ARTIFICIAL VIRAL CAPSID PROTEINS MIMICKING NATURE TO DELIVER MEDICINES

LIONE WILLEMS

Propositions

1. Design of virus-inspired transport vehicles requires the precise tuning of individual protein blocks.
   (this thesis)
2. Flexible coronas covering nanoparticles are not ideal for ensuring optimal activity of conjugated functional groups.
   (this thesis)
3. Scientifically accepted terms to characterise food products, such as “diet” and “light”, are misleading for the public.
4. Using medicines influences the evolution of mankind.
5. Pharmaceutical industries should care more about patients than patents.
6. The motorcyclist is the better car driver.
7. Lighting a fireplace or messing around with forklift trucks is an ineffective way to promote shorttrack speedskating.

Propositions belonging to the thesis entitled:

Artificial Viral Capsid Proteins

MIMICKING NATURE TO DELIVER MEDICINES

Lione Willems

Wageningen, 24 May 2019

Artificial Viral Capsid Proteins

MIMICKING NATURE TO DELIVER MEDICINES

Lione Willems

Thesis committee

Promotors
Dr Renko J. de Vries
Associate professor at the Physical Chemistry and Soft Matter Group
Wageningen University & Research

Prof. Dr John van der Oost
Personal chair at the Laboratory of Microbiology
Wageningen University & Research

Co-promotor
Prof. Dr Enrico Mastrobattista
Professor of Pharmaceutical Biotechnology and Delivery
Utrecht University

Other members
Prof. Dr J.A. Kortekaas, Wageningen University & Research
Prof. Dr W.H. Roos, University of Groningen
Dr M.A. Kostiainen, Aalto University, Finland
Dr J.C. Hohlbein, Wageningen University & Research

This research was conducted under the auspices of the Graduate School VLAG (Advanced studies in Food Technology, Agrobiotechnology, Nutrition and Health Sciences).

Artificial Viral Capsid Proteins

MIMICKING NATURE TO DELIVER MEDICINES

Lione Willems

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 24 May 2019
at 4 p.m. in the Aula.

Lione Willems
Artificial Viral Capsid Proteins - Mimicking nature to deliver medicines
142 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2019)
With references, with summary in English

ISBN: 978-94-6343-427-0
DOI: https://doi.org/10.18174/471210

Contents

1

General Introduction

Viruses are small infectious particles that can infect living cells and hijack their molecular machinery in order to multiply themselves. They are the causative agents of many diseases affecting members of all three domains of life: eukaryotes, bacteria and archaea. Although viral diseases such as the flu, measles and smallpox have caused troubles for millions of years[1], the fact that they are caused by small particles, now called viruses, was only discovered about 120 years ago[2]. Three scientists playing a key role in this discovery are Adolf Mayer, Dimitri Ivanovsky and Martinus Beijerinck. In 1886, the German chemist Adolf Mayer, who at that time was employed by the Agricultural School of Wageningen, recognised that the sap extracted from tobacco plants suffering from tobacco mosaic disease was an infectious substance. He applied the sap to healthy plants, which then showed signs of the disease within a few weeks. He could, however, not identify a contagious agent within the sap with the use of a light microscope. Next, in 1892, the Russian scientist Dmitri Ivanovsky reported that sap passed through a porcelain filter such that bacteria had been removed still remained infectious. He proposed the presence of bacterial toxins in the infectious sap that cause the disease. Finally, the Dutch scientist, Martinus Beijerinck, published a paper that is probably the seminal paper about the discovery of viruses. He learned about the tobacco mosaic disease when working as a colleague of Adolf Mayer at the Agricultural School of Wageningen. When he became a teacher at the Polytechnical Institute of Delft, he carried out a series of experiments, leading to the publication in 1898 about a contagious living fluid. Beijerinck discovered that a large number of healthy plants could be infected by a single diseased plant and therefore concluded that the infectious agent, which he believed was a contagium vivum fluidum (contagious living fluid), could reproduce itself in living plants[3]. For the contagious living fluid he coined the name “virus”, which is Latin for poisonous liquid. Only in the 1930s, it would be discovered that the infectious agent was not actually the fluid as a whole, but nm-sized particles within the fluid[2].

Viruses as a tool

As its Latin translation already suggests, viruses can cause illnesses, sometimes leading to the death of the organism. Indeed, outbreaks of viruses, like Zika in 2015[4] and Influenza (a.k.a. swine flu) in 2009[5], indicate the global impact viruses can have. In the eyes of most people therefore, the word “virus” mostly has a very negative connotation.

However, they are also remarkable nanoparticles that can be of great value in medicine and biotechnology[6]. Their ability to efficiently deliver their nucleic acid content into specific cells within a host organism is an incredible useful feature for applications such as gene therapy in which (genetic) disorders are treated via the delivery of therapeutic nucleic acids[7]. In addition, viruses come in a broad range of shapes and sizes making them useful as nano-containers and nano-delivery devices in a broad range of applications. Spherical particles, for example, are highly suitable for storage of nucleic acids, drugs or enzymes in their internal cavity, while rod-shaped particles, though having less internal storage space, tend

to penetrate deeper into tissues leading to better therapeutic efficiencies when used as a delivery device[8,9]. Next, we discuss the most important strategies currently used in the development of virus-derived and virus-inspired delivery vehicles (also see Figure 1.1).

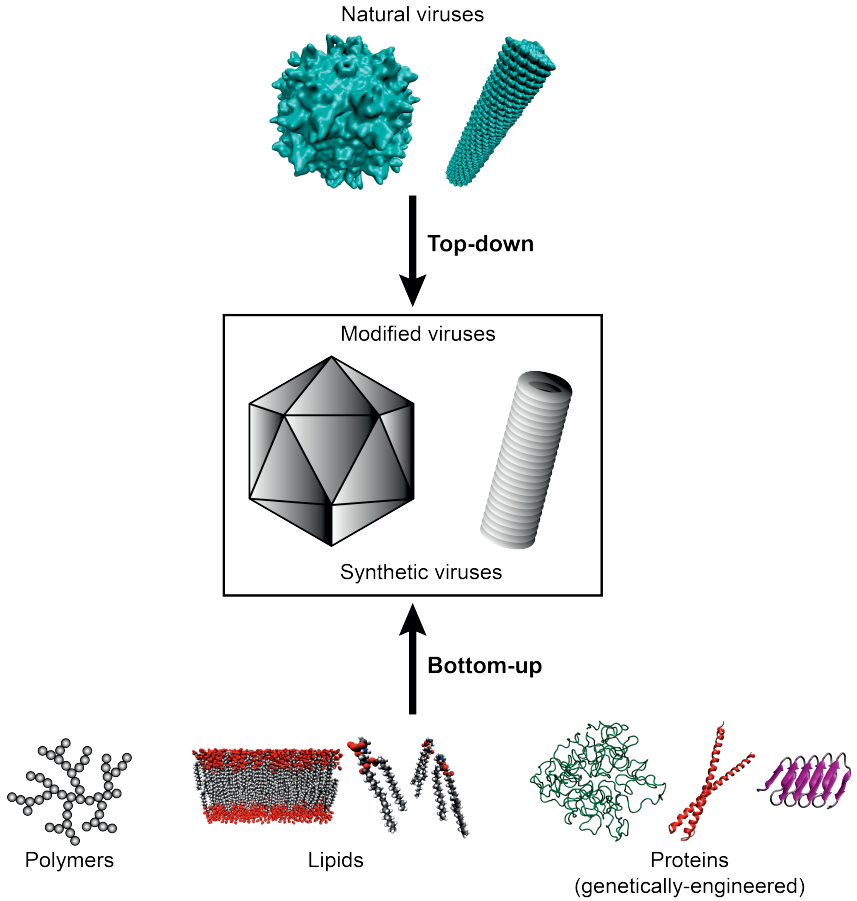


Figure 1.1 – Design of viral tools. Viral tools can be developed either via a top-down approach (top part of this figure) or via a bottom-up approach (bottom part of this figure). The top-down approach involves the modification of natural viruses in order to make them safe and more suitable for the intended application. Possible modifications can be the introduction of mutations in or the complete deletion of viral genes. In the bottom-up approach so-called synthetic viruses are created by using small non-viral materials as their building blocks. These materials can be based on polymers, lipids and (genetically-engineered) proteins.

Top-down design: modified natural viruses

The top-down approach consists in molecularly modifying natural viruses to remedy the real and perceived risks associated with their use in specific applications. To achieve this, in general, viral genes involved in replication and assembly are deleted from the encapsulated genome[10]. Viruses modified in this way are incapable of replicating in their host cells to form new virus particles and can therefore be used more safely, for example to deliver therapeutic nucleic acids into patients, such as in gene therapy. The first virus-based gene therapy that was approved in Europe or the United States is known as Glybera[11]. Glybera is composed of an adeno-associated virus that carries and delivers an active version of the lipoprotein lipase gene to patients with the rare genetic disease lipoprotein lipase deficiency[12]. Approval of this highly effective treatment by the European Commission in 2012 was a milestone in the field of gene therapy and shows the tremendous potential of virus-based vectors in medicine.

Virus particles have also been modified to make them suitable for use in many other applications such as for drug delivery and as nanoreactors for the production of chemicals. Virus particles, sometimes composed of the capsid protein only, can be loaded with therapeutic drugs, either on the inside in the internal cavity or via chemical coupling on the outside of the particles[13]. This way, the viral capsids may serve as drug-delivery devices, for example for the delivery of chemotherapeutics for the treatment of cancer. They may also be used as nanoreactors, by loading the viral capsids with enzymes that catalyse certain reactions in the production of useful chemicals[14]. In that case, the role of the virus is to act as a nano-container, for example by keeping in close proximity a set of enzymes and small molecules that together form a reaction cascade.

Finally, a form of modified viruses that everybody knows of is the vaccine. Vaccines prime our immune system already before an infection with certain natural viruses has taken place. Various types of vaccines exist that are based on modified natural viruses, for example, so-called live attenuated viruses, inactivated viruses, or so-called sub-unit vaccines[15]. That vaccination is a key strategy for the prevention of viral diseases is evident from the fact that vaccination programs have led to the world-wide eradication of smallpox[16] and to massive reductions in the number of reported cases for severe diseases such as polio, measles and tetanus (99.8%, 95.9% and 89.1% reduction since 1980 respectively)[17].

Despite the efficiency of virus-based tools for many applications, for each of these applications there are also risks and limitations in their use. Viral capsid proteins typically assemble into capsids with fixed dimensions, limiting the available space in their internal cavity for storage of molecules such as nucleic acids, drugs and enzymes. Furthermore, in many cases virus-based particles lack the required target-cell specificity, which may result in delivery of the contents into off-target cells and may have unintended side effects. Also, their safe use is often hampered by their inherent immunogenicity. Additionally, for viral tools that promote the incorporation of therapeutic DNA into the host cell genomes, there is a risk of insertional mutagenesis. This may cause the aberrant expression of host proteins and may eventually even result in the development of cancers[18,19]. Finally, there

are also real problems in the safe and efficient production of virus-based particles for therapeutic applications in view of the many requirements that they should obey[20].

Bottom-up design: synthetic viruses

Therefore, as an alternative many researchers have embarked on an ambitious long-term effort to start developing artificial virus-like particles from scratch, as a safe, more easy to be produced and better controlled alternative to virus-like particles derived from natural viruses. Indeed, synthetic viruses cannot reproduce in host cells, their safe and efficient production is much easier since they are built from a finite set of precisely controlled components, and the risk of insertional mutagenesis is minimal[21,22].

In the bottom-up approach, synthetic (or artificial) virus particles are developed that should mimic natural viruses in their ability to encapsulate and deliver nucleic acid molecules (or other therapeutic molecules). They are built by combining a finite number of well-defined molecular building blocks, like lipids and (protein) polymers, in a way such that it is clear what the function is of each of the components[21,22]. Such simplicity allows for better control and higher engineerability of synthetic viruses. As yet, the same simplicity unfortunately also means that in general they are not yet as efficient in delivery of nucleic acids and other cargoes as compared to the much more complex natural viruses. Nevertheless, it is widely believed that the bottom-up design of synthetic viruses will on the one hand help us to increase our understanding of how natural viruses work and on the other hand eventually lead to delivery devices that are safe and that rival the natural viruses in efficiency of delivery.

Going beyond synthetic viruses, the same molecular building blocks such as lipids and polymers are now not only used to design synthetic nanostructures that mimic natural viruses, but also to design synthetic nanostructures mimicking other biological structures such as organelles and even complete cells[23]. By building synthetic structures that mimic the structures in living organisms from scratch, we obtain valuable insights in what components would be minimally required for certain biological structures and hence for certain biological processes.

Finally, while in this thesis we focus only on synthetic materials for the delivery of nucleic acids, we wish to emphasise that the same design principles are also applicable for drug delivery vehicles more generally. Next, we describe in more detail the design challenges for synthetic viruses, and give an overview of what has been achieved in this area so far.

Requirements for the delivery of nucleic acids using synthetic viruses

Bottom-up designed synthetic viruses have to overcome several challenges[21,22,24] that natural viruses have successfully learnt to deal with during the course of evolution (Figure 1.2). Based on these challenges, a list of requirements can be put together for synthetic viruses as nucleic acid delivery vehicles:

1) Protection of cargo from degradation
2) Non-immunogenic and low clearance by kidneys and liver
3) Non-toxic
4) Facilitate cell binding and entry
5) Facilitate endosomal escape
6) In the case of DNA, delivery to the nucleus
7) Easy to produce

We discuss the different requirements one by one. First of all, the synthetic virus should protect its nucleic acid cargo against degradation (requirement 1). Naked nucleic acid molecules are readily cleaved by nucleases[25], causing low therapeutic efficiency. Synthetic viruses should thus shield their nucleic acid cargo from nucleases which are found in both blood and tissues. At the same time, the synthetic virus itself should not elicit immune responses (requirement 2), unless this is the purpose of the therapy as for vaccinations or cancer immunotherapy. Upon administration, nanoparticles may be recognised and internalised by immune cells, such as macrophages and dendritic cells, that eliminate foreign particles[26]. In addition, specific antibody responses may be generated which results in the rapid clearance of the nanoparticles upon repeated administration. Depending on the route of administration, nanoparticles are also subjected to clearance from the blood by the kidneys and the liver[27]. Particles smaller than 5 nm are excreted to the urine via filtration in the kidneys, while bigger nanoparticles may experience hepatic clearance due to uptake by hepatocytes or Kupffer cells (the macrophages of the liver)[27,28]. The synthetic viruses themselves and their degradation products should in addition not be toxic to cells (requirement 3). Toxicity may lead to cell death and unwanted side effects.

