. Ahmad, andJ. Pichtel, eds. (Springer New York), pp 415-445.
47. Miele, S.A., Garavaglia, M.J., Belaich, M.N., and Ghiringhelli, P.D. (2011). Baculovirus: molecular insights on their diversity and conservation. International journal of evolutionary biology 2011, 379424.
48. Fraser, M.J. (1986). Ultrastructural observations of virion maturation in Autographa californica nuclear polyhedrosis virus infected Spodoptera frugiperda cell cultures. Journal of Ultrastructure and Molecular Structure Research 95, 189-195.
49. Luckow, V.A., Lee, S.C., Barry, G.F., and Olins, P.O. (1993). Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in Escherichia coli. Journal of virology 67, 4566-4579.
50. van Oers, M.M. (2011). Opportunities and challenges for the baculovirus expression system. Journal of invertebrate pathology 107 Suppl, S3-15.
51. Liu, F., Wu, X., Li, L., Liu, Z., and Wang, Z. (2013). Use of baculovirus expression system for generation of virus-like particles: successes and challenges. Protein expression and purification 90, 104-116.
52. Joosten, C.E., and Shuler, M.L. (2003). Production of a sialylated N-linked glycoprotein in insect cells: role of glycosidases and effect of harvest time on glycosylation. Biotechnology progress 19, 193-201.
53. Au, S., Wu, W., and Pante, N. (2013). Baculovirus nuclear import: open, nuclear pore complex (NPC) sesame. Viruses 5, 1885-1900.
54. O'reilly, D.R., Miller, L.K., and Luckov, V.A. (1994). Baculovirus expression vectors, a laboratory manual.(Oxford University Press).
55. Washburn, J.O., Chan, E.Y., Volkman, L.E., Aumiller, J.J., and Jarvis, D.L. (2003). Early synthesis of budded virus envelope fusion protein GP64 enhances Autographa californica multicapsid nucleopolyhedrovirus virulence in orally infected Heliothis virescens. Journal of virology 77, 280-290.
56. Keddie, B.A., Aponte, G.W., and Volkman, L.E. (1989). The pathway of infection of Autographa californica nuclear polyhedrosis virus in an insect host. Science 243, 1728-1730.

57. Wang, P., and Granados, R.R. (1997). An intestinal mucin is the target substrate for a baculovirus enhancin. Proceedings of the National Academy of Sciences of the United States of America 94, 6977-6982.
58. Horton, H.M., and Burand, J.P. (1993). Saturable attachment sites for polyhedron-derived baculovirus on insect cells and evidence for entry via direct membrane fusion. Journal of virology 67, 1860-1868.
59. Ohkawa, T., Volkman, L.E., and Welch, M.D. (2010). Actin-based motility drives baculovirus transit to the nucleus and cell surface. The Journal of cell biology 190, 187-195.
60. Passarelli, A.L., and Miller, L.K. (1993). Three baculovirus genes involved in late and very late gene expression: ie-1, ie-n, and lef-2. Journal of virology 67, 2149-2158.
61. Blissard, G.W., and Wenz, J.R. (1992). Baculovirus gp64 envelope glycoprotein is sufficient to mediate pH-dependent membrane fusion. Journal of virology 66, 6829-6835.
62. Volkman, L.E. (1997). Nucleopolyhedrovirus interactions with their insect hosts. Advances in virus research 48, 313-348.
63. Contreras-Gomez, A., Sanchez-Miron, A., Garcia-Camacho, F., Molina-Grima, E., and Chisti, Y. (2014). Protein production using the baculovirus-insect cell expression system. Biotechnology progress 30, 1-18.
64. Merrington, C.L., Bailey, M.J., and Possee, R.D. (1997). Manipulation of baculovirus vectors. Molecular biotechnology 8, 283-297.
65. Possee, R.D., Hitchman, R.B., Richards, K.S., Mann, S.G., Siaterli, E., Nixon, C.P., Irving, H., Assenberg, R., Alderton, D., Owens, R.J., et al. (2008). Generation of baculovirus vectors for the high-throughput production of proteins in insect cells. Biotechnology and bioengineering 101, 1115-1122.
66. Haase, S., Ferrelli, L., Pidre, M.L., and (, V.R. (2013). Genetic Engineering of Baculoviruses, Current Issues in Molecular Virology - Viral Genetics and Biotechnological Applications.
67. Kitts, P.A., Ayres, M.D., and Possee, R.D. (1990). Linearization of baculovirus DNA enhances the recovery of recombinant virus expression vectors. Nucleic acids research 18, 5667-5672.
68. Vialard, J.E., and Richardson, C.D. (1993). The 1,629-nucleotide open reading frame located downstream of the Autographa californica nuclear polyhedrosis virus polyhedrin gene encodes a nucleocapsid-associated phosphoprotein. Journal of virology 67, 5859-5866.
69. Kitts, P.A., and Possee, R.D. (1993). A method for producing recombinant baculovirus expression vectors at high frequency. BioTechniques 14, 810-817.
70. Jarvis, D.L. (2009). Baculovirus-insect cell expression systems. Methods in enzymology 463, 191-222.
71. Lynn, D.E. (2007). Available lepidopteran insect cell lines. Methods Mol Biol 388, 117-138.
72. Aucoin, M.G., Mena, J.A., and Kamen, A.A. (2010). Bioprocessing of baculovirus vectors: a review. Current gene therapy 10, 174-186.
73. van Oers, M.M., and Lynn, D.E. (2001). Insect Cell Culture. In eLS. (John Wiley & Sons, Ltd.
74. Grace, T.D. (1962). Establishment of four strains of cells from insect tissues grown in vitro. Nature 195, 788-789.
75. Vaughn, J.L., Goodwin, R.H., Tompkins, G.J., and McCawley, P. (1977). The establishment of two cell lines from the insect Spodoptera frugiperda (Lepidoptera; Noctuidae). In vitro 13, 213-217.
76. Wickham, T.J., Davis, T., Granados, R.R., Shuler, M.L., and Wood, H.A. (1992). Screening of insect cell lines for the production of recombinant proteins and infectious virus in the baculovirus expression system. Biotechnology progress 8, 391-396.
77. Beas-Catena, A., Sanchez-Miron, A., Garcia-Camacho, F., and Molina-Grima, E. (2011). Adaptation of the Se301 insect cell line to suspension culture. Effect of