If the synthetic viruses succeed in avoiding the immune system, their next challenge is to deliver the nucleic acids to target cells. In order to do so, they should specifically bind to the plasma membrane of the target cell and facilitate entry, e.g. via passage through the plasma membrane or via endosomal uptake (requirement 4)[21,22,24]. Specific binding to target cells also limits off-target delivery and associated side effects[29].

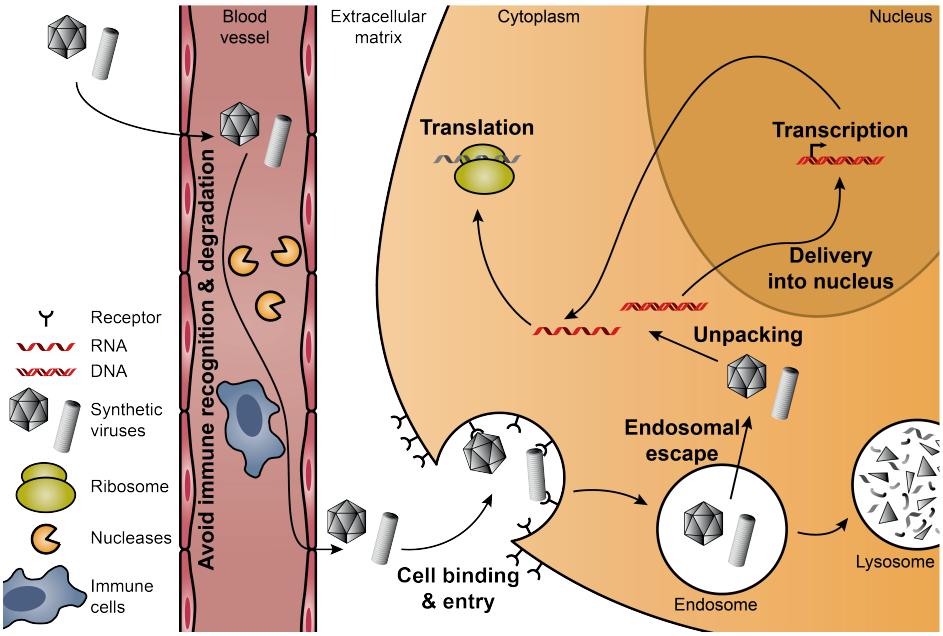


Figure 1.2 – Challenges during the delivery of nucleic acids by synthetic viruses. In order to deliver their nucleic acid cargo to the target cells, synthetic viruses need to overcome several challenges that otherwise (severely) reduce efficiency of delivery. Upon systemic administration (top-left), synthetic viruses should *avoid recognition by immune cells*, such as macrophages, and *prevent degradation* of their nucleic acid cargo by nucleases. When they arrive at the target site, they ideally specifically *bind to the target cells* only, after which they should be *internalised by the cells*. In general for synthetic viruses, endocytosis is the uptake pathway, so that they need to *escape from the endosome* to prevent degradation in the lysosome. Next, they should *release or unpack the nucleic acids*, and in case of DNA also facilitate *delivery into the nucleus*.

The next step in delivery process is to deliver the nucleic acids to the appropriate site within the cell, viz. to the nucleus for DNA and the cytoplasm for mRNA. Depending on the entry pathway and the type of nucleic acids that are delivered, various challenges exist. Most nanoparticles are internalised by the cell via endocytosis. The synthetic viruses should escape from these endosomes into the cytoplasm (requirement 5)[21,22,24]. If they fail to do so, the endosomes will fuse with lysosomes, resulting in the degradation of the synthetic viruses including the nucleic acids, such that there is no therapeutic effect. If the synthetic viruses do manage to enter the cytoplasm, the nucleic acid cargo should be released from the delivery vehicle. mRNA molecules can now immediately be used as templates for protein synthesis and this way exert their therapeutic effect. DNA molecules, however, have to overcome an additional challenge, as they should also pass the nuclear membrane prior to gene expression in the nucleus (requirement 6).

Taken together, synthetic viruses have to deal with many biological barriers. Not overcoming these barriers may reduce or even abolish the therapeutic efficiency of the delivery vehicles. In addition to these biological challenges, synthetic viruses should also be easy to produce (requirement 7). This requirement allows for the safe and efficient production of delivery vehicles at low cost which is in great contrast to natural viruses for which safe production is complex and expensive[20–22].

Current designs for synthetic viruses

Viruses are a major source of inspiration for nanomedicine and biotechnology, and many groups have worked and are currently working on the bottom-up design of virus-inspired synthetic materials for the encapsulation of nucleic acids. Various types of molecules are used as building blocks for the construction of these synthetic viruses, including lipids, polymers and proteins.

Lipid-based synthetic viruses

Given that many of the biological barriers to nucleic acid delivery are caused by the presence of lipid membranes, it is not surprising that lipids are one of the key types of molecules that have been explored for synthetic viruses. The lipids allow for fusion with the cellular membranes and for release of the cargo into the cytoplasm. In fact, the same mechanisms are exploited by many natural viruses that are enveloped by lipid membranes. Lipids are amphiphilic molecules possessing both hydrophilic and hydrophobic characteristics, and as a consequence have the ability to self-assemble into bilayer vesicles named liposomes. Liposomes carrying and protecting nucleic acid molecules can be easily produced (requirements 1 and 7) by simple self-assembly in the presence of the nucleic acids[30,31]. They have been shown to have good *in vitro* and *in vivo* transfection efficiencies and the therapeutic feasibility of several formulations is currently tested in clinical trials[32]. In fact, the first synthetic delivery vehicle approved in the US and Europe (Onpattro; in 2018) for the delivery of a small interfering RNA (siRNA) is based on lipids[33].

The high transfection efficiency for some formulations is thought to be specifically related to the positive charge on their headgroups. On the one hand this facilitates loading with negatively charged nucleic acids, and on the other hand it also facilitates adsorption onto the negatively charged cell membranes which is thought to promote endosomal escape via destabilisation of the endosomal membrane (requirement 4 and 5)[34,35]. Incorporation of fusogenic lipids such as dioleoylphosphatidylethanolamine (DOPE) likely further enhances membrane disruption upon acidification in the endosome[35–38]. Overall, liposomes tick many of the requirement boxes for the delivery of nucleic acids. Nevertheless, a major issue exists regarding their biodistribution (requirement 2). Systemically administered liposomes are mainly found in the liver where they are internalised by hepatocytes via receptor-mediated uptake, or cleared from the blood by Kupffer cells[39]. Unless the liver is the target of delivery, it remains extremely challenging to target specific locations within the body via the systemic route using lipid-based delivery vehicles.

Polymer-based synthetic viruses

Two well-known examples of polymers that are used for the protection and delivery of nucleic acids (requirement 1) are polyethylenimine (PEI) and polylysine[24,32]. These strongly positively charged polymers complex with nucleic acids through electrostatic interactions and, as is the case for cationic lipid particles, the excess positive charge also facilitates binding of the complexes to the negatively charged cell membranes (requirement 4). The polymer-based synthetic viruses are internalised via endocytosis[40], but their ability to also escape from the endosomes (requirement 5) into the cytoplasm is dependent on the type of polymer that is used. For PEI-based particles, the proton sponge hypothesis was proposed as the mechanism for endosomal escape[24,41,42]. At neutral pH, PEI is only partly protonated so that it can act as a buffer by binding the protons that are imported into the endosome. Consequently, acidification of the endosome is limited and as the proton influx is coupled to the influx of chloride anions, the ionic concentration increases. This will then lead to the influx of water and, due to the increased osmotic pressure, the swelling of the endosome. Swelling causes the endosomal membrane to become leaky, with endosomal escape of the PEI-based particles as the result. Polymers that do not exhibit the buffering capacity that PEI shows, such as polylysine which is already fully protonated at neutral pH, often do not allow for efficient endosomal escape and show much lower transfection efficiencies[43,44]. Modifying such polymers with ligands or reagents that facilitate endosomal escape may then be an effective strategy to improve transfection.

Although (modified) polymers do show the ability to deliver nucleic acids into cells, they are also associated with immune activation[45,46] and high cytotoxicity[47] (requirements 2 and 3). Immune activation results in clearance of the polymer complexes and is related to the charge of the polymers[45]. Also the degree of toxicity is related to the polymer characteristics, where an increase in particle size or branching typically results in higher cytotoxicity[48]. Furthermore, red blood cells and negatively charged proteins in the blood can interact with the positively charged polymers, causing the formation of toxic aggregates that inhibit delivery to the target tissues and that are rapidly cleared from the blood[47–49].

Often, transfection efficiencies of both polymer-based and lipid-based synthetic viruses can be enhanced through the coupling of the hydrophilic polymer polyethylene glycol (PEG) to the polymers or lipids[21,45,46]. Upon complexation with nucleic acids, the PEG forms a hydrophilic corona around the nanoparticle. The corona provides colloidal stability and, by shielding the positive charges, prevents binding of the complexes to red blood cells and negatively charged proteins in the blood[48].

Protein-based synthetic viruses

The design of protein-based (non-viral) synthetic viruses is relatively new in the fields of nanomedicine and biotechnology, so that the available data is limited compared to lipid- and polymer-based synthetic viruses. Nonetheless, proteins have the potential to become a highly important material for the design of synthetic viruses and delivery vehicles in general. The fact that they are based on proteins means that they can be closer mimics of natural viruses, for which the capsids

are also made out of protein. Furthermore, protein-based synthetic viruses are likely to be less toxic to cells as they are biodegradable (requirement 3). The third, and maybe most important reason why proteins are so interesting for the design of synthetic viruses, is the fact that they can be very precisely genetically engineered. Genetic engineering provides an excellent way to precisely control the amino acid sequence, length and molecular weight, which are crucial determinants for the biodistribution, clearance, transport and immunogenicity[50–52]. In addition, functionalities such as cell binding and endosomal escape (requirements 4 and 5) can be easily incorporated in protein designs via genetic engineering[53–55]. Protein components of synthetic viruses can be produced in production hosts, such as bacteria, yeast and plants[56–58], to allow for the easy manufacturing of protein-based synthetic viruses (requirement 7).

The self-assembly of proteins into nanostructures is not only typical for viral assembly, but is also a key process in the organisation of biological systems. Self-assembly of non-viral proteins, for example, provides structural rigidity to cells and tissues via the formation of a cytoskeleton and the extracellular matrix. For creating synthetic protein-based virus-like particles, we need not to only use proteins of viral origin, but we can be inspired by any natural protein, including structural proteins such as collagen[59] and its hydrolysed variant gelatin[59–61]. Also, non-viral proteins such as lumazine synthase[62] are being explored as inspiration and building blocks for delivery vehicles. The latter protein is in fact an enzyme that in some species assembles into icosahedral protein cages. Using genetic engineering, such proteins can be optimised for encapsulation of molecules such as green fluorescent protein and nucleic acids[63–65]. An interesting next step would be to evaluate the ability of these protein assemblies to deliver nucleic acids into cells.

Despite our current ability to precisely genetically engineer all kinds of proteins, the above mentioned natural proteins are not necessarily optimal building blocks for the design of synthetic viruses. Often, small changes to amino-acid sequences of natural self-assembling proteins have large consequences for their self-assembly and functional properties, which are difficult to predict or understand[66–69]. If we could *de novo* design the protein sequences for self-assembling protein capsids, one would hope that for such sequences we have much better control over the protein structure-function relationships. The design of such sequences may of course still be inspired by natural proteins.

In particular, sequence motifs found in structural proteins in nature, such as collagen, elastin and silk, can be used to generate proteins for novel purposes that combine a range of desired functions. In addition, it is now also possible to *de novo* design new protein sequences using computational methods. Using software, such as Rosetta, new protein sequences can be generated with the desired architectures and functions[36]. For example, Baker and colleagues recently reported the computational design of two proteins that co-assemble into a 120-subunit icosahedral protein nanostructure[70]. These proteins were able to encapsulate green fluorescent protein molecules through electrostatic interactions[70]. Interestingly, the design of these proteins could in addition be optimised for the encapsulation of mRNA by using directed evolution[71] and, in fact, this is the first time scientists report the design of a non-viral protein nanostructure that can encapsulate its own

genome. In the next section, we will focus on the *de novo* design of proteins using sequence motifs taken from natural structural proteins such as collagen, silk and elastin.

De novo design using sequence motifs from natural structural proteins

Often, natural structural proteins consist of highly repetitive sequences with many repeats of small motifs that adopt specific conformations and have specific physical properties. For example, the amino acid sequence of silk proteins from the silkworm *Bombyx mori* consists of an alternation of crystallisable and amorphous blocks[72], where the crystallisable blocks are repeats of characteristic glycine-alanine sequences such as (GAGAGS), and the amorphous blocks are very rich in prolines and glycines. Such characteristic short sequence motifs can be identified in many structural proteins. Using recombinant DNA technology, repeats of these motifs can be expressed in production hosts resulting in the production of recombinant protein polymers that behave similarly to the functional domains within the natural proteins, but are more homogeneous in their sequence. If multiple sequence motifs are combined, new recombinant protein block copolymers can be generated that have a new set of properties. In the next paragraphs, common sequence motifs in natural proteins and some examples of designed protein block copolymers are discussed (also see Figure 1.3a).

Collagen-like motif

Collagen is the main component of the extracellular matrix, providing structural support and tensile strength to tissues such as tendons, cartilage, bone and skin. It forms triple helices that (in the case of collagen type I) ultimately bundle into collagen fibres[73]. Characteristic for collagens is the repeat sequence (GXaaYaa) that is now known to be required for triple helix formation. The Xaa and Yaa positions within this sequence motif are often occupied by the amino acids proline and hydroxyproline respectively[74,75]. Hydroxylation of the proline at the Yaa position is one of the key processes in the production of functional collagens. The enzyme required for proline hydroxylation is not found in the common production hosts such as yeast and bacteria. Hence, functional recombinant protein polymers based on the collagen (GXaaYaa) motif were initially only produced in the less productive mammalian cells that do naturally express the enzyme. Later, the common production hosts such as yeast and bacteria were engineered to co-express collagen and the prolyl hydroxylase enzyme, such that functional collagens could also be produced in more productive hosts[76,77]. Recombinant collagens have now been developed that mimic the ability of natural collagens to provide support and tensile strength to tissues and have been used in various biomedical applications such as drug delivery and tissue engineering[76,78].

A collagen-inspired protein polymer that was first reported by Werten *et al.*[79], has been extensively used in the work that precedes the work described in this

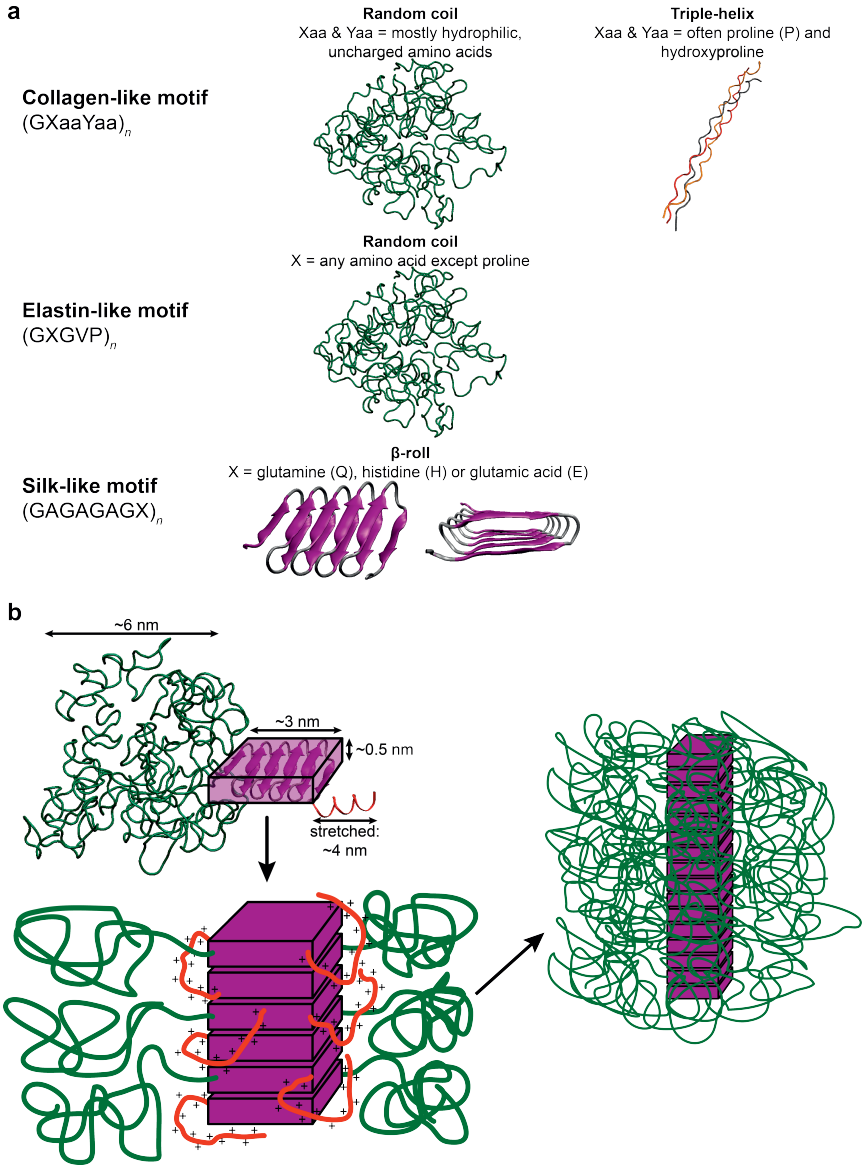


Figure 1.3 – *De novo* design of synthetic viruses based on natural protein motifs. a) Three examples of simple repeat motifs found in natural structural proteins (collagen-, elastin- and silk-like), including the amino acid sequence where *n* is the number of repeats and the adopted conformation of a protein block based on that sequence motif. b) *De novo* designed artificial viral capsid protein that mimics the capsid proteins of Tobacco Mosaic Virus. The protein consists of three blocks: an oligolysine in red, a silk-like motif in purple, and a random coil block in green that is based on either the collagen-like or the elastin-like motif.

thesis[80–82]. It also plays a role in this thesis since it is a potentially interesting protein polymer for the design of protein block copolymers that encapsulate and protect nucleic acids and other polyanionic molecules[83–85]. This polymer is a repeat of the (GXaaYaa) motif found in collagen but now the Xaa and Yaa positions are mostly occupied by hydrophilic, uncharged amino acids[79]. Unlike the triple helices formed by functional collagen, this highly hydrophilic protein polymer adopts a random coil conformation such that its solution structure more resembles the denatured or hydrolysed form of collagen known as gelatin.

Elastin-like motif

Similar to collagen, elastin is a structural protein that is abundantly present in the extracellular matrix of vertebrates including humans. It provides elasticity and strength to tissues such as arteries, skin, lungs and ligaments. The amino acid sequence of the precursor of elastin, tropoelastin, features an alternation of hydrophobic and hydrophilic blocks[86]. Upon secretion into the extracellular space, lysine residues within the hydrophilic domains are crosslinked to form the insoluble elastin fibrils that provide strength to tissues[86]. Elasticity of the fibrils was found to arise from the flexibility of the hydrophobic domains[87,88]. Several repeat sequences exist in the tropoelastin hydrophobic domains, but the (GXGVP) motif is the most commonly used to generate recombinant repeat proteins named elastin-like polypeptides (ELPs). Position X in this motif is the so-called guest residue. It can be occupied by any amino acid except proline.

ELPs show lower critical solution temperature (LCST) transition behaviour in aqueous solutions, which means that they transition from being in a one-phase state composed of soluble polymer chains at low temperature to a two-phase state with aggregated chains in a solution at high temperature. In the latter case, one phase is very dilute in the ELP, while the other phase contains the aggregated ELPs in solution. This type of phase separation, which is fully reversible, is sometimes also called coacervation. The temperature at which phase separation occurs is very precisely defined and is called the inverse transition temperature (T_t)[89,90]. Values of the T_t are sensitive to the chemical details of the ELPs (e.g. amino acid sequence and polymer chain length)[91–93] and to environmental conditions (e.g. salt concentration)[94]. Hence, by changing the chemical details of the ELPs it is possible to precisely tune the T_t.

At temperatures below the T_t, ELPs are intrinsically disordered, although several studies suggest the presence of β-turns within the structure of the soluble ELPs[90,95,96]. For simplicity we here consider ELPs below their LCST to be randomly coiled flexible polymers. Above their LCST on the other hand, ELPs are assumed to adopt more secondary structures like β-sheets, but the exact conformation is still under debate[90]. Because of their tunable LCST-type behaviour, ELPs are an interesting class of *de novo* designed proteins that are being used in for example protein purification, drug delivery and tissue engineering[97–99]. Another favourable property of ELPs for biomedical applications is that they have been found to be biocompatible[97,100].

Silk-like motif

Silk is a unique material that is mostly built from protein, and which has a mechanical strength that matches that of engineering materials such as Kevlar and steel[72,101]. Silks are produced by several insects and spiders, who use these materials for purposes such as protection and prey catching. For centuries, silk has been used by humans in the textile industry. More recently, it has also become an important material for biomedical applications[102–107]. Amino acid sequences of silk proteins feature an alternation of amorphous domains and crystallisable domains. The latter mediate local self-assembly of the silk proteins into nano-fibrils via the formation of β-sheets, thereby acting as cross-links and imparting strength to the material. On the other hand, the amorphous domains are extensible and make the silk material not only strong, but also difficult to break, or in other words tough. Depending on the species producing the silk, different sequence motifs can be identified in both the amorphous and the crystallisable domains. As for the crystallisable domains, while alanine repeats are commonly found in spiders, in the *Bombyx mori* silkworm the common motif consists of glycine-alanine repeats[72,101]. Based on the glycine-alanine (GA) repeats found in *B. mori*, Tirrell and coworkers designed and produced, in bacteria, protein-based polymers with the sequence $[(AG)_xEG]_n$. These were shown to form silk-like β-sheet assemblies[108].

In work at Wageningen University preceding the present work, various protein block copolymers were designed and produced in yeast that also contain sequence motifs inspired by the sequence of *B. mori* silks. This work used the motif (GAGAGAGX), with either glutamine, glutamic acid or histidine at position X. Proteins with repeats of this octapeptide were shown to self-assemble into fibrils due to the formation and stacking of β-rolls, wherein the first six amino acids of each octapeptide form the β-sheets and the seventh and eight GX residues are present in the turns[109–111]. Interestingly, the self-assembly behaviour could be controlled by choosing the residue at position X. Incorporation of glutamic acid or histidine results in a pH sensitive crystallisation, with glutamic acid promoting self-assembly at low pH[112,113] while silk-based polymers with histidine are soluble in acidic solutions and show self-assembly at neutral pH[114]. For glutamine at the X position, which is the choice made for the silk-like domains used in the present thesis, we cannot control the crystallisation of the β-sheets and the resulting fibril formation by tuning the pH. These proteins start crystallising and forming fibrils as soon as they are dissolved.

Protein block copolymers

By using recombinant DNA technology, many groups have designed protein-based block copolymers that combine silk-like blocks with either elastin-like or collagen-like blocks. Materials made out of such protein-based block copolymers combine the properties of the blocks from which they are made. For example, silk-elastin-like polypeptides (SELPs) combine the strength of the crystallisable blocks of *B. mori* silk with the flexibility of elastin, resulting in a tough biomaterial[115]. The properties of the biomaterial can be precisely tuned by changing for example the relative lengths of the silk- and elastin-like blocks[116], the type of guest residue X

used in the elastin-like block[92], or the type of corner residue X used in the silk-like crystallisable block[112–114]. Similarly, the molecular structures into which the protein polymers assemble can be tuned from assembly into micelles, to assembly into fibrils, to assembly into hydrogels[81,116,117]. All of these protein-polymer structures can be exploited in many ways for biomedical applications[100,118,119], for example in tissue engineering, drug delivery and more specifically, in gene delivery, which is the application closest to the work described in this thesis.

A *de novo* designed artificial viral capsid protein mimicking TMV capsid proteins

The work in the present thesis builds on earlier work done at Wageningen University, in which a silk-like, a collagen-like, and a simple polycationic motif were combined into a protein-based block copolymer that mimics the co-assembly of nucleic acids with viral capsid proteins[83]. The design of this protein was inspired especially by the Tobacco Mosaic Virus (TMV).

The protein consists of three blocks (see Figure 1.3b) each coding for a specific physical property of the TMV capsid proteins: nucleic acid binding, self-assembly and colloidal stability. For colloidal stability, the collagen-like sequence $C = (\mathrm{GXaaYaa})_{132}$ was used, which adopts a random coil conformation. For self-assembly into a rod-shaped particle, mimicking the rod-shaped TMV, a silk-like domain was used: $S^Q{}_{10} = (\mathrm{GAGAGAGQ})_{10}$. Finally, for binding to nucleic acid templates, an oligolysine K_{12} was used, since these are well known to strongly bind to nucleic acids (as well as to other polyanionic polymers)[83,84].

As for TMV, the final triblock protein C-$S^Q{}_{10}$-K_{12} self-assembles into rod-shaped, synthetic virus-like particles (VLPs) upon binding to polyanionic templates such as DNA[83,84]. The rod-shaped core of the VLPs is composed of the β-roll forming silk-like blocks $S^Q{}_{10}$ that stack on top of each other (Figure 1.3b). The polyanionic template is encapsulated in the rod-shaped particles through electrostatic interactions with the flexible K_{12} blocks attached to the rod-shaped silk-inspired core of the VLP. Finally, the amorphous C blocks form a corona around this silk-like core and the encapsulated template (Figure 1.3b). For encapsulated DNA and RNA templates[83,85], it was shown that they are protected from degradation by nucleases, which we hypothesise is due to the shielding effect of the amorphous C blocks.

Interestingly, encapsulation of polyanionic templates in the synthetic VLPs also leads to their condensation. It was found that for double-stranded linear DNA the ratio between the contour length of the template and the length of the rod-shaped VLPs was a constant factor (referred to as the packing factor) of $\sim$3 [83]. The silk-inspired block, that gives rise to the distinct rod-shape of the VLPs was found to be essential for condensation: proteins lacking the silk-like block (C-K_{12}), or having silk-like blocks too small to be able to allow for self-assembly into a rod-shaped fibril (C-$S^Q{}_2$-K_{12} and C-$S^Q{}_4$-K_{12}) fail to condense nucleic acid templates. Instead they simply coat them, while preserving the original contour length of the templates.

The *C*-S^{Q}_{10}-K_{12} protein triblock copolymers, also referred to as artificial viral capsid proteins, are potentially interesting for applications such as gene therapy and nucleic acid-based vaccinations. In addition, their simple design make them ideal model systems for studying the self-assembly, and maybe also other aspects, of natural viruses.

Improvements and design rules

In order to improve the design of the artificial viral capsid protein, so that it can be optimised for nanomedicine and biotechnology applications, we will first discuss whether the current design meets with the requirements discussed earlier in this introduction, and then discuss some other design rules proposed in the literature.

The *C*-S^{Q}_{10}-K_{12} proteins are produced in yeast, which is a relatively easy expression host (requirement 7). In addition, the design based on protein motifs allows for genetic engineering so that new designs can also be produced rather easily. As already mentioned above, the produced proteins are capable of protecting encapsulated nucleic acid molecules from nuclease activity[83,85], so that these proteins also meet with the first requirement discussed earlier (requirement 1). If the VLPs, formed after encapsulation of nucleic acids, are meant to be used inside the body, these VLPs should be non-immunogenic, non-toxic and exhibit a low clearance rate (requirements 2 and 3). Although experiments to determine these properties have not been performed yet for the *C*-S^{Q}_{10}-K_{12} VLPs, we believe there is a good chance that they are not very immunogenic or toxic since the VLP corona formed by the *C* block, is based on natural collagen of which plenty can be found in the human body.