turbulence on growth and on production of nucleopolyhedrovius (SeMNPV). Cytotechnology 63, 543-552.
78. Beas-Catena, A., Sanchez-Miron, A., Garcia-Camacho, F., Contreras-Gomez, A., and Molina-Grima, E. (2013). Adaptation of the Spodoptera exigua Se301 insect cell line to grow in serum-free suspended culture. Comparison of SeMNPV productivity in serum-free and serum-containing media. Applied microbiology and biotechnology 97, 3373-3381.
79. Elias, C.B., Jardin, B., and Kamen, A. (2007). Recombinant protein production in large-scale agitated bioreactors using the baculovirus expression vector system. Methods Mol Biol 388, 225-246.
80. Kadwell, S.H., and Hardwicke, P.I. (2007). Production of baculovirus-expressed recombinant proteins in wave bioreactors. Methods Mol Biol 388, 247-266.
81. Hollister, J.R., Shaper, J.H., and Jarvis, D.L. (1998). Stable expression of mammalian beta 1,4-galactosyltransferase extends the N-glycosylation pathway in insect cells. Glycobiology 8, 473-480.
82. Hollister, J., Grabenhorst, E., Nimtz, M., Conradt, H., and Jarvis, D.L. (2002). Engineering the protein N-glycosylation pathway in insect cells for production of biantennary, complex N-glycans. Biochemistry 41, 15093-15104.
83. Fath-Goodin, A., Kroemer, J., Martin, S., Reeves, K., and Webb, B.A. (2006). Polydnavirus genes that enhance the baculovirus expression vector system. Advances in virus research 68, 75-90.
84. Sokolenko, S., George, S., Wagner, A., Tuladhar, A., Andrich, J.M., and Aucoin, M.G. (2012). Co-expression vs. co-infection using baculovirus expression vectors in insect cell culture: Benefits and drawbacks. Biotechnology advances 30, 766-781.
85. Boer, E., Steinborn, G., Kunze, G., and Gellissen, G. (2007). Yeast expression platforms. Applied microbiology and biotechnology 77, 513-523.
86. Estes, S., and Melville, M. (2014). Mammalian cell line developments in speed and efficiency. Advances in biochemical engineering/biotechnology 139, 11-33.
87. Punt, P.J., van Biezen, N., Conesa, A., Albers, A., Mangnus, J., and van den Hondel, C. (2002). Filamentous fungi as cell factories for heterologous protein production. Trends in biotechnology 20, 200-206.
88. Cherry, J.R., and Fidantsef, A.L. (2003). Directed evolution of industrial enzymes: an update. Current opinion in biotechnology 14, 438-443.
89. Janssen, M., Tramper, J., Mur, L.R., and Wijffels, R.H. (2003). Enclosed outdoor photobioreactors: light regime, photosynthetic efficiency, scale-up, and future prospects. Biotechnology and bioengineering 81, 193-210.
90. Griesbeck, C., Kobl, I., and Heitzer, M. (2006). Chlamydomonas reinhardtii: a protein expression system for pharmaceutical and biotechnological proteins. Molecular biotechnology 34, 213-223.
91. Rosenberg, J.N., Oyler, G.A., Wilkinson, L., and Betenbaugh, M.J. (2008). A green light for engineered algae: redirecting metabolism to fuel a biotechnology revolution. Current opinion in biotechnology 19, 430-436.
92. Pauling, L., and Corey, R.B. (1951). Configurations of Polypeptide Chains With Favored Orientations Around Single Bonds: Two New Pleated Sheets. Proceedings of the National Academy of Sciences of the United States of America 37, 729-740.
93. Perutz, M.F., Rossmann, M.G., Cullis, A.F., Muirhead, H., Will, G., and North, A.C. (1960). Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. Nature 185, 416-422.
94. Anfinsen, C.B. (1973). Principles that govern the folding of protein chains. Science 181, 223-230.
95. Levinthal, C. (1968). Are there pathways for protein folding? J Chim Phys, 44-45.
96. Wetlaufer, D.B. (1973). Nucleation, rapid folding, and globular intrachain regions in proteins. Proceedings of the National Academy of Sciences of the United States of America 70, 697-701.

97. Dill, K.A. (1985). Theory for the folding and stability of globular proteins. Biochemistry 24, 1501-1509.
98. Bryngelson, J.D., and Wolynes, P.G. (1987). Spin glasses and the statistical mechanics of protein folding. Proceedings of the National Academy of Sciences of the United States of America 84, 7524-7528.
99. Leopold, P.E., Montal, M., and Onuchic, J.N. (1992). Protein folding funnels: a kinetic approach to the sequence-structure relationship. Proceedings of the National Academy of Sciences of the United States of America 89, 8721-8725.
100. Wolynes, P.G., Onuchic, J.N., and Thirumalai, D. (1995). Navigating the folding routes. Science 267, 1619-1620.
101. Harrison, S.C., and Durbin, R. (1985). Is there a single pathway for the folding of a polypeptide chain? Proceedings of the National Academy of Sciences of the United States of America 82, 4028-4030.
102. Dill, K.A., and MacCallum, J.L. (2012). The protein-folding problem, 50 years on. Science 338, 1042-1046.
103. Montelione, G.T., and Scheraga, H.A. (1989). Formation of local structures in protein folding. Accounts of Chemical Research 22, 70-76.
104. Dobson, C.M. (2004). Principles of protein folding, misfolding and aggregation. Seminars in cell & developmental biology 15, 3-16.
105. Fink, A.L. (1998). Protein aggregation: folding aggregates, inclusion bodies and amyloid. Folding & design 3, R9-23.
106. Bullock, A.N., and Fersht, A.R. (2001). Rescuing the function of mutant p53. Nature reviews Cancer 1, 68-76.
107. Gallati, S. (2014). Disease-modifying genes and monogenic disorders: experience in cystic fibrosis. The application of clinical genetics 7, 133-146.
108. Agorogiannis, E.I., Agorogiannis, G.I., Papadimitriou, A., and Hadjigeorgiou, G.M. (2004). Protein misfolding in neurodegenerative diseases. Neuropathology and applied neurobiology 30, 215-224.
109. Horwich, A. (2002). Protein aggregation in disease: a role for folding intermediates forming specific multimeric interactions. The Journal of clinical investigation 110, 1221-1232.
110. Kelly, J.W. (1998). The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. Current opinion in structural biology 8, 101-106.
111. Uversky, V.N., and Eliezer, D. (2009). Biophysics of Parkinson's disease: structure and aggregation of alpha-synuclein. Current protein & peptide science 10, 483-499.
112. Ellis, R.J., and Minton, A.P. (2006). Protein aggregation in crowded environments. Biological chemistry 387, 485-497.
113. Ellis, R.J., and van der Vies, S.M. (1991). Molecular chaperones. Annual review of biochemistry 60, 321-347.
114. Frydman, J., Nimmesgern, E., Ohtsuka, K., and Hartl, F.U. (1994). Folding of nascent polypeptide chains in a high molecular mass assembly with molecular chaperones. Nature 370, 111-117.
115. Hartl, F.U., and Hayer-Hartl, M. (2002). Molecular chaperones in the cytosol: from nascent chain to folded protein. Science 295, 1852-1858.
116. Bukau, B., Weissman, J., and Horwich, A. (2006). Molecular chaperones and protein quality control. Cell 125, 443-451.
117. Doyle, S.M., Genest, O., and Wickner, S. (2013). Protein rescue from aggregates by powerful molecular chaperone machines. Nature reviews Molecular cell biology 14, 617-629.
118. Laskey, R.A., Honda, B.M., Mills, A.D., and Finch, J.T. (1978). Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. Nature 275, 416-420.
119. Thomas, J.G., Ayling, A., and Baneyx, F. (1997). Molecular chaperones, folding catalysts, and the recovery of active recombinant proteins from E. coli. To fold or to refold. Applied biochemistry and biotechnology 66, 197-238.
120. Ellis, J. (1987). Proteins as molecular chaperones. Nature 328, 378-379.