When the purpose of the artificial viral capsid proteins is to deliver cargo to cells, they should also facilitate cell binding and entry (requirement 4), endosomal escape (requirement 5), and in the case of DNA cargo, also delivery to the nucleus (requirement 6). Limited data is currently available on the efficiency of the *C*-S^{Q}_{10}-K_{12} VLPs to deliver their nucleic acid cargo inside cells. For delivery of mRNA, a very low efficiency of transfection was observed in HEK293 cells, and in HeLa cells the efficiency was too low to be detected[85]. The particles have also been reported to allow for transfection of HeLa cells with plasmid DNA[83], but as was later established (MSc thesis of Hanna de Jong), the efficiency of transfection is again very low. Successful transfection requires a number of steps to each be successful (cell attachment, cell entry, endosomal escape, entry in nucleus for DNA) and it is not known which step, or which steps are problematic for the VLPs made with the artificial viral capsid proteins. Only after this has been established we would be able to further improve the properties of the artificial viral capsid proteins as delivery agents of nucleic acids.

Design rules for protein-based synthetic viruses

To implement requirements that are not yet covered by the current design of our artificial viral capsid protein or to optimise the design for other applications, various design rules should be taken into account. Design rules for viral and non-viral

nanoparticles are extensively discussed by Steinmetz *et al.*[6,120] and will be briefly summed-up here.

Firstly, the charge of a nanoparticle affects its binding to cell membranes, thereby influencing the probability to bind and enter target cells. Specifically, positively charged particles bind more efficiently to cells via electrostatic interactions since cell membranes are negatively charged.

An important consideration is also the size and shape of the nanoparticles. Larger nanoparticles have bigger surface areas with which they can make contacts with cells and this may facilitate cell binding. Larger nanoparticles, however, may also be more susceptible to hepatic clearance which would lower their efficiency. The size should thus be optimised for the application. Regarding the shape of a nanoparticle, rod-shaped architectures seem to contribute to higher delivery efficiencies as they seem to be able to penetrate deeper into tissues than spherical particles[8,9]. On the other hand, their narrow internal cavities may also be less suitable for cargoes other than nucleic acids.

An effective strategy to avoid immune recognition, reduce clearance and to protect the cargo is by shielding the nanoparticle from proteins in the blood and from components of the immune system. This can be achieved by applying a coat at the surface of the nanoparticle. The coats that work best are hydrophilic and flexible, and have a high enough density to fully cover the surface[121].

Most factors determining the efficiency of a nanoparticle, such as clearance and cell binding, are dependent on more than one of the above mentioned design properties. It is thus essential to optimise each of them in the context of each other. For example, if the design of a nanoparticle is improved with a shielding domain that further reduces the clearance of the particles, it may be possible to at the same time increase the size a bit. Increasing the size may lead to more clearance, but this is counteracted by the better shielding. Increasing the size, however, also may lead to better cell binding and subsequent entry.

Outline of this thesis

In this thesis we aim to increase our understanding of the physical (self- and co-assembly) behaviour of the artificial viral capsid proteins and its interaction with cells *in vitro*, in order to eventually improve its design for applications in nucleic acid delivery. We explore the modularity of the design by either replacing single protein blocks for other ones or attaching new functional groups, and by analysing how these changes influence the behaviour of the whole protein.

The original protein C-$S^Q{}_{10}$-K_{12} was produced previously in the yeast *Pichia Pastoris*. To facilitate more convenient production of a wide range of new artificial viral capsid proteins, we have first aimed at producing artificial capsid proteins in the bacterium *Escherichia coli*. To this end, in newly designed and produced proteins, the amorphous collagen-inspired C block was replaced by the amorphous elastin-inspired $E^S{}_m$ block with the amino acid sequence $(GSGVP)_m$. This obviously may have consequences for the self-assembly of the artificial viral capsid proteins, and this is the first topic that is addressed in this thesis.

First, in Chapter 2 we explore how amorphous elastin-like and silk-like crystallising blocks influence each other's behaviour in order to better understand how to optimise the design of the artificial viral capsid protein for template-induced assembly. To that purpose we designed and produced a series of nine $E^S{}_m$-$S^Q{}_n$ diblock proteins with varying lengths of the amorphous $E^S{}_m$ block and the crystallisable $S^Q{}_n$ block, where m = 40, 60 or 80 and n = 6, 10 or 14. We show that for self-assembly a minimum length of the crystallisable block is required and that just above this length self-assembly of the crystallisable block is suppressed with increasing lengths of the amorphous block. For longer lengths of the crystallisable block, there is not much influence of the amorphous block length on the self-assembly of the crystallisable block. We conclude therefore, that for template-induced assembly, suitable proteins should have a crystallisable $S^Q{}_n$ block length just above the minimum length required for crystallisation, such that the assembly behaviour can be fine-tuned by changing the amorphous $E^S{}_m$ block length.

Next, in Chapter 3 we study the effect of replacing the amorphous *C* block for the amorphous $E^S{}_m$ block on the self-assembly and co-assembly of the artificial viral capsid triblock proteins. We produced the new triblock protein $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} for which the amino acid length of the $E^S{}_{80}$ block is almost equal to the length of the *C* block. We show that despite the approximately equal lengths of the amorphous blocks, self-assembly without a template is much stronger for $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} than for *C*-$S^Q{}_{10}$-K_{12}. Presumably this is due to differences in the hydrophilicity of the two amorphous blocks. Co-assembly with a DNA template was however very similar for both proteins. Therefore we conclude that replacing the amorphous block with the elastin-inspired sequence does not affect the co-assembly behaviour of these proteins.

In Chapter 4 we continue to explore the modularity of the artificial viral capsid proteins by replacing the non-specific DNA binding domain K_{12} for a sequence-specific DNA binding domain, with the aim of making the capsid proteins sensitive to a packaging signal in the DNA template. This is a further step in mimicking relevant properties of natural viruses, and it could help to allow for packaging of specific sequences in complex environments such as the cytoplasm of living cells. The binding domain that we focus on is based on the zinc finger (ZnF) found in the *Drosophila melanogaster* GAGA factor. Similarly to some natural viral capsid proteins, we intend this $E^S{}_{80}$-$S^Q{}_{10}$-ZnF protein to first bind to a specific sequence in the DNA template (the packaging signal) which then triggers encapsulation of the rest of the template. We here show our work towards the development of such a packaging-signal sensitive version of the artificial viral capsid protein.

In Chapter 5, we focus on one of the key steps in the delivery of nucleic acids to cells for therapeutic applications, viz. the cell entry of the artificial virus particles. We visualise cell entry by fluorescently labelling both the *C*-$S^Q{}_{10}$-K_{12} proteins and the mRNA templates. In addition, we investigate whether cell entry is improved by conjugating an RGD cell-binding peptide to the N-terminus of the amorphous block of the *C*-$S^Q{}_{10}$-K_{12} proteins. We find that VLPs with mRNA readily enter the HeLa cells and presumably end up in endosomes. The level of uptake is found to only slightly increase by the presence of the RGD peptides, presumably, because of the shielding effects of the large and flexible amorphous *C* block.

Finally, in Chapter 6 we discuss the conclusions of our work, which we formulate as two new design rules for self-assembling synthetic viruses, and as additional recommendations for further improvements of their design. We also speculate on the possible conformation of a DNA encapsulated by the artificial viral capsid proteins. Finally we discuss how versions of the artificial viral capsid proteins (improved as suggested by us) could be used in a broad range of applications.

References

[1] B. A. Cunha, "Smallpox and measles: historical aspects and clinical differentiation," *Infectious Disease Clinics of North America*, vol. 18, no. 1, pp. 79–100, 2004.

[2] A. Lustig and A. J. Levine, "One Hundred Years of Virology," *Journal of Virology*, vol. 66, no. 8, pp. 4629–4631, 1992.

[3] M. Beijerinck, "Ueber ein Contagium vivum fluidum als Ursache der Fleckenkrankheit der Tabaksblätter.," *Verhandelingen der Koninklijke akademie van Wetenschappen te Amsterdam*, vol. 5, no. 2, pp. 3–21, 1898. Translated in English by J. Johnson as "Concerning a contagium vivum fluidum as cause of the sport disease of tobacco leaves," Phytopathologival Classics Number 7, 1942.

[4] A. S. Fauci and D. M. Morens, "Zika virus in the Americas - yet another arbovirus threat," *The New England Journal of Medicine*, vol. 374, no. 7, pp. 601–604, 2016.

[5] F. S. Dawood, A. D. Iuliano, C. Reed, M. I. Meltzer, D. K. Shay, P. Y. Cheng, D. Bandaranayake, R. F. Breiman, W. A. Brooks, P. Buchy, D. R. Feikin, K. B. Fowler, A. Gordon, N. T. Hien, P. Horby, Q. S. Huang, M. A. Katz, A. Krishnan, R. Lal, J. M. Montgomery, K. Mølbak, R. Pebody, A. M. Presanis, H. Razuri, A. Steens, Y. O. Tinoco, J. Wallinga, H. Yu, S. Vong, J. Bresee, and M. A. Widdowson, "Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: A modelling study," *The Lancet Infectious Diseases*, vol. 12, no. 9, pp. 687–695, 2012.

[6] A. M. Wen and N. F. Steinmetz, "Design of virus-based nanomaterials for medicine, biotechnology, and energy," *Chemical Society Reviews*, vol. 45, no. 15, pp. 4074–4126, 2016.

[7] K. Lundstrom, "Viral Vectors in Gene Therapy," *Diseases*, vol. 6, no. 2, p. 42, 2018.

[8] V. P. Chauhan, Z. Popović, O. Chen, J. Cui, D. Fukumura, M. G. Bawendi, and R. K. Jain, "Fluorescent nanorods and nanospheres for real-time in vivo probing of nanoparticle shape-dependent tumor penetration," *Angewandte Chemie - International Edition*, vol. 50, no. 48, pp. 11417–11420, 2011.

[9] R. Fernandes, N. R. Smyth, O. L. Muskens, S. Nitti, A. Heuer-Jungemann, M. R. Ardern-Jones, and A. G. Kanaras, "Interactions of skin with gold nanoparticles of different surface charge, shape, and functionality," *Small*, vol. 11, no. 6, pp. 713–721, 2015.

[10] D. Sharon and A. Kamen, "Advancements in the design and scalable production of viral gene transfer vectors," *Biotechnology and Bioengineering*, vol. 115, no. 1, pp. 25–40, 2018.

[11] S. L. Ginn, A. K. Amaya, I. E. Alexander, M. Edelstein, and M. R. Abedi, "Gene Therapy Clinical Trials Worldwide To 2017 - an Update," *The Journal of Gene Medicine*, no. January, p. e3015, 2018.

[12] J. J. Kastelein, C. J. Ross, and M. R. Hayden, "From Mutation Identification to Therapy: Discovery and Origins of the First Approved Gene Therapy in the Western World," *Human Gene Therapy*, vol. 24, no. 5, pp. 472–478, 2013.

[13] Y. Ma, R. J. Nolte, and J. J. Cornelissen, "Virus-based nanocarriers for drug delivery," *Advanced Drug Delivery Reviews*, vol. 64, no. 9, pp. 811–825, 2012.

[14] N. Carette, H. Engelkamp, E. Akpa, S. J. Pierre, N. R. Cameron, P. C. Christianen, J. C. Maan, J. C. Thies, R. Weberskirch, A. E. Rowan, R. J. Nolte, T. Michon, and J. C. Van Hest, "A virus-based biocatalyst," *Nature Nanotechnology*, vol. 2, no. 4, pp. 226–229, 2007.

[15] J. O. Josefsberg and B. Buckland, "Vaccine process technology," *Biotechnology and Bioengineering*, vol. 109, no. 6, pp. 1443–1460, 2012.

[16] "Smallpox," https://www.who.int/csr/disease/smallpox/en/, 14 Dec 2018.

[17] WHO, "Global immunization profile," tech. rep., 2018.

[18] C. E. Thomas, A. Ehrhardt, and M. A. Kay, "Progress and problems with the use of viral vectors for gene therapy," *Nature Reviews Genetics*, vol. 4, no. 5, pp. 346–358, 2003.

[19] D. Bouard, N. Alazard-Dany, and F. L. Cosset, "Viral vectors: From virology to transgene expression," *British Journal of Pharmacology*, vol. 157, no. 2, pp. 153–165, 2009.

[20] J. C. M. van der Loo and J. F. Wright, "Progress and challenges in viral vector manufacturing," *Human molecular genetics*, vol. 25, no. R1, pp. R42–R52, 2016.

[21] H. Yin, R. L. Kanasty, A. A. Eltoukhy, A. J. Vegas, J. R. Dorkin, and D. G. Anderson, "Non-viral vectors for gene-based therapy," *Nature Reviews Genetics*, vol. 15, no. 8, pp. 541–555, 2014.

[22] E. Mastrobattista, M. A. E. M. van der Aa, W. E. Hennink, and D. J. A. Crommelin, "Artificial viruses: a nanotechnological approach to gene delivery.," *Nature reviews. Drug discovery*, vol. 5, pp. 115–121, 2006.

[23] K. Powell, "How biologists are creating life-like cells from scratch.," *Nature*, vol. 563, no. 7730, pp. 172–175, 2018.

[24] U. Lächelt and E. Wagner, "Nucleic Acid Therapeutics Using Polyplexes : A Journey of 50 Years (and Beyond)," *Chemical Reviews*, vol. 115, no. 19, pp. 11043–11078, 2015.

[25] N. B. Tsui, E. K. Ng, and Y. M. Lo, "Stability of endogenous and added RNA in blood specimens, serum, and plasma," *Clinical Chemistry*, vol. 48, no. 10, pp. 1647–1653, 2002.

[26] M. A. Dobrovolskaia and S. E. McNeil, "Immunological properties of engineered nanomaterials," *Nature Nanotechnology*, vol. 2, no. 8, pp. 469–478, 2007.

[27] B. Wang, X. He, Z. Zhang, Y. Zhao, and W. Feng, "Metabolism of nanomaterials in vivo: Blood circulation and organ clearance," *Accounts of Chemical Research*, vol. 46, no. 3, pp. 761–769, 2013.

[28] K. M. Tsoi, S. A. Macparland, X. Z. Ma, V. N. Spetzler, J. Echeverri, B. Ouyang, S. M. Fadel, E. A. Sykes, N. Goldaracena, J. M. Kaths, J. B. Conneely, B. A. Alman, M. Selzner, M. A. Ostrowski, O. A. Adeyi, A. Zilman, I. D. McGilvray, and W. C. Chan, "Mechanism of hard-nanomaterial clearance by the liver," *Nature Materials*, vol. 15, no. 11, pp. 1212–1221, 2016.