121. Baneyx, F. (2004). Keeping up with protein folding. Microbial cell factories 3, 6.
122. Langer, T., Lu, C., Echols, H., Flanagan, J., Hayer, M.K., and Hartl, F.U. (1992). Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. Nature 356, 683-689.
123. Siegers, K., Waldmann, T., Leroux, M.R., Grein, K., Shevchenko, A., Schiebel, E., and Hartl, F.U. (1999). Compartmentation of protein folding in vivo: sequestration of non-native polypeptide by the chaperonin-GimC system. The EMBO journal 18, 75-84.
124. Hesterkamp, T., Hauser, S., Lutcke, H., and Bukau, B. (1996). Escherichia coli trigger factor is a prolyl isomerase that associates with nascent polypeptide chains. Proceedings of the National Academy of Sciences of the United States of America 93, 4437-4441.
125. Houry, W.A. (2001). Chaperone-assisted protein folding in the cell cytoplasm. Current protein & peptide science 2, 227-244.
126. Patzelt, H., Rudiger, S., Brehmer, D., Kramer, G., Vorderwulbecke, S., Schaffitzel, E., Waitz, A., Hesterkamp, T., Dong, L., Schneider-Mergener, J., et al. (2001). Binding specificity of Escherichia coli trigger factor. Proceedings of the National Academy of Sciences of the United States of America 98, 14244-14249.
127. Kandror, O., and Goldberg, A.L. (1997). Trigger factor is induced upon cold shock and enhances viability of Escherichia coli at low temperatures. Proceedings of the National Academy of Sciences of the United States of America 94, 4978-4981.
128. Bukau, B., and Horwich, A.L. (1998). The Hsp70 and Hsp60 chaperone machines. Cell 92, 351-366.
129. Gao, X.C., Zhou, C.J., Zhou, Z.R., Wu, M., Cao, C.Y., and Hu, H.Y. (2012). The C-terminal helices of heat shock protein 70 are essential for J-domain binding and ATPase activation. The Journal of biological chemistry 287, 6044-6052.
130. Rudiger, S., Schneider-Mergener, J., and Bukau, B. (2001). Its substrate specificity characterizes the DnaJ co-chaperone as a scanning factor for the DnaK chaperone. The EMBO journal 20, 1042-1050.
131. Sha, B., Lee, S., and Cyr, D.M. (2000). The crystal structure of the peptide-binding fragment from the yeast Hsp40 protein Sis1. Structure 8, 799-807.
132. Harrison, C.J., Hayer-Hartl, M., Di Liberto, M., Hartl, F., and Kuriyan, J. (1997). Crystal structure of the nucleotide exchange factor GrpE bound to the ATPase domain of the molecular chaperone DnaK. Science 276, 431-435.
133. Teter, S.A., Houry, W.A., Ang, D., Tradler, T., Rockabrand, D., Fischer, G., Blum, P., Georgopoulos, C., and Hartl, F.U. (1999). Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains. Cell 97, 755-765.
134. Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A., and Bukau, B. (1999). Trigger factor and DnaK cooperate in folding of newly synthesized proteins. Nature 400, 693-696.
135. PDB, P.J. In. ((http://pdbj.org/eprots/index_en.cgi?PDB%3A3a6m)
136. van Bergen en Henegouwen, P.M., Berbers, G., Linnemans, W.A., and van Wijk, R. (1987). Subcellular localization of the 84,000 dalton heat-shock protein in mouse neuroblastoma cells: evidence for a cytoplasmic and nuclear location. European journal of cell biology 43, 469-478.
137. Katschinski, D.M., Le, L., Heinrich, D., Wagner, K.F., Hofer, T., Schindler, S.G., and Wenger, R.H. (2002). Heat induction of the unphosphorylated form of hypoxia-inducible factor-1alpha is dependent on heat shock protein-90 activity. The Journal of biological chemistry 277, 9262-9267.
138. Tapia, H., and Morano, K.A. (2010). Hsp90 nuclear accumulation in quiescence is linked to chaperone function and spore development in yeast. Molecular biology of the cell 21, 63-72.
139. Mollapour, M., and Neckers, L. (2012). Post-translational modifications of Hsp90 and their contributions to chaperone regulation. Biochimica et biophysica acta 1823, 648-655.

140. Erlejman, A.G., Lagadari, M., Toneatto, J., Piwien-Pilipuk, G., and Galigniana, M.D. (2014). Regulatory role of the 90-kDa-heat-shock protein (Hsp90) and associated factors on gene expression. Biochimica et biophysica acta 1839, 71-87.
141. Eckl, J.M., and Richter, K. (2013). Functions of the Hsp90 chaperone system: lifting client proteins to new heights. International journal of biochemistry and molecular biology 4, 157-165.
142. Baneyx, F., and Mujacic, M. (2004). Recombinant protein folding and misfolding in Escherichia coli. Nature biotechnology 22, 1399-1408.
143. Sherman, M.Y., and Goldberg, A.L. (1996). Involvement of molecular chaperones in intracellular protein breakdown. Exs 77, 57-78.
144. Chandu, D., and Nandi, D. (2004). Comparative genomics and functional roles of the ATP-dependent proteases Lon and Clp during cytosolic protein degradation. Research in microbiology 155, 710-719.
145. Botos, I., Melnikov, E.E., Cherry, S., Tropea, J.E., Khalatova, A.G., Rasulova, F., Dauter, Z., Maurizi, M.R., Rotanova, T.V., Wlodawer, A., et al. (2004). The catalytic domain of Escherichia coli Lon protease has a unique fold and a Ser-Lys dyad in the active site. The Journal of biological chemistry 279, 8140-8148.
146. Fischer, H., and Glockshuber, R. (1994). A point mutation within the ATP-binding site inactivates both catalytic functions of the ATP-dependent protease La (Lon) from Escherichia coli. FEBS letters 356, 101-103.
147. Ebel, W., Skinner, M.M., Dierksen, K.P., Scott, J.M., and Trempy, J.E. (1999). A conserved domain in Escherichia coli Lon protease is involved in substrate discriminator activity. Journal of bacteriology 181, 2236-2243.
148. Swamy, K.H., and Goldberg, A.L. (1981). E. coli contains eight soluble proteolytic activities, one being ATP dependent. Nature 292, 652-654.
149. Tsilibaris, V., Maenhaut-Michel, G., and Van Melderen, L. (2006). Biological roles of the Lon ATP-dependent protease. Research in microbiology 157, 701-713.
150. Yu, A.Y., and Houry, W.A. (2007). ClpP: a distinctive family of cylindrical energy-dependent serine proteases. FEBS letters 581, 3749-3757.
151. Kress, W., Maglica, Z., and Weber-Ban, E. (2009). Clp chaperone-proteases: structure and function. Research in microbiology 160, 618-628.
152. Mahato, R.I., Takakura, Y., and Hashida, M. (1997). Development of targeted delivery systems for nucleic acid drugs. Journal of drug targeting 4, 337-357.
153. Jacobs, P. (1980). Pioneer genetic implants revealed. Los Ageles Times.
154. Rosenberg, S.A., Aebersold, P., Cornetta, K., Kasid, A., Morgan, R.A., Moen, R., Karson, E.M., Lotze, M.T., Yang, J.C., Topalian, S.L., et al. (1990). Gene transfer into humans--immunotherapy of patients with advanced melanoma, using tumor-infiltrating lymphocytes modified by retroviral gene transduction. The New England journal of medicine 323, 570-578.
155. Blaese, R.M., Culver, K.W., Chang, L., Anderson, W.F., Mullen, C., Nienhuis, A., Carter, C., Dunbar, C., Leitman, S., Berger, M., et al. (1993). Treatment of severe combined immunodeficiency disease (SCID) due to adenosine deaminase deficiency with CD34+ selected autologous peripheral blood cells transduced with a human ADA gene. Amendment to clinical research project, Project 90-C-195, January 10, 1992. Human gene therapy 4, 521-527.
156. Kay, M.A., Manno, C.S., Ragni, M.V., Larson, P.J., Couto, L.B., McClelland, A., Glader, B., Chew, A.J., Tai, S.J., Herzog, R.W., et al. (2000). Evidence for gene transfer and expression of factor IX in haemophilia B patients treated with an AAV vector. Nature genetics 24, 257-261.
157. Cavazzana-Calvo, M., Hacein-Bey, S., de Saint Basile, G., Gross, F., Yvon, E., Nusbaum, P., Selz, F., Hue, C., Certain, S., Casanova, J.L., et al. (2000). Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. Science 288, 669-672.
158. Khuri, F.R., Nemunaitis, J., Ganly, I., Arseneau, J., Tannock, I.F., Romel, L., Gore, M., Ironside, J., MacDougall, R.H., Heise, C., et al. (2000). a controlled trial of intratumoral ONYX-015, a selectively-replicating adenovirus, in combination with

cisplatin and 5-fluorouracil in patients with recurrent head and neck cancer. Nature medicine 6, 879-885.
159. Hacein-Bey-Abina, S., von Kalle, C., Schmidt, M., Le Deist, F., Wulffraat, N., McIntyre, E., Radford, I., Villeval, J.L., Fraser, C.C., Cavazzana-Calvo, M., et al. (2003). A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. The New England journal of medicine 348, 255-256.
160. Pearson, S., Jia, H., and Kandachi, K. (2004). China approves first gene therapy. Nature biotechnology 22, 3-4.
161. Fischer, A., Hacein-Bey-Abina, S., and Cavazzana-Calvo, M. (2010). 20 years of gene therapy for SCID. Nature immunology 11, 457-460.
162. Ferrua, F., Brigida, I., and Aiuti, A. (2010). Update on gene therapy for adenosine deaminase-deficient severe combined immunodeficiency. Current opinion in allergy and clinical immunology 10, 551-556.
163. LeWitt, P.A., Rezai, A.R., Leehey, M.A., Ojemann, S.G., Flaherty, A.W., Eskandar, E.N., Kostyk, S.K., Thomas, K., Sarkar, A., Siddiqui, M.S., et al. (2011). AAV2-GAD gene therapy for advanced Parkinson's disease: a double-blind, sham-surgery controlled, randomised trial. The Lancet Neurology 10, 309-319.
164. Cartier, N., and Aubourg, P. (2010). Hematopoietic stem cell transplantation and hematopoietic stem cell gene therapy in X-linked adrenoleukodystrophy. Brain Pathol 20, 857-862.
165. Ledford, H. (2011). Cell therapy fights leukaemia. Nature.
166. Richards, S. (2012). Gene therapy arrives in europe. TheScientist.
167. uniQure. In. ((http://www.uniqure.com/products/glybera/)
168. Wiley. In. ((http://www.wiley.com/legacy/wileychi/genmed/clinical/)
169. Unifg. In. ((http://www.medicina.unifg.it)
170. Frank, K.M., Hogarth, D.K., Miller, J.L., Mandal, S., Mease, P.J., Samulski, R.J., Weisgerber, G.A., and Hart, J. (2009). Investigation of the Cause of Death in a Gene-Therapy Trial. New England Journal of Medicine 361, 161-169.
171. Somia, N., and Verma, I.M. (2000). Gene therapy: trials and tribulations. Nature reviews Genetics 1, 91-99.
172. Senior, J.H. (1987). Fate and behavior of liposomes in vivo: a review of controlling factors. Critical reviews in therapeutic drug carrier systems 3, 123-193.
173. Takakura, Y., and Hashida, M. (1996). Macromolecular carrier systems for targeted drug delivery: pharmacokinetic considerations on biodistribution. Pharmaceutical research 13, 820-831.
174. Lin, M.T., Pulkkinen, L., Uitto, J., and Yoon, K. (2000). The gene gun: current applications in cutaneous gene therapy. International journal of dermatology 39, 161-170.
175. Nishikawa, M., and Hashida, M. (2002). Nonviral approaches satisfying various requirements for effective in vivo gene therapy. Biological & pharmaceutical bulletin 25, 275-283.
176. Davis, M.E. (2002). Non-viral gene delivery systems. Current opinion in biotechnology 13, 128-131.
177. Greelish, J.P., Su, L.T., Lankford, E.B., Burkman, J.M., Chen, H., Konig, S.K., Mercier, I.M., Desjardins, P.R., Mitchell, M.A., Zheng, X.G., et al. (1999). Stable restoration of the sarcoglycan complex in dystrophic muscle perfused with histamine and a recombinant adeno-associated viral vector. Nature medicine 5, 439-443.
178. Ziello, J.E., Huang, Y., and Jovin, I.S. (2010). Cellular endocytosis and gene delivery. Mol Med 16, 222-229.
179. Sonawane, N.D., Szoka, F.C., Jr., and Verkman, A.S. (2003). Chloride accumulation and swelling in endosomes enhances DNA transfer by polyamine-DNA polyplexes. The Journal of biological chemistry 278, 44826-44831.
180. Midoux, P., Pichon, C., Yaouanc, J.J., and Jaffres, P.A. (2009). Chemical vectors for gene delivery: a current review on polymers, peptides and lipids containing histidine or imidazole as nucleic acids carriers. British journal of pharmacology 157, 166-178.