[29] B. Yu, Y. Mao, L.-Y. Bai, S. E. Herman, X. Wang, A. Ramanunni, Y. Jin, X. Mo, C. Cheney, K. K. Chan, D. Jarjoura, G. Marcucci, R. J. Lee, J. C. Byrd, L. J. Lee, and N. Muthusamy, "Targeted Nanoparticle Delivery Overcomes Off-Target Immunostimulatory Effects of Oligonucleotides and Improves Therapeutic Efficacy in Chronic Lymphocytic Leukemia," *Blood*, vol. 121, no. 1, pp. 136–147, 2013.

[30] R. M. Levine, T. R. Pearce, M. Adil, and E. Kokkoli, "Preparation and characterization of liposome-encapsulated plasmid DNA for gene delivery," *Langmuir*, vol. 29, no. 29, pp. 9208–9215, 2013.

[31] A. A. Mokhtarieh, J. Lee, S. Kim, and M. K. Lee, "Preparation of siRNA encapsulated nanoliposomes suitable for siRNA delivery by simply discontinuous mixing," *Biochimica et Biophysica Acta - Biomembranes*, vol. 1860, no. 6, pp. 1318–1325, 2018.

[32] K. A. Hajj and K. A. Whitehead, "Tools for translation: Non-viral materials for therapeutic mRNA delivery," *Nature Reviews Materials*, vol. 2, pp. 1–17, 2017.

[33] S. M. Hoy, "Patisiran : First Global Approval," *Drugs*, vol. 78, no. 15, pp. 1625–1631, 2018.

[34] O. Zelphati and F. C. Szoka, "Mechanism of oligonucleotide release from cationic liposomes.," *Proceedings of the National Academy of Sciences*, vol. 93, no. 21, pp. 11493–11498, 1996.

[35] I. M. Hafez, N. Maurer, and P. R. Cullis, "On the mechanism whereby cationic lipids promote intracellular delivery of polynucleic acids," *Gene Therapy*, vol. 8, no. 15, pp. 1188–1196, 2001.

[36] H. Farhood, N. Serbina, and L. Huang, "The role of dioleoyl phosphatidylethanolamine in cationic liposome mediated gene transfer," *BBA - Biomembranes*, vol. 1235, no. 2, pp. 289–295, 1995.

[37] S. Mochizuki, N. Kanegae, K. Nishina, Y. Kamikawa, K. Koiwai, H. Masunaga, and K. Sakurai, "The role of the helper lipid dioleoylphosphatidylethanolamine (DOPE) for DNA transfection cooperating with a cationic lipid bearing ethylenediamine," *Biochimica et Biophysica Acta - Biomembranes*, vol. 1828, no. 2, pp. 412–418, 2013.

[38] K. J. Kauffman, J. R. Dorkin, J. H. Yang, M. W. Heartlein, F. Derosa, F. F. Mir, O. S. Fenton, and D. G. Anderson, "Optimization of Lipid Nanoparticle Formulations for mRNA Delivery in Vivo with Fractional Factorial and Definitive Screening Designs," *Nano Letters*, vol. 15, no. 11, pp. 7300–7306, 2015.

[39] X. Yan, F. Kuipers, L. M. Havekes, R. Havinga, B. Dontje, K. Poelstra, G. L. Scherphof, and J. A. Kamps, "The role of apolipoprotein E in the elimination of liposomes from blood by hepatocytes in the mouse," *Biochemical and Biophysical Research Communications*, vol. 328, no. 1, pp. 57–62, 2005.

[40] W. Godbey, K. K. Wu, and A. Mikos, "Tracking the intracellular path of poly(ethylenimine)/DNA complexes for gene delivery," *Proc. Nat. Acad. Sci. USA*, vol. 96, no. 9, pp. 5177–5181, 1999.

[41] O. Boussif, F. Lezoualc'h, M. A. Zanta, M. D. Mergny, D. Scherman, B. Demeneix, and J.-P. Behr, "A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: Polyethylenimine," *Proc. Nat. Acad. Sci. USA*, vol. 92, no. 16, pp. 7297–7301, 1995.

[42] J.-P. Behr, "The Proton Sponge: a Trick to Enter Cells the Viruses Did Not Exploit," *Chimia*, vol. 51, pp. 34–36, 1997.

[43] A. Akinc, M. Thomas, A. M. Klibanov, and R. Langer, "Exploring polyethylenimine-mediated DNA transfection and the proton sponge hypothesis," *The Journal of Gene Medicine*, vol. 7, no. 5, pp. 657–663, 2005.

[44] K. Itaka, A. Harada, Y. Yamasaki, K. Nakamura, H. Kawaguchi, and K. Kataoka, "In situ single cell observation by fluorescence resonance energy transfer reveals fast intra-cytoplasmic delivery and easy release of plasmid DNA complexed with linear polyethylenimine," *The Journal of Gene Medicine*, vol. 6, no. 1, pp. 76–84, 2004.

[45] C. Plank, K. Mechtler, F. C. Szoka Jr., and E. Wagner, "Activation of the complement system by synthetic DNA complexes: a potential barrier for intravenous gene delivery," *Human Gene Therapy*, vol. 7, no. 12, pp. 1437–1446, 1996.

[46] O. M. Merkel, R. Urbanics, P. Bedocs, Z. Rozsnyay, L. Rosivall, M. Toth, T. Kissel, and J. Szebeni, "In vitro and in vivo complement activation and related anaphylactic effects associated with polyethylenimine and polyethylenimine- graft -poly (ethylene glycol) block copolymers," *Biomaterials*, vol. 32, no. 21, pp. 4936–4942, 2011.

[47] H. Lv, S. Zhang, B. Wang, S. Cui, and J. Yan, "Toxicity of cationic lipids and cationic polymers in gene delivery," *Journal of Controlled Release*, vol. 114, no. 1, pp. 100–109, 2006.

[48] F. Alexis, E. Pridgen, L. K. Molnar, and O. C. Farokhzad, "Factors affecting the clearance and biodistribution of polymeric nanoparticles," *Molecular Pharmaceutics*, vol. 5, no. 4, pp. 505–515, 2008.

[49] E. Moreau, M. Domurado, P. Chapon, M. Vert, and D. Domurado, "Biocompatibility of polycations: In vitro agglutination and lysis of red blood cells and in vivo toxicity," *Journal of Drug Targeting*, vol. 10, no. 2, pp. 161–173, 2002.

[50] a. Nagarsekar and H. Ghandehari, "Genetically engineered polymers for drug delivery," *Journal of drug targeting*, vol. 7, no. 1, pp. 11–32, 1999.

[51] D. Chow, M. L. Nunalee, D. W. Lim, A. J. Simnick, and A. Chilkoti, "Peptide-based biopolymers in biomedicine and biotechnology," *Materials Science and Engineering R: Reports*, vol. 62, no. 4, pp. 125–155, 2008.

[52] O. S. Rabotyagova, P. Cebe, and D. L. Kaplan, "Protein-based block copolymers," *Biomacromolecules*, vol. 12, pp. 269–289, feb 2011.

[53] W. Li, F. Nicol, and F. C. Szoka, "GALA: A designed synthetic pH-responsive amphipathic peptide with applications in drug and gene delivery," *Advanced Drug Delivery Reviews*, vol. 56, no. 7, pp. 967–985, 2004.

[54] M. B. Elsner, H. M. Herold, S. Müller-Herrmann, H. Bargel, and T. Scheibel, "Enhanced cellular uptake of engineered spider silk particles," *Biomaterials Science*, vol. 3, no. 3, pp. 543–551, 2015.

[55] K. Numata, J. Hamasaki, B. Subramanian, and D. L. Kaplan, "Gene delivery mediated by recombinant silk proteins containing cationic and cell binding motifs," *Journal of Controlled Release*, vol. 146, no. 1, pp. 136–143, 2010.

[56] W. Hassouneh, T. Christensen, and A. Chilkoti, "Elastin-like polypeptides as a purification tag for recombinant proteins," in *Current Protocols in Protein Science*, pp. 6.11.1–6.11.16, aug 2010.

[57] R. Schipperus, R. L. Teeuwen, M. W. Werten, G. Eggink, and F. A. De Wolf, "Secreted production of an elastin-like polypeptide by Pichia pastoris," *Applied Microbiology and Biotechnology*, vol. 85, no. 2, pp. 293–301, 2009.

[58] J. Scheller, K.-H. Gührs, F. Grosse, and U. Conrad, "Production of spider silk proteins in tobacco and potato," *Nature Biotechnology*, vol. 19, no. 6, pp. 573–577, 2001.

[59] J. M. Dang and K. W. Leong, "Natural polymers for gene delivery and tissue engineering," *Advanced Drug Delivery Reviews*, vol. 58, no. 4, pp. 487–499, 2006.

[60] C. Bourquin, C. Wurzenberger, S. Heidegger, S. Fuchs, D. Anz, S. Weigel, N. Sandholzer, G. Winter, C. Coester, and S. Endres, "Delivery of Immunostimulatory RNA Oligonucleotides by Gelatin Nanoparticles Triggers an Efficient Antitumoral Response," *Journal of Immunotherapy*, vol. 33, no. 9, pp. 935–944, 2010.

[61] I.-D. Kim, E. Sawicki, H.-K. Lee, E.-H. Lee, H. J. Park, P.-L. Han, K. K. Kim, H. Choi, and J.-K. Lee, "Robust neuroprotective effects of intranasally delivered iNOS siRNA encapsulated in gelatin nanoparticles in the postischemic brain," *Nanomedicine: Nanotechnology, Biology, and Medicine*, vol. 12, no. 5, pp. 1219–1229, 2016.

[62] Y. Wei, P. Kumar, N. Wahome, N. J. Mantis, and C. R. Middaugh, "Biomedical Applications of Lumazine Synthase," *Journal of Pharmaceutical Sciences*, vol. 107, no. 9, pp. 2283–2296, 2018.

[63] F. P. Seebeck, K. J. Woycechowsky, W. Zhuang, J. P. Rabe, and D. Hilvert, "A Simple Tagging System for Protein Encapsulation," *Journal of the American Chemical Society*, vol. 128, no. 14, pp. 4516–4517, 2006.

[64] S. Lilavivat, D. Sardar, S. Jana, G. C. Thomas, and K. J. Woycechowsky, "In Vivo Encapsulation of Nucleic Acids Using an Engineered Nonviral Protein Capsid," *Journal of the American Chemical Society*, vol. 134, no. 32, pp. 13152–13155, 2012.

[65] N. Terasaka, Y. Azuma, and D. Hilvert, "Laboratory evolution of virus-like nucleocapsids from nonviral protein cages," *Proceedings of the National Academy of Sciences*, vol. 115, no. 21, pp. 5432–5437, 2018.

[66] N. A. Carter, X. Geng, and T. Z. Grove, *Design of self-assembling protein-polymer conjugates*, vol. 940. 2016.

[67] J. W. Bryson, S. F. Betz, H. S. Lu, D. J. Suich, H. X. Zhou, K. T. O'Neil, and W. F. DeGrado, "Protein Design: A Hierarchic Approach," *Science*, vol. 270, no. 5238, pp. 935–941, 1995.

[68] P. Burkhard, J. Stetefeld, and S. V. Strelkov, "Coiled coils: A highly versatile protein folding motif," *Trends in Cell Biology*, vol. 11, no. 2, pp. 82–88, 2001.

[69] W. F. DeGrado, C. M. Summa, V. Pavone, F. Nastri, and A. Lombardi, "De novo design and structural characterization of proteins and metalloproteins," *Annual Review of Biochemistry*, vol. 68, pp. 779–819, 1999.

[70] J. B. Bale, S. Gonen, Y. Liu, W. Sheffler, D. Ellis, C. Thomas, D. Cascio, T. O. Yeates, T. Gonen, N. P. King, and D. Baker, "Accurate design of megadalton-scale two-component icosahedral protein complexes," *Science (New York, N.Y.)*, vol. 353, no. 6297, pp. 389–394, 2016.

[71] G. L. Butterfield, M. J. Lajoie, H. H. Gustafson, D. L. Sellers, U. Nattermann, D. Ellis, J. B. Bale, S. Ke, G. H. Lenz, A. Yehdego, R. Ravichandran, S. H. Pun, N. P. King, and D. Baker, "Evolution of a designed protein assembly encapsulating its own RNA genome," *Nature*, vol. 552, no. 7685, pp. 415–420, 2017.

[72] C. Vepari and D. L. Kaplan, "Silk as a biomaterial," *Progress in Polymer Science (Oxford)*, vol. 32, no. 8-9, pp. 991–1007, 2007.

[73] K. Gelse, E. Pöschl, and T. Aigner, "Collagens - structure, function, and biosynthesis," *Advanced Drug Delivery Reviews*, vol. 55, no. 12, pp. 1531–1546, 2003.

[74] K. E. Kadler, C. Baldock, J. Bella, and R. P. Boot-Handford, "Collagens at a glance," *Journal of Cell Science*, vol. 120, pp. 1955–1958, 2007.

[75] M. K. Gordon and R. A. Hahn, "Collagens," *Cell and Tissue Research*, vol. 339, no. 1, pp. 247–257, 2010.

[76] D. Olsen, C. Yang, M. Bodo, R. Chang, S. Leigh, J. Baez, D. Carmichael, M. Perälä, E.-R. Hämäläinen, M. Jarvinen, and J. Polarek, "Recombinant collagen and gelatin for drug delivery," *Advanced Drug Delivery Reviews*, vol. 55, no. 12, pp. 1547–1567, 2003.

[77] C. Rutschmann, S. Baumann, J. Cabalzar, K. B. Luther, and T. Hennet, "Recombinant expression of hydroxylated human collagen in Escherichia coli.," *Applied microbiology and biotechnology*, vol. 98, pp. 4445–55, may 2014.

[78] C. Yang, P. J. Hillas, J. A. Báez, M. Nokelainen, J. Balan, J. Tang, R. Spiro, and J. W. Polarek, "The Application of Recombinant Human Collagen in Tissue Engineering," *BioDrugs*, vol. 18, no. 2, pp. 103–119, 2004.

[79] M. W. T. Werten, W. H. Wisselink, T. J. Jansen-van den Bosch, E. C. de Bruin, and F. a. de Wolf, "Secreted production of a custom-designed, highly hydrophilic gelatin in Pichia pastoris," *Protein Engineering Design and Selection*, vol. 14, pp. 447–454, jun 2001.