181. Nishimura, Y., Takeda, K., Ezawa, R., Ishii, J., Ogino, C., and Kondo, A. (2014). A display of pH-sensitive fusogenic GALA peptide facilitates endosomal escape from a Bio-nanocapsule via an endocytic uptake pathway. Journal of nanobiotechnology 12, 11.
182. Adam, E.J., and Adam, S.A. (1994). Identification of cytosolic factors required for nuclear location sequence-mediated binding to the nuclear envelope. The Journal of cell biology 125, 547-555.
183. Jimenez-Sanchez, G., Childs, B., and Valle, D. (2001). Human disease genes. Nature 409, 853-855.
184. Weiss, B., Davidkova, G., and Zhou, L.W. (1999). Antisense RNA gene therapy for studying and modulating biological processes. Cellular and molecular life sciences : CMLS 55, 334-358.
185. Dorsett, Y., and Tuschl, T. (2004). siRNAs: applications in functional genomics and potential as therapeutics. Nature reviews Drug discovery 3, 318-329.
186. Lambeth, L.S., and Smith, C.A. (2013). Short hairpin RNA-mediated gene silencing. Methods Mol Biol 942, 205-232.
187. Benitec. In. ((http://www.benitec.com/rna-interference-rnai.php)
188. Nilsen, T.W., and Graveley, B.R. (2010). Expansion of the eukaryotic proteome by alternative splicing. Nature 463, 457-463.
189. Bauman, J., Jearawiriyapaisarn, N., and Kole, R. (2009). Therapeutic potential of splice-switching oligonucleotides. Oligonucleotides 19, 1-13.
190. Times, T.N.Y. (1999). The Biotech Death of Jesse Gelsinger. In. (http://www.nytimes.com/1999/11/28/magazine/the-biotech-death-of-jesse-gelsinger.html)
191. Domingo-Espín, J. (2013). Development and characterization of artificial viruses for gene therapy. PhD Thesis, Universitat Autónoma de Barcelona.
192. Kay, M.A., Glorioso, J.C., and Naldini, L. (2001). Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics. Nature medicine 7, 33-40.
193. Ibraheem, D., Elaissari, A., and Fessi, H. (2014). Gene therapy and DNA delivery systems. International journal of pharmaceutics 459, 70-83.
194. Dani, S.U. (1999). The challenge of vector development in gene therapy. Brazilian journal of medical and biological research = Revista brasileira de pesquisas medicas e biologicas / Sociedade Brasileira de Biofisica [et al] 32, 133-145.
195. Cavazzana-Calvo, M., Payen, E., Negre, O., Wang, G., Hehir, K., Fusil, F., Down, J., Denaro, M., Brady, T., Westerman, K., et al. (2010). Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia. Nature 467, 318-322.
196. Bainbridge, J.W., Smith, A.J., Barker, S.S., Robbie, S., Henderson, R., Balaggan, K., Viswanathan, A., Holder, G.E., Stockman, A., Tyler, N., et al. (2008). Effect of gene therapy on visual function in Leber's congenital amaurosis. The New England journal of medicine 358, 2231-2239.
197. Wang, W., Li, W., Ma, N., and Steinhoff, G. (2013). Non-viral gene delivery methods. Current pharmaceutical biotechnology 14, 46-60.
198. Gao, X., Kim, K.S., and Liu, D. (2007). Nonviral gene delivery: what we know and what is next. The AAPS journal 9, E92-104.
199. Newman, C.M., and Bettinger, T. (2007). Gene therapy progress and prospects: ultrasound for gene transfer. Gene therapy 14, 465-475.
200. Romero, N.B., Braun, S., Benveniste, O., Leturcq, F., Hogrel, J.Y., Morris, G.E., Barois, A., Eymard, B., Payan, C., Ortega, V., et al. (2004). Phase I study of dystrophin plasmid-based gene therapy in Duchenne/Becker muscular dystrophy. Human gene therapy 15, 1065-1076.
201. Krzysztof Gawrychowski, M.M., Grzegorz Szewczyk, Grzegorz Panek, Wacław Śmiertka, Katarzyna Roszkowska-Purska, Mariusz Bidziński, Małgorzata Sadowska, Joanna Surowińska, Monika Dąbrowska, Anna Górniak, Ryszard Krynicki, Karolina Hajdukiewicz. (2010). Gene therapy of locally advanced vulvar cancer with psFLT construct. Contemporary oncology 14, 4.

202. Kaneda, Y. (2001). Gene therapy: a battle against biological barriers. Current molecular medicine 1, 493-499.
203. Aied, A., Greiser, U., Pandit, A., and Wang, W. (2013). Polymer gene delivery: overcoming the obstacles. Drug discovery today 18, 1090-1098.
204. Parhiz, H., Shier, W.T., and Ramezani, M. (2013). From rationally designed polymeric and peptidic systems to sophisticated gene delivery nano-vectors. International journal of pharmaceutics 457, 237-259.
205. Yin, H., Kanasty, R.L., Eltoukhy, A.A., Vegas, A.J., Dorkin, J.R., and Anderson, D.G. (2014). Non-viral vectors for gene-based therapy. Nature reviews Genetics 15, 541-555.
206. Stopeck, A.T., Jones, A., Hersh, E.M., Thompson, J.A., Finucane, D.M., Gutheil, J.C., and Gonzalez, R. (2001). Phase II study of direct intralesional gene transfer of allovectin-7, an HLA-B7/beta2-microglobulin DNA-liposome complex, in patients with metastatic melanoma. Clinical cancer research : an official journal of the American Association for Cancer Research 7, 2285-2291.
207. Ferrer-Miralles, N., Vazquez, E., and Villaverde, A. (2008). Membrane-active peptides for non-viral gene therapy: making the safest easier. Trends in biotechnology 26, 267-275.
208. Aris, A., and Villaverde, A. (2000). Molecular organization of protein-DNA complexes for cell-targeted DNA delivery. Biochemical and biophysical research communications 278, 455-461.
209. Peluffo, H., Aris, A., Acarin, L., Gonzalez, B., Villaverde, A., and Castellano, B. (2003). Nonviral gene delivery to the central nervous system based on a novel integrin-targeting multifunctional protein. Human gene therapy 14, 1215-1223.
210. Dahiyat, B.I., and Mayo, S.L. (1997). De novo protein design: fully automated sequence selection. Science 278, 82-87.
211. Khoury, G.A., Smadbeck, J., Kieslich, C.A., and Floudas, C.A. (2014). Protein folding and de novo protein design for biotechnological applications. Trends in biotechnology 32, 99-109.
212. Glover, D.J., Ng, S.M., Mechler, A., Martin, L.L., and Jans, D.A. (2009). Multifunctional protein nanocarriers for targeted nuclear gene delivery in nondividing cells. FASEB journal : official publication of the Federation of American Societies for Experimental Biology 23, 2996-3006.
213. D'Souza, S.E., Ginsberg, M.H., and Plow, E.F. (1991). Arginyl-glycyl-aspartic acid (RGD): a cell adhesion motif. Trends in biochemical sciences 16, 246-250.
214. Domingo-Espin, J., Petegnief, V., de Vera, N., Conchillo-Sole, O., Saccardo, P., Unzueta, U., Vazquez, E., Cedano, J., Negro, L., Daura, X., et al. (2012). RGD-based cell ligands for cell-targeted drug delivery act as potent trophic factors. Nanomedicine : nanotechnology, biology, and medicine 8, 1263-1266.
215. Demeule, M., Regina, A., Che, C., Poirier, J., Nguyen, T., Gabathuler, R., Castaigne, J.P., and Beliveau, R. (2008). Identification and design of peptides as a new drug delivery system for the brain. The Journal of pharmacology and experimental therapeutics 324, 1064-1072.
216. Chakrabarti, A., Witsenburg, J.J., Sinzinger, M.D., Richter, M., Wallbrecher, R., Cluitmans, J.C., Verdurmen, W.P., Tanis, S., Adjobo-Hermans, M.J., Rademann, J., et al. (2014). Multivalent presentation of the cell-penetrating peptide nona-arginine on a linear scaffold strongly increases its membrane-perturbing capacity. Biochimica et biophysica acta 1838, 3097-3106.
217. Unzueta, U., Cespedes, M.V., Ferrer-Miralles, N., Casanova, I., Cedano, J., Corchero, J.L., Domingo-Espin, J., Villaverde, A., Mangues, R., and Vazquez, E. (2012). Intracellular CXCR4(+) cell targeting with T22-empowered protein-only nanoparticles. International journal of nanomedicine 7, 4533-4544.
218. Ferrer-Miralles, N., Corchero, J.L., Kumar, P., Cedano, J.A., Gupta, K.C., Villaverde, A., and Vazquez, E. (2011). Biological activities of histidine-rich peptides; merging biotechnology and nanomedicine. Microbial cell factories 10, 101.

219. Aris, A., and Villaverde, A. (2003). Engineering nuclear localization signals in modular protein vehicles for gene therapy. Biochemical and biophysical research communications 304, 625-631.
220. Ramqvist, T., Andreasson, K., and Dalianis, T. (2007). Vaccination, immune and gene therapy based on virus-like particles against viral infections and cancer. Expert opinion on biological therapy 7, 997-1007.
221. Slilaty, S.N., and Aposhian, H.V. (1983). Gene transfer by polyoma-like particles assembled in a cell-free system. Science 220, 725-727.
222. Forstova, J., Krauzewicz, N., Sandig, V., Elliott, J., Palkova, Z., Strauss, M., and Griffin, B.E. (1995). Polyoma virus pseudocapsids as efficient carriers of heterologous DNA into mammalian cells. Human gene therapy 6, 297-306.
223. Saccardo, P., Rodriguez-Carmona, E., Villaverde, A., and Ferrer-Miralles, N. (2014). Effect of the DnaK chaperone on the conformational quality of JCV VP1 virus-like particles produced in Escherichia coli. Biotechnology progress 30, 744-748.
224. Bundy, B.C., Franciszkowicz, M.J., and Swartz, J.R. (2008). Escherichia coli-based cell-free synthesis of virus-like particles. Biotechnology and bioengineering 100, 28-37.
225. Huang, Z., Elkin, G., Maloney, B.J., Beuhner, N., Arntzen, C.J., Thanavala, Y., and Mason, H.S. (2005). Virus-like particle expression and assembly in plants: hepatitis B and Norwalk viruses. Vaccine 23, 1851-1858.
226. Pan, X., Lu, P., Zhang, M., Hu, Z., Li, G., Ma, S., Feng, F., Chen, K., and Yao, Q. (2014). Expression analysis of Bombyx mori bidensovirus structural proteins and assembly of virus-like particles in insect cells. Current microbiology 69, 567-573.
227. Chang, D., Fung, C.Y., Ou, W.C., Chao, P.C., Li, S.Y., Wang, M., Huang, Y.L., Tzeng, T.Y., and Tsai, R.T. (1997). Self-assembly of the JC virus major capsid protein, VP1, expressed in insect cells. The Journal of general virology 78 (Pt 6), 1435-1439.
228. Tome-Amat, J., Fleischer, L., Parker, S.A., Bardliving, C., and Batt, C. (2014). Secreted production of assembled Norovirus virus-like particles from Pichia pastoris. Microbial cell factories 13, 134.
229. Hale, A.D., Bartkeviciute, D., Dargeviciute, A., Jin, L., Knowles, W., Staniulis, J., Brown, D.W., and Sasnauskas, K. (2002). Expression and antigenic characterization of the major capsid proteins of human polyomaviruses BK and JC in Saccharomyces cerevisiae. Journal of virological methods 104, 93-98.
230. Oka, T., Hansman, G.S., Katayama, K., Ogawa, S., Nagata, N., Miyamura, T., and Takeda, N. (2006). Expression of sapovirus virus-like particles in mammalian cells. Archives of virology 151, 399-404.
231. Fontana, D., Kratje, R., Etcheverrigaray, M., and Prieto, C. (2014). Rabies virus-like particles expressed in HEK293 cells. Vaccine 32, 2799-2804.
232. Teunissen, E.A., de Raad, M., and Mastrobattista, E. (2013). Production and biomedical applications of virus-like particles derived from polyomaviruses. Journal of controlled release : official journal of the Controlled Release Society 172, 305-321.
233. Hahne, T., Palaniyandi, M., Kato, T., Fleischmann, P., Watzig, H., and Park, E.Y. (2014). Characterization of human papillomavirus 6b L1 virus-like particles isolated from silkworms using capillary zone electrophoresis. Journal of bioscience and bioengineering 118, 311-314.
234. Niikura, M., Takamura, S., Kim, G., Kawai, S., Saijo, M., Morikawa, S., Kurane, I., Li, T.C., Takeda, N., and Yasutomi, Y. (2002). Chimeric recombinant hepatitis E virus-like particles as an oral vaccine vehicle presenting foreign epitopes. Virology 293, 273-280.
235. Brandenburg, B., Stockl, L., Gutzeit, C., Roos, M., Lupberger, J., Schwartlander, R., Gelderblom, H., Sauer, I.M., Hofschneider, P.H., and Hildt, E. (2005). A novel system for efficient gene transfer into primary human hepatocytes via cell-permeable hepatitis B virus-like particle. Hepatology 42, 1300-1309.
236. Kessans, S.A., Linhart, M.D., Matoba, N., and Mor, T. (2013). Biological and biochemical characterization of HIV-1 Gag/dgp41 virus-like particles expressed in Nicotiana benthamiana. Plant biotechnology journal 11, 681-690.