[80] M. K. Włodarczyk-Biegun, K. Farbod, M. W. Werten, C. J. Slingerland, F. A. De Wolf, J. J. Van Den Beucken, S. C. Leeuwenburgh, M. A. Cohen Stuart, and M. Kamperman, "Fibrous hydrogels for cell encapsulation: A modular and supramolecular approach," *PLoS ONE*, vol. 11, no. 5, 2016.

[81] L. H. Beun, I. M. Storm, M. W. T. Werten, F. A. De Wolf, M. A. Cohen Stuart, and R. De Vries, "From micelles to fibers: Balancing self-assembling and random coiling domains in pH-responsive silk-collagen-like protein-based polymers," *Biomacromolecules*, vol. 15, no. 9, pp. 3349–3357, 2014.

[82] N. E. Domeradzka, M. W. Werten, F. A. De Wolf, and R. De Vries, "Cross-Linking and Bundling of Self-Assembled Protein-Based Polymer Fibrils via Heterodimeric Coiled Coils," *Biomacromolecules*, vol. 17, no. 12, pp. 3893–3901, 2016.

[83] A. Hernandez-Garcia, D. J. Kraft, A. F. J. Janssen, P. H. H. Bomans, N. A. J. M. Sommerdijk, D. M. E. Thies-Weesie, M. E. Favretto, R. Brock, F. a. de Wolf, M. W. T. Werten, P. van der Schoot, M. C. Stuart, and R. de Vries, "Design and self-assembly of simple coat proteins for artificial viruses," *Nature Nanotechnology*, vol. 9, pp. 698–702, aug 2014.

[84] A. Hernandez-Garcia, M. A. Cohen Stuart, and R. de Vries, "Templated co-assembly into nanorods of polyanions and artificial virus capsid proteins," *Soft Matter*, vol. 14, no. 1, pp. 132–139, 2018.

[85] S. Jekhmane, R. de Haas, O. Paulino da Silva Filho, A. H. van Asbeck, M. E. Favretto, A. Hernandez Garcia, R. Brock, and R. de Vries, "Virus-Like Particles of mRNA with Artificial Minimal Coat Proteins: Particle Formation, Stability, and Transfection Efficiency," *Nucleic Acid Therapeutics*, vol. 27, no. 3, pp. 159–167, 2017.

[86] B. Vrhovski and A. S. Weiss, "Biochemistry of tropoelastin," *European Journal of Biochemistry*, vol. 258, no. 1, pp. 1–18, 1998.

[87] B. Bochicchio, A. Pepe, and A. M. Tamburro, "Investigating by CD the Molecular Mechanism of Elasticity of Elastomeric Proteins," *Chirality*, vol. 20, no. 9, pp. 985–994, 2008.

[88] S. Rauscher and R. Pomès, "Structural Disorder and Protein Elasticity," in *Fuzziness: Structural Disorder in Protein Complexes* (M. Fuxreiter and P. Tompa, eds.), vol. 1, pp. 159–183, New York, NY: Springer US, 2012.

[89] D. W. Urry, "Physical Chemistry of Biological Free Energy Transduction As Demonstrated by Elastic Protein-Based Polymers ," *The Journal of Physical Chemistry B*, vol. 101, no. 51, pp. 11007–11028, 1997.

[90] N. K. Li, F. Garcia Quiroz, C. K. Hall, A. Chilkoti, and Y. G. Yingling, "Molecular description of the LCST behavior of an elastin-like polypeptide.," *Biomacromolecules*, 2014.

[91] D. W. Urry, C.-H. Luan, T. M. Parker, D. C. Gowda, K. U. Prasad, M. C. Reid, and A. Safavy, "Temperature of polypeptide inverse temperature transition depends on mean residue hydrophobicity," *Journal of the American Chemical Society*, vol. 113, no. 11, pp. 4346–4348, 1991.

[92] J. R. Mcdaniel, D. C. Radford, and A. Chilkoti, "A unified model for de novo design of elastin-like polypeptides with tunable inverse transition temperatures," *Biomacromolecules*, vol. 14, pp. 2866–2872, 2013.

[93] D. E. Meyer and A. Chilkoti, "Quantification of the effects of chain length and concentration on the thermal behavior of elastin-like polypeptides," *Biomacromolecules*, vol. 5, no. 3, pp. 846–851, 2004.

[94] Y. Cho, Y. Zhang, T. Christensen, L. B. Sagle, A. Chilkoti, and P. S. Cremer, "Effects of hofmeister anions on the phase transition temperature of elastin-like polypeptides," *Journal of Physical Chemistry B*, vol. 112, no. 44, pp. 13765–13771, 2008.

[95] N. K. Li, S. Roberts, F. G. Quiroz, A. Chilkoti, and Y. G. Yingling, "Sequence Directionality Dramatically Affects LCST Behavior of Elastin-Like Polypeptides," *Biomacromolecules*, vol. 19, no. 7, pp. 2496–2505, 2018.

[96] M. Hong, D. Isailovic, R. A. Mcmillan, and V. P. Conticello, "Structure of an Elastin-Mimetic Polypeptide by Solid-State NMR Chemical," *Biopolymers*, vol. 70, no. 2, pp. 158–168, 2003.

[97] S. R. MacEwan and A. Chilkoti, "Elastin-like polypeptides: biomedical applications of tunable biopolymers.," *Biopolymers*, vol. 94, pp. 60–77, jan 2010.

[98] J. R. McDaniel, D. J. Callahan, and A. Chilkoti, "Drug delivery to solid tumors by elastin-like polypeptides.," *Advanced drug delivery reviews*, vol. 62, pp. 1456–1467, dec 2010.

[99] D. E. Meyer and a. Chilkoti, "Purification of recombinant proteins by fusion with thermally-responsive polypeptides.," *Nature biotechnology*, vol. 17, no. 11, pp. 1112–1115, 1999.

[100] W. Huang, A. Rollett, and D. L. Kaplan, "Silk-elastin-like protein biomaterials for the controlled delivery of therapeutics," *Expert Opinion on Drug Delivery*, vol. 12, no. 5, pp. 779–791, 2015.

[101] J. G. Hardy, L. M. Romer, and T. R. Scheibel, "Polymeric materials based on silk proteins," *Polymer*, vol. 49, no. 20, pp. 4309–4327, 2008.

[102] G. H. Altman, F. Diaz, C. Jakuba, T. Calabro, R. L. Horan, J. Chen, H. Lu, J. Richmond, and D. L. Kaplan, "Silk-based biomaterials," *Biomaterials*, vol. 24, no. 3, pp. 401–416, 2003.

[103] A. S. Lammel, X. Hu, S. H. Park, D. L. Kaplan, and T. R. Scheibel, "Controlling silk fibroin particle features for drug delivery," *Biomaterials*, vol. 31, no. 16, pp. 4583–4591, 2010.

[104] V. E. Bosio, J. Brown, M. J. Rodriguez, and D. L. Kaplan, "Biodegradable porous silk microtubes for tissue vascularization," *J. Mater. Chem. B*, vol. 5, no. 6, pp. 1227–1235, 2017.

[105] E. S. Gil, B. Panilaitis, E. Bellas, and D. L. Kaplan, "Functionalized Silk Biomaterials for Wound Healing," *Advanced Healthcare Materials*, vol. 2, no. 1, pp. 206–217, 2013.

[106] W. Zhang, C. Zhu, D. Ye, L. Xu, X. Zhang, Q. Wu, X. Zhang, D. L. Kaplan, and X. Jiang, "Porous silk scaffolds for delivery of growth factors and stem cells to enhance bone regeneration," *PLoS ONE*, vol. 9, no. 7, pp. 1–9, 2014.

[107] A. E. Thurber, F. G. Omenetto, and D. L. Kaplan, "In vivo bioresponses to silk proteins," *Biomaterials*, vol. 71, pp. 145–157, 2015.

[108] M. T. Krejchi, E. D. T. Atkins, A. J. Waddon, M. J. Fournier, T. L. Mason, and D. A. Tirrell, "Chemical sequence control of B-sheet assembly in macromolecular crystals of periodic polypeptides," *Science (New York, N.Y.)*, vol. 265, pp. 1427–1432, 1994.

[109] M. Schor, A. a. Martens, F. a. DeWolf, M. a. Cohen Stuart, and P. G. Bolhuis, "Prediction of solvent dependent β-roll formation of a self-assembling silk-like protein domain," *Soft Matter*, vol. 5, no. 13, pp. 2658–2665, 2009.

[110] B. Zhao, M. A. Cohen Stuart, and C. K. Hall, "Dock n roll: folding of a silk-inspired polypeptide into an amyloid-like beta solenoid," *Soft Matter*, 2016.

[111] B. Zhao, M. A. Cohen Stuart, and C. K. Hall, "Navigating in foldonia: Using accelerated molecular dynamics to explore stability, unfolding and self-healing of the β-solenoid structure formed by a silk-like polypeptide," *PLOS Computational Biology*, vol. 13, no. 3, p. e1005446, 2017.

[112] A. a. Martens, G. Portale, M. W. T. Werten, R. J. de Vries, G. Eggink, M. A. Cohen Stuart, and F. A. de Wolf, "Triblock protein copolymers forming supramolecular nanotapes and pH-responsive gels," *Macromolecules*, vol. 42, pp. 1002–1009, feb 2009.

[113] M. D. Golinska, M. K. Włodarczyk-Biegun, M. W. Werten, M. A. Stuart, F. A. De Wolf, and R. De Vries, "Dilute self-healing hydrogels of silk-collagen-like block copolypeptides at neutral pH," *Biomacromolecules*, vol. 15, no. 3, pp. 699–706, 2014.

[114] L. H. Beun, X. J. Beaudoux, J. M. Kleijn, F. a. De Wolf, and M. a. Cohen Stuart, "Self-assembly of silk-collagen-like triblock copolymers resembles a supramolecular living polymerization," *ACS Nano*, vol. 6, no. 1, pp. 133–140, 2012.

[115] W. Teng, J. Cappello, and X. Wu, "Recombinant silk-elastinlike protein polymer displays elasticity comparable to elastin," *Biomacromolecules*, vol. 10, no. 11, pp. 3028–3036, 2009.

[116] X.-x. Xia, Q. Xu, X. Hu, G. Qin, and D. L. Kaplan, "Tunable Self-Assembly of Genetically Engineered Silk Elastin-like Protein Polymers," 2011.

[117] M. D. Golinska, T. T. H. Pham, M. W. T. Werten, F. A. D. Wolf, M. A. C. Stuart, and J. V. D. Gucht, "Fibril Formation by pH and Temperature Responsive Silk-Elastin Block Copolymers," *Biomacromolecules*, vol. 14, no. 1, pp. 48–55, 2013.

[118] W. Qiu, Y. Huang, W. Teng, C. M. Cohn, J. Cappello, and X. Wu, "Complete Recombinant Silk-Elastinlike Protein-Based Tissue Scaffold," *Biomacromolecules*, vol. 11, no. 12, pp. 3219–3227, 2010.

[119] Z. Megeed, M. Haider, D. Li, B. W. O'Malley, J. Cappello, and H. Ghandehari, "In vitro and in vivo evaluation of recombinant silk-elastinlike hydrogels for cancer gene therapy," *Journal of Controlled Release*, vol. 94, pp. 433–445, feb 2004.

[120] A. M. Wen, P. H. Rambhia, R. H. French, and N. F. Steinmetz, "Design rules for nanomedical engineering: From physical virology to the applications of virus-based materials in medicine," *Journal of Biological Physics*, vol. 39, no. 2, pp. 301–325, 2013.

[121] A. Vonarbourg, C. Passirani, P. Saulnier, and J. P. Benoit, "Parameters influencing the stealthiness of colloidal drug delivery systems," *Biomaterials*, vol. 27, no. 24, pp. 4356–4373, 2006.

2

Inducible fibril formation of silk-elastin diblocks

Silk-elastin block copolymers have physical and biological properties that make them attractive biomaterials for applications ranging from tissue regeneration to drug delivery. Silk-elastin block copolymers that only assemble into fibrils at high concentrations can be used for template-induced fibril assembly. This can be achieved by additionally including template-binding blocks that promote high local concentrations of polymers on the template, leading to template-induced fibril assembly. We hypothesise that template inducible silk-fibril formation, and hence high critical concentrations for fibril formation require careful tuning of the block lengths, to be close to a critical set of block lengths that separates fibril forming from non-fibril forming polymer architectures. Therefore, we explore herein the impact of tuning block lengths for silk-elastin diblock proteins on fibril formation. For silk-elastin diblocks $E^S{}_m$-$S^Q{}_n$ *, in which the elastin pentamer repeat is* E^S *= GSGVP and the crystallisable silk octamer repeat is* S^Q *= GAGAGAGQ, we find that no fibril formation occurs for* n *= 6, but that the* n *= 10 and 14 diblocks do show concentration-dependent fibril formation. For* n *= 14 diblocks no effect is observed of the length* m *(with* m *= 40, 60, 80) of the amorphous block on the lengths of the fibrils. In contrast, for the* n *= 10 diblocks that are closest to the critical boundary for fibril formation, we find that long amorphous blocks (*m *= 80) oppose the growth of fibrils at low concentrations, making them suitable for engineering template-inducible fibril formation.*

A modified version of this chapter is submitted as:
L. Willems, S. Roberts, I. Weitzhandler, A. Chilkoti, E. Mastrobattista, J. van der Oost and R.J. de Vries, *Inducible fibril formation of silk-elastin diblocks*.

Introduction

Silk is a naturally occurring biomaterial with a unique combination of elasticity and strength approaching the mechanical properties of materials like Kevlar[1,2]. For millennia, humans have repurposed naturally occurring silk in textiles, and the unique properties of this material have led to its widespread use as a biomaterial[3]. Silk films have been used as wound covers to protect against infections[4], silk hydrogels and sponges, and porous microtubes have been used as scaffolds for the delivery of cells and/or cytokines[5–7] and for tissue regeneration[8], and silk capsules have been proven useful for drug delivery[9,10].