237. Rodriguez-Limas, W.A., Tyo, K.E., Nielsen, J., Ramirez, O.T., and Palomares, L.A. (2011). Molecular and process design for rotavirus-like particle production in Saccharomyces cerevisiae. Microbial cell factories 10, 33.
238. Sanchez-Rodriguez, S.P., Munch-Anguiano, L., Echeverria, O., Vazquez-Nin, G., Mora-Pale, M., Dordick, J.S., and Bustos-Jaimes, I. (2012). Human parvovirus B19 virus-like particles: In vitro assembly and stability. Biochimie 94, 870-878.
239. Gleiter, S., and Lilie, H. (2001). Coupling of antibodies via protein Z on modified polyoma virus-like particles. Protein science : a publication of the Protein Society 10, 434-444.
240. Bioinformatics, S.I.o. (2010). In. (viralzone.expasy.org)
241. Viola, J.R., El-Andaloussi, S., Oprea, II, and Smith, C.I. (2010). Non-viral nanovectors for gene delivery: factors that govern successful therapeutics. Expert opinion on drug delivery 7, 721-735.
242. Desai, M.P., Labhasetwar, V., Walter, E., Levy, R.J., and Amidon, G.L. (1997). The mechanism of uptake of biodegradable microparticles in Caco-2 cells is size dependent. Pharmaceutical research 14, 1568-1573.
243. Sciencedirect. In. (http://www.sciencedirect.com/science/article/pii/S1674200108000357)
244. VJ Mohanraj, Y.C. (2006). Nanoparticles-A review. Trop J Pharm Res 5, 12.
245. Mohsen Jahanshahi, Z.B. (2008). Protein nanoparticle: A unique system as drug delivery vehicles. African Journal of Biotechnology 7, 8.
246. Domingo-Espin, J., Unzueta, U., Saccardo, P., Rodriguez-Carmona, E., Corchero, J.L., Vazquez, E., and Ferrer-Miralles, N. (2011). Engineered biological entities for drug delivery and gene therapy protein nanoparticles. Progress in molecular biology and translational science 104, 247-298.
247. Goldmann, C., Stolte, N., Nisslein, T., Hunsmann, G., Luke, W., and Petry, H. (2000). Packaging of small molecules into VP1-virus-like particles of the human polyomavirus JC virus. Journal of virological methods 90, 85-90.
248. Vazquez, E., Roldan, M., Diez-Gil, C., Unzueta, U., Domingo-Espin, J., Cedano, J., Conchillo, O., Ratera, I., Veciana, J., Daura, X., et al. (2010). Protein nanodisk assembling and intracellular trafficking powered by an arginine-rich (R9) peptide. Nanomedicine (Lond) 5, 259-268.
249. Middaugh, C.R., Evans, R.K., Montgomery, D.L., and Casimiro, D.R. (1998). Analysis of plasmid DNA from a pharmaceutical perspective. Journal of pharmaceutical sciences 87, 130-146.
250. Mahato, R.I. (1999). Non-viral peptide-based approaches to gene delivery. Journal of drug targeting 7, 249-268.
251. Eickbush, T.H., and Moudrianakis, E.N. (1978). The compaction of DNA helices into either continuous supercoils or folded-fiber rods and toroids. Cell 13, 295-306.
252. Balls, M. (2009). The origins and early days of the Three Rs concept. Alternatives to laboratory animals : ATLA 37, 255-265.
253. Manno, C.S., Pierce, G.F., Arruda, V.R., Glader, B., Ragni, M., Rasko, J.J., Ozelo, M.C., Hoots, K., Blatt, P., Konkle, B., et al. (2006). Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. Nature medicine 12, 342-347.
254. Sibbald, B. (2001). Death but one unintended consequence of gene-therapy trial. CMAJ : Canadian Medical Association journal, journal de l'Association medicale canadienne 164, 1612.
255. Baldo, A., van den Akker, E., Bergmans, H.E., Lim, F., and Pauwels, K. (2013). General considerations on the biosafety of virus-derived vectors used in gene therapy and vaccination. Current gene therapy 13, 385-394.
256. Chen, L.S., Wang, M., Ou, W.C., Fung, C.Y., Chen, P.L., Chang, C.F., Huang, W.S., Wang, J.Y., Lin, P.Y., and Chang, D. (2010). Efficient gene transfer using the human JC virus-like particle that inhibits human colon adenocarcinoma growth in a nude mouse model. Gene therapy 17, 1033-1041.
257. Bozaykut, P., Ozer, N.K., and Karademir, B. (2014). Regulation of protein turnover by heat shock proteins. Free radical biology & medicine.