Applications of silk proteins and engineered silk-like polymers are to a large extent determined by their unique self-assembly behaviour which is controlled by an alternation of crystallisable and amorphous sequence motifs. In early studies on the structure-property relationship of silk-like polymers, Tirrell and coworkers synthesised and studied protein-based polymers with the general sequence $[(AG)_xEG]_n$ inspired by the crystallisable blocks of *Bombyx mori* silks. Next, mimicking the alternation of crystallisable and amorphous blocks of natural silks, various groups have developed block copolymers with both crystallisable and amorphous blocks. In particular, silk-elastin-like polypeptides (SELPs)[11,12] have been developed that are composed of multiple crystallisable silk-like blocks such as $(GAGAGS)_n$ and amorphous elastin-like $(GXGVP)_m$ blocks, where X is the so-called guest residue in the consensus motif for elastin-like polypeptides (ELPs). Another type of amorphous block that has been explored extensively for use in silk block copolymers[13–18] is a collagen-like polypeptide $(GXaaYaa)_m$ with the Xaa and Yaa residues chosen to be mostly uncharged and hydrophilic such that these blocks adopt random coils rather than triple helical configurations[19].

Silk-like polymers often assemble into fibrils and for many applications it would be advantageous if silk fibril formation could be engineered to be induced by specific templates. One approach is to locally concentrate the precursors on the template, and use silk-like polymers that only self-assemble into fibrils at high concentrations. This approach - that of driving fibril formation by a change in concentration - has been realised for the specific case of the virus-like encapsulation of single DNA molecules, using a silk-like polymer endowed with a K_{12} oligolysine DNA binding domain[20]. For the triblock copolymer $(GXaaYaa)_{132}$-$(GAGAGAGQ)_{10}$-K_{12} it was shown that in the absence of a DNA template it only forms fibrils at high concentrations (above 80 μM). However, already at much lower concentrations and when mixed with DNA, the triblock spontaneously forms rod-shaped virus-like particles. Each virus-like particle encapsulates a single DNA molecule, and has a core formed by a silk fibril.

A remaining outstanding question, and one that we attempt to answer herein for silk-elastin diblocks, is what the requirements are on the silk and elastin block lengths to arrive at silk polymers that only assemble into long fibrils at rather high concentrations. Normally concentrations are much lower, but fibril formation could be induced by an external template onto which the protein accumulates. With this in mind, we study the concentration-dependent fibril formation of silk-elastin-like diblocks, $(GSGVP)_m$-$(GAGAGAGQ)_n$ for a range of block lengths n and m.

Materials & Methods

Materials

pET24a(+) was previously modified for PRe-RDL (recursive directional ligation by plasmid reconstruction) cloning[21]. PRe-RDL vectors containing the genes encoding $E^S{}_m$ = $(GSGVP)_m$ (with m = 40, 60 and 80) were developed previously[22]. Custom oligonucleotides for the construction of the silk blocks $S^Q{}_n$ = $(GAGAGAGQ)_n$ were synthesised by Integrated DNA Technologies Inc. Restriction enzymes, calf-intestinal phosphatase (CIP) and Quick Ligation kit were ordered from New England Biolabs, and T4 DNA ligase buffer was purchased from Invitrogen. The DNA miniprep, gel purification and PCR purification kits were purchased from Qiagen Inc. Chemically competent *E. coli* cells (EB5Alpha and BL21 (DE3)) were purchased from EdgeBioSystems, and the TBdry growth media was purchased from MO BIO Laboratories Inc. The 4-20% Ready gel Tris-HCl precast gels and Precision Plus Protein Kaleidoscope Prestained Protein Standard were all purchased from BioRad.

Concatemerisation of silk-like repeat

To construct vectors carrying the silk-like repeat $S^Q{}_n$ (with n = 6, 10 or 14), two complementary custom oligonucleotides were designed. They encode two repeats of the silk-like sequence, $S^Q{}_2$ = $(GAGAGAGQ)_2$: 5'-cGCTGGTGCTGGTGCTGGTCAAGGAGCCGGTGCTGGAGCCGGCCAAgg-3' and 5'-TTGGCCGGCTCCAGCACCGGCTCCTTGACCAGCACCAGCACCAGCgcc-3'. The two oligonucleotides were 5' phosphorylated to allow efficient ligation into the vector. They were annealed at a concentration of 2 μM in T4 DNA ligase buffer by heating to 95°C for 2 min followed by slowly cooling the solution to room temperature over 3 hours. The pET24a(+) PRe-RDL cloning vector, was linearised by digesting approximately 2 μg of the vector with 5 U of *BseRI* for 2 hours at 37°C, followed by dephosphorylation of the 5' ends with 10 U of CIP for 30 min at 37°C. The linearised vector was purified using a PCR purification kit before ligation with the annealed oligonucleotides. Ligation and concatemerisation was conducted by incubation of an excess amount of the annealed oligonucleotides with the vector at room temperature for 5 min in the presence of Quick ligase and 1x Quick ligase buffer. Due to the design of the oligonucleotides, multiple copies of the $S^Q{}_2$ annealed oligonucleotides can sequentially ligate into the vector, thereby creating sequences that encode amongst others $S^Q{}_6$, $S^Q{}_{10}$ and $S^Q{}_{14}$. Finally, the ligation products were transformed into EB5Alpha *E. coli* chemically competent cells and the cells were plated on TBdry-Agar plates supplemented with 45 μg/mL kanamycin. Colony PCR was performed to select for those colonies that contained plasmids with one or more copies of the sequence encoding the silk-like repeat $S^Q{}_2$ and identity of the plasmids was confirmed by DNA sequencing.

Gene construction by PRe-RDL

To construct genes encoding for a set of $E^S{}_m$-$S^Q{}_n$ diblock proteins with various lengths of the ELP and silk-like blocks, we used the PRe-RDL approach previously developed[21]. The pET24a(+) PRe-RDL cloning vectors containing genes for various numbers of the ELP repeat ($E^S{}_m$, where m is 40, 60 or 80) were previously developed[22]. They were digested for 3 hours at 37°C using *BglI* and *AcuI* enzymes. *BglI* x *AcuI* fragments containing the $E^S{}_m$ genes, the so-called A-fragments, were purified using gel electrophoresis by separating the digestion products on a 1% agarose gel. To obtain the B-fragments, vectors containing the genes for a number of repeats of the silk-like sequence ($S^Q{}_n$, where n is 6, 10 or 14) were digested for 3 hours at 37°C using *BglI* and *BseRI* enzymes and the fragments containing the genes for the $S^Q{}_n$ sequences were purified by gel purification. Genes for $E^S{}_m$-$S^Q{}_n$ diblocks were then constructed by ligation of the A- and B-fragments using Quick ligase for 5 min at room temperature in a 1x Quick ligase buffer. The ligation products were transformed into *E. coli* EB5Alpha chemically competent cells and the cells were plated on TBdry-Agar plates supplemented with 45 μg/mL kanamycin. All sequences were confirmed by DNA sequencing.

Protein expression

Between 100 and 150 ng of the plasmids encoding the $E^S{}_m$-$S^Q{}_n$ diblocks were transformed into BL21(DE3) *E. coli* chemically competent cells. These cells were used to inoculate starter cultures of 10 mL TB medium supplemented with 45 μg/mL kanamycin. Starter cultures were incubated at 37°C overnight on a shaker at 200 rpm and used to inoculate cultures of 2 L TB medium with 45 μg/mL kanamycin. Cells were grown at 37°C on a shaker at 200 rpm for a total of 24 hours. Protein expression was induced in these cells 8 hours after inoculation by supplying isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 1 mM to the medium.

Protein purification

After 24 hours, cells were centrifuged at 3000 x *g* for 10 min at 10°C and the pellet was resuspended in 25 mL of cold PBS. The cells were lysed on ice by 18 sonication cycles of 10 seconds with 40 seconds intervals (Sonicator 3000, Misonix). Inclusion bodies were pelleted by centrifugation at 29000 x *g* for 12 min at 4°C and the supernatant (soluble lysate), containing the protein, was collected. Next, nucleic acids were removed by mixing the soluble lysate with 4 mL of 10% polyethylenimine and centrifugation at 29000 x *g* for 12 min at 4°C. Proteins in the remaining supernatant (cleared lysate) were further purified using inverse transition cycling (ITC)[23,24]. During each round of ITC, proteins are subjected to the following four steps: (1) addition of 1.0 to 2.5 M of ammonium sulphate or sodium chloride to lower the transition temperature of the ELP-containing proteins. (2) Incubation at 37°C for 15 min to induce phase separation of the ELP-containing proteins, and precipitation of the ELP-containing proteins by centrifugation at 40°C for 12 min at

29000 x *g*. (3) Resuspension in cold PBS to solubilise the ELP-containing proteins. (4) Removal of any remaining insoluble matter by centrifugation at 4°C for 12 min at 29000 x *g*. $E^S{}_m$-$S^Q{}_n$ proteins were purified by multiple rounds of ITC, but it was found that this did not lead to the desired level of purity of the proteins. To further increase purity, we also used a "bake-out" procedure: samples were incubated at 95°C for 15 min to induce denaturation of any remaining cellular proteins. Next, the samples were cooled down on ice for 30 min to ensure that the ELP-containing proteins dissolved, and finally the remaining denatured cellular proteins were removed by centrifugation at 4°C for 12 min at 29000 x *g*. A bake-out only was found to be insufficient for protein purification (as determined spectroscopically and with SDS-PAGE), so that this procedure was always combined with several rounds or ITC. For the $E^S{}_m$-$S^Q{}_{10}$ proteins, three rounds of ITC were used, followed by a bake-out. For the $E^S{}_m$-$S^Q{}_6$ proteins, three rounds of ITC were used with a bake-out after the 1st and 2nd rounds of ITC. Finally, for the $E^S{}_m$-$S^Q{}_{14}$ proteins, bake-out was integrated in the ITC: after centrifugation of aggregated ELP-containing proteins (ITC step 2), the pellet was resuspended and solubilised in PBS by incubation for 15 min at 95°C (step 3). To remove denatured proteins and other insoluble matter, the sample was cooled down on ice for 30 min and centrifuged at 4°C for 12 min at 29000 x *g* (step 4). Two rounds of ITC including bake-out were performed to purify the $E^S{}_m$-$S^Q{}_{14}$ proteins. Finally, all proteins were dialysed against Milli-Q water, lyophilised and stored at room temperature.

Protein characterisation

Protein purity was assessed with SDS-PAGE and matrix-assisted laser desorption ionisation time-of-flight mass spectrometry (MALDI-TOF MS). SDS-PAGE was carried out using 4-20% Ready gel Tris-HCl precast gels, 1x Laemmli running buffer and Precision Plus Protein Kaleidoscope Prestained Protein Standard. Gels were stained with PageBlue. MALDI-TOF MS was carried out on an UltrafleXtreme mass spectrometer (Bruker). Weighed portions of freeze-dried proteins were mixed with the matrix to a final concentration of 0.3 mg/mL and dried on the 800 μm spots of an MTP AnchorChip 384 target (Bruker). After mixing, final concentrations of the matrix components were 5 mg/mL 2.5-dihydroxyacetophenone, 1.5 mg/mL di-ammonium hydrogen citrate, 25% (v/v) ethanol and 1% (v/v) trifluoroacetic acid.

Protein self-assembly

For Atomic Force Microscopy experiments, protein stock solutions of 100 μM were prepared by vortexing weighed portions of freeze-dried proteins in 10 mM phosphate buffer (pH 7.4) at 95°C for 10 min. Solutions at lower concentrations were prepared by immediately diluting the 100 μM protein solutions in 10 mM phosphate buffer (pH 7.4). Next, samples were incubated at room temperature for 24 hours to allow self-assembly of the silk-elastin fibrils. For cryo-TEM, weighed portions of freeze-dried $E^S{}_{40}$-$S^Q{}_{10}$ and $E^S{}_{80}$-$S^Q{}_{10}$ proteins were solubilised in PBS to a concentration of 100 μM by a combination of vortexing and heating to 65°C

for 15 min. Solubilised proteins were incubated for 16 hours at room temperature to allow for the assembly of the silk-elastin fibrils.

Atomic Force Microscopy (AFM)

Samples for AFM imaging were prepared by depositing 5 μL of 24 hours incubated protein solutions onto clean silicon surfaces. Salts and non-absorbed particles were removed after 2 min by rinsing the surfaces with 1 mL of Milli-Q water. Surfaces were slowly dried under a N_2 stream. Samples were analysed in air in the ScanAsyst (PeakForce Tapping) imaging mode on a NanoScope MultiMode 8 system (Bruker) and using ScanAsyst-Air cantilevers (Bruker). The scan area was set to 5 x 5 μm and images were recorded at 512 x 512 pixels and a line rate of 0.977 Hz. A second-order flattening was performed on all images using NanoScope Analysis 1.40 software. Lengths of the self-assembled silk-elastin fibrils were measured using FiberApp software[25] with settings: alpha = 0, beta = 500, gamma = 20, kappa1 = 20, kappa2 = 10, step = 1 pixel, iterations = 100, "Use A* pathfinding algorithm". Lengths obtained with FiberApp software were binned to obtain length distributions, and to calculate the number and weight averages of lengths of the self-assembled silk-elastin fibrils (assuming fibril mass is proportional to fibril length). Standard error of the mean for the weight averages were estimated using the method of block averages: fibril length observations were split into 5 blocks, and the standard error of the mean was taken to be the standard deviation of the weight averaged fibril lengths for the 5 blocks.

Cryogenic Transmission Electron Microscopy (cryo-TEM)

Solutions containing 100 μM of $E^S{}_{40}$-$S^Q{}_{10}$ or $E^S{}_{80}$-$S^Q{}_{10}$, that were incubated for 16 hours at room temperature, were prepared for cryo-TEM using a FEI Vitrobot Mark III and imaged on a FEI Tecnai G2 Twin.

Results

Protein production

To study the effect of block lengths on the self-assembly behaviour of silk-elastin diblocks, genes for a set of diblock proteins $(GSGVP)_m$-$(GAGAGAGQ)_n$ were prepared using the PRe-RDL strategy as described previously[21]. Herein, we use the nomenclature E^S = GSGVP for the elastin pentamer motif, and S^Q = GAGAGAGQ for the silk motif; diblock sequences are abbreviated as $E^S{}_m$-$S^Q{}_n$. Genes were prepared for diblocks with m = 40, 60 or 80 $E^S{}_m$ pentamer repeats and n = 6, 10 or 14 $S^Q{}_n$ silk repeats. The final PRe-RDL step of combining genes for silk and elastin blocks to form genes for diblocks is illustrated in Figure 2.1, together with a depiction of the protein architecture. We depict the silk block in Figure 2.1b as a β-solenoid configuration: this is the configuration that the silk blocks adopt in fibrils according to recent computer simulations[26]. Fibrils are formed by β-solenoids

stacking on top of each other. Stacking is thought to be driven by the formation of multiple hydrogen bonds[26].