258. Ou, W.C., Chen, L.H., Wang, M., Hseu, T.H., and Chang, D. (2001). Analysis of minimal sequences on JC virus VP1 required for capsid assembly. Journal of neurovirology 7, 298-301.
259. Goldmann, C., Petry, H., Frye, S., Ast, O., Ebitsch, S., Jentsch, K.D., Kaup, F.J., Weber, F., Trebst, C., Nisslein, T., et al. (1999). Molecular cloning and expression of major structural protein VP1 of the human polyomavirus JC virus: formation of virus-like particles useful for immunological and therapeutic studies. Journal of virology 73, 4465-4469.
260. Lima, M.A. (2013). Progressive multifocal leukoencephalopathy: new concepts. Arquivos de neuro-psiquiatria 71, 699-702.
261. Chou, M.I., Hsieh, Y.F., Wang, M., Chang, J.T., Chang, D., Zouali, M., and Tsay, G.J. (2010). In vitro and in vivo targeted delivery of IL-10 interfering RNA by JC virus-like particles. Journal of biomedical science 17, 51.
262. Chuan, Y.P., Fan, Y.Y., Lua, L.H., and Middelberg, A.P. (2010). Virus assembly occurs following a pH- or Ca2+-triggered switch in the thermodynamic attraction between structural protein capsomeres. Journal of the Royal Society, Interface / the Royal Society 7, 409-421.
263. Mukherjee, S., Abd-El-Latif, M., Bronstein, M., Ben-nun-Shaul, O., Kler, S., and Oppenheim, A. (2007). High cooperativity of the SV40 major capsid protein VP1 in virus assembly. PloS one 2, e765.
264. Kedzierska, S. (2005). [Role of Escherichia coli molecular chaperones in the protection of bacterial cells against irreversible aggregation induced by heat shock]. Postepy biochemii 51, 146-153.
265. Martinez-Alonso, M., Toledo-Rubio, V., Noad, R., Unzueta, U., Ferrer-Miralles, N., Roy, P., and Villaverde, A. (2009). Rehosting of bacterial chaperones for high-quality protein production. Applied and environmental microbiology 75, 7850-7854.
266. Saibil, H. (2013). Chaperone machines for protein folding, unfolding and disaggregation. Nature reviews Molecular cell biology 14, 630-642.
267. Kolaj, O., Spada, S., Robin, S., and Wall, J.G. (2009). Use of folding modulators to improve heterologous protein production in Escherichia coli. Microbial cell factories 8, 9.
268. Roy, P., and Noad, R. (2012). Use of bacterial artificial chromosomes in baculovirus research and recombinant protein expression: current trends and future perspectives. ISRN microbiology 2012, 628797.
269. Doll, T.A., Raman, S., Dey, R., and Burkhard, P. (2013). Nanoscale assemblies and their biomedical applications. Journal of the Royal Society, Interface / the Royal Society 10, 20120740.
270. Domingo-Espin, J., Vazquez, E., Ganz, J., Conchillo, O., Garcia-Fruitos, E., Cedano, J., Unzueta, U., Petegnief, V., Gonzalez-Montalban, N., Planas, A.M., et al. (2011). Nanoparticulate architecture of protein-based artificial viruses is supported by protein-DNA interactions. Nanomedicine (Lond) 6, 1047-1061.
271. Unzueta, U., Ferrer-Miralles, N., Cedano, J., Zikung, X., Pesarrodona, M., Saccardo, P., Garcia-Fruitos, E., Domingo-Espin, J., Kumar, P., Gupta, K.C., et al. (2012). Non-amyloidogenic peptide tags for the regulatable self-assembling of protein-only nanoparticles. Biomaterials 33, 8714-8722.
272. Cespedes, M.V., Unzueta, U., Tatkiewicz, W., Sanchez-Chardi, A., Conchillo-Sole, O., Alamo, P., Xu, Z., Casanova, I., Corchero, J.L., Pesarrodona, M., et al. (2014). In vivo architectonic stability of fully de novo designed protein-only nanoparticles. ACS nano 8, 4166-4176.
273. Cohen, P. (2000). The regulation of protein function by multisite phosphorylation--a 25 year update. Trends in biochemical sciences 25, 596-601.
274. Musselman, C.A., Khorasanizadeh, S., and Kutateladze, T.G. (2014). Towards understanding methyllysine readout. Biochimica et biophysica acta 1839, 686-693.
275. Glozak, M.A., Sengupta, N., Zhang, X., and Seto, E. (2005). Acetylation and deacetylation of non-histone proteins. Gene 363, 15-23.

276. Yokoyama, A., Fujiki, R., Ohtake, F., and Kato, S. (2011). Regulated histone methyltransferase and demethylase complexes in the control of genes by nuclear receptors. Cold Spring Harbor symposia on quantitative biology 76, 165-173.
277. Rudd, P.M., and Dwek, R.A. (1997). Glycosylation: heterogeneity and the 3D structure of proteins. Critical reviews in biochemistry and molecular biology 32, 1-100.
278. Crowell, D.N., and Huizinga, D.H. (2009). Protein isoprenylation: the fat of the matter. Trends in plant science 14, 163-170.
279. Nadolski, M.J., and Linder, M.E. (2007). Protein lipidation. The FEBS journal 274, 5202-5210.
280. Resh, M.D. (2012). Targeting protein lipidation in disease. Trends in molecular medicine 18, 206-214.
281. Casey, P.J. (1995). Protein lipidation in cell signaling. Science 268, 221-225.
282. Strieter, E.R., and Korasick, D.A. (2012). Unraveling the complexity of ubiquitin signaling. ACS chemical biology 7, 52-63.
283. Hay, R.T. (2005). SUMO: a history of modification. Molecular cell 18, 1-12.
284. Hannoun, Z., Greenhough, S., Jaffray, E., Hay, R.T., and Hay, D.C. (2010). Post-translational modification by SUMO. Toxicology 278, 288-293.
285. Rogers, L.D., and Overall, C.M. (2013). Proteolytic post-translational modification of proteins: proteomic tools and methodology. Molecular & cellular proteomics : MCP 12, 3532-3542.

Departament de Genètica i Microbiologia
Facultat de biociències

Paolo Saccardo

PhD Thesis 2015