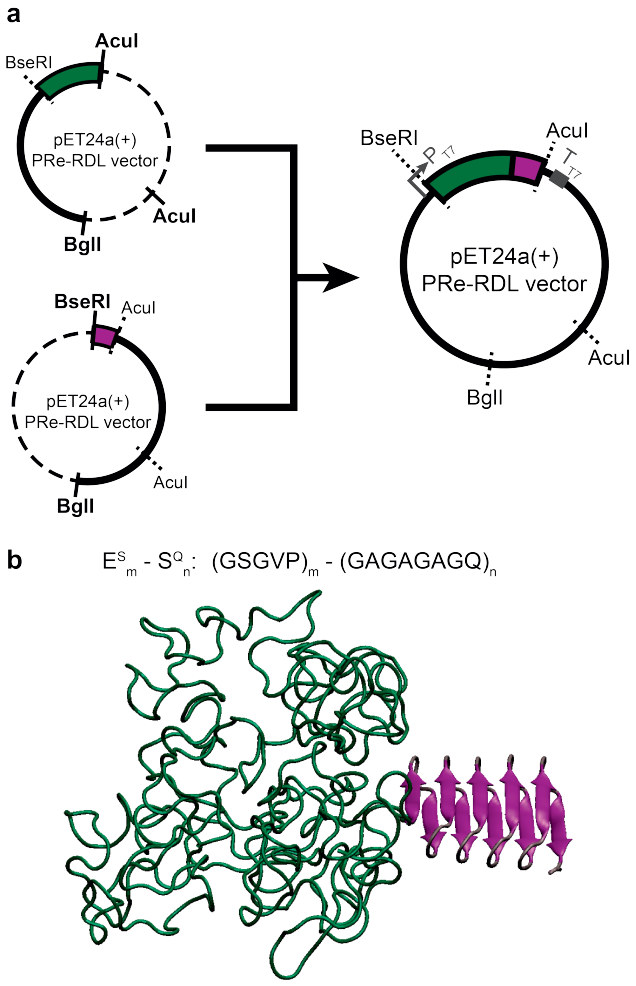


Figure 2.1 – PRe-RDL cloning of silk-elastin diblock proteins. a) PRe-RDL plasmids harbouring the genes for ELP ($E^S{}_m$; green) or silk ($S^Q{}_n$; purple) are digested using *BglI* x *AcuI* or *BglI* x *BseRI* respectively. Fragments containing the desired genes are ligated to recover the PRe-RDL plasmid containing the $E^S{}_m$-$S^Q{}_n$ genes. b) Full amino-acid sequence and depiction of the protein architecture of the $E^S{}_m$-$S^Q{}_n$ diblock with $m = 80$ and $n = 10$.

Representative SDS-PAGE for the $E^S{}_m$-$S^Q{}_n$ diblocks after purification is shown in Figure 2.2a. Note that silk-elastin diblocks stain rather poorly by the coomassie-based PageBlue staining. Also, as observed for many non-globular protein-based polymers with poor SDS binding[19,20,27–29], the diblocks have anomalous electrophoretic mobility as compared with the typically used molecular weight markers, such that we cannot directly estimate molar masses on the basis of the SDS-PAGE by comparing with molecular weight markers. Precise molar masses were obtained from MALDI-TOF MS. An example spectrum for the selected diblocks is shown in Figure 2.2b, and experimental molar masses for all diblocks are given in Table 2.1. Assuming full removal of N-terminal methionines[30], experimentally determined masses agree with the expected ones to within the experimental error.

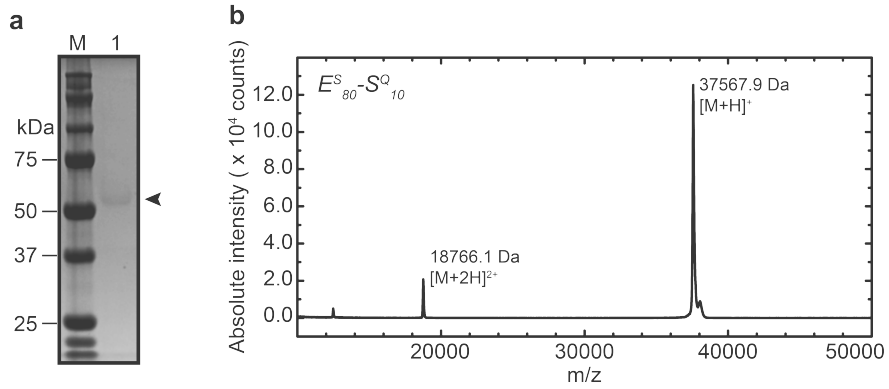


Figure 2.2 – Representative results for purification and characterisation of $E^S{}_m$-$S^Q{}_n$ diblocks. a) PageBlue-stained SDS-PAGE gel of the supernatant after three rounds of ITC and a bake-out containing pure $E^S{}_{80}$-$S^Q{}_{10}$ protein. Arrowhead indicates the protein and lane M shows the molecular weight marker. b) MALDI-TOF spectrum of purified $E^S{}_{80}$-$S^Q{}_{10}$, which is representative for the spectra of the other $E^S{}_m$-$S^Q{}_n$ proteins.

Table 2.1 – Expected and experimentally determined molar masses (MALDI-TOF) of $E^S{}_m$-$S^Q{}_n$ diblocks.

	Expected MW (Da)	Experimental MW(Da)
$E^S{}_{40}$-$S^Q{}_6$	19389.77	19386.4
$E^S{}_{60}$-$S^Q{}_6$	27338.39	27335.8
$E^S{}_{80}$-$S^Q{}_6$	35287.02	35290.9
$E^S{}_{40}$-$S^Q{}_{10}$	21668.07	21660.4
$E^S{}_{60}$-$S^Q{}_{10}$	29616.69	29620.0
$E^S{}_{80}$-$S^Q{}_{10}$	37565.32	37567.9
$E^S{}_{40}$-$S^Q{}_{14}$	23946.36	23944.6
$E^S{}_{60}$-$S^Q{}_{14}$	31894.99	31900.7
$E^S{}_{80}$-$S^Q{}_{14}$	39843.62	39850.3

Dependence of self-assembly of silk-elastin diblocks on block lengths

We used atomic force microscopy (AFM) and cryo-TEM imaging to explore the dependence of the self-assembly of silk-elastin diblocks on block length. Solutions of $E^S{}_m$-$S^Q{}_n$ diblocks with $m = 40$ or 80 and $n = 6$, 10 or 14 were prepared from purified protein powders and incubated at room temperature for 24 h (for AFM) or for 16 h (for cryo-TEM) to allow for their self-assembly into fibrils. In our previous studies on silk fibril templating on DNA by the $(GXaaYaa)_{132}$-$(GAGAGAGQ)_{10}$-K_{12} triblock copolymer we found that templated fibril assembly already occurred at a low concentration of 1.8 μM, but in the absence of a template, fibrils were only found at concentrations above 80 μM[20]. Here we study the assembly of the silk-elastin diblocks at the same two concentrations. Results for AFM and cryo-TEM imaging are shown in Figure 2.3.

We first examined the dependence of self-assembly on the length of the silk block with the elastin block fixed at $m = 80$. We find that a silk block length of $n = 6$ is not enough to allow for fibril formation at either the low or high concentration that we have used in our experiments (see Figure 2.3a,e). For silk block lengths of $n = 10$ and 14, we find concentration dependent fibril formation: at a protein concentration of 1.8 μM both diblocks self-assemble into relatively short fibrils (Figure 2.3b,c), whereas at a high concentration of 100 μM they self-assemble into very long fibrils (Figure 2.3f,g).

Next, we examine the influence of the length m of the amorphous elastin blocks on the self-assembly of the $E^S{}_m$-$S^Q{}_n$ silk-elastin diblocks. At high concentrations, all fibrils for the diblocks with $n = 10$ or 14 are very long irrespective of the length m of the amorphous block (Figure 2.3f,g,h), therefore we focus on the case of low concentrations, comparing the $n = 10$ diblocks for lengths m of the amorphous blocks of $m = 40$ (Figure 2.3b) and $m = 80$ (Figure 2.3d). From the images, it appears that the fibrils become shorter as the length of the amorphous block increases from $m = 40$ to $m = 80$, ascertaining this requires a more quantitative analysis that we perform below. To exclude the possibility that the AFM sample preparation somehow affects our observations on fibril growth, we also used cryo-TEM to visualise fibrils in solution, at a protein concentration of 100 μM. Representative micrographs for a silk block length of $n = 10$, and amorphous block lengths of $m = 40$ or 80 are shown in Figure 2.3i,j. As observed using AFM for the same conditions, long fibrils are formed in both cases (indicated by white arrowheads).

Quantitative Analysis of fibril lengths

To quantitatively characterise the subtle differences in the assembly of the diblocks into fibrils at the low concentrations of 1.8 μM, for the diblocks with $n = 10$ or 14 and $m = 40$, 60 or 80, lengths of fibrils were determined from AFM images. Histograms of the fibril lengths for $E^S{}_{80}$-$S^Q{}_{10}$, $E^S{}_{80}$-$S^Q{}_{14}$ and $E^S{}_{40}$-$S^Q{}_{10}$ are shown in Figure 2.4. The distributions are very broad, with many short fibrils and a smaller number of longer fibrils. Despite their lower number, the longer fibrils represent a significant fraction of the total protein mass. Therefore the length distributions

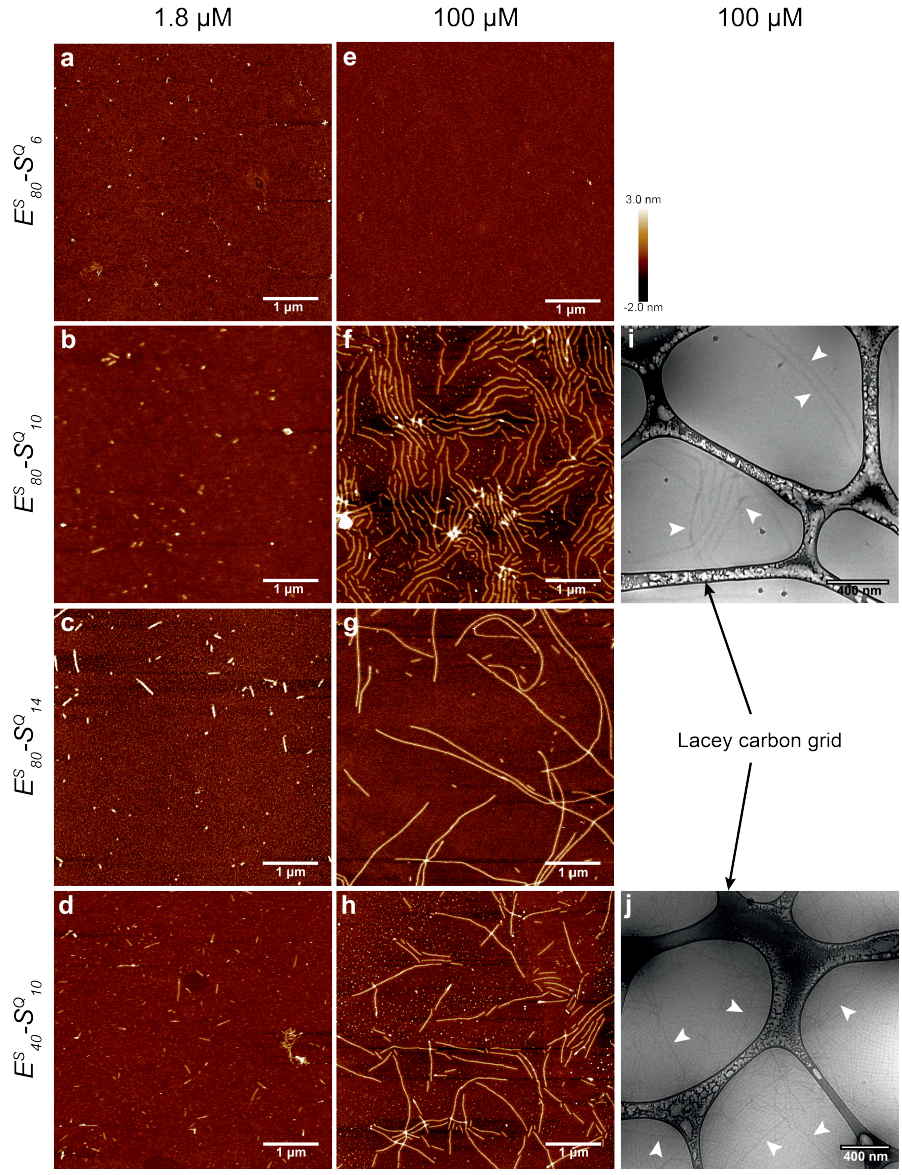


Figure 2.3 – Dependence of the self-assembly of E^S_m-S^Q_n diblocks on silk and elastin block lengths. For AFM, silk-elastin diblocks were incubated for 24h, subsequently adsorbed on silicon substrates and imaged under dry conditions. Protein solutions were prepared at a low concentration of 1.8 μM: a) E^S_{80}-S^Q_6, b) E^S_{80}-S^Q_{10}, c) E^S_{80}-S^Q_{14}, and d) E^S_{40}-S^Q_{10}; or at a high concentration of 100 μM: e) E^S_{80}-S^Q_6, f) E^S_{80}-S^Q_{10}, g) E^S_{80}-S^Q_{14}, and h) E^S_{40}-S^Q_{10}. For cryo-TEM, diblocks were prepared at 100 μM and incubated for 16h: i) E^S_{80}-S^Q_{10}, j) E^S_{40}-S^Q_{10}. Black arrows indicate the lacey carbon grid, silk-based block copolymer fibrils are indicated by white arrowheads.

are given as the fraction of total protein mass (weight fraction) versus fibril length. Again, in view of the broad distributions, we give both the number averaged fibril lengths $<L>_n$ and the weight averaged fibril lengths $<L>_w$ for the diblocks with $n = 10$ or 14 and $m = 40$, 60 or 80, in Table 2.2. Finally, Figure 2.5 shows the number averaged fibril lengths $<L>_n$ for all diblocks with silk block lengths $n = 10$ and $n = 14$.

First, we examined the effect of the silk block length n on fibril lengths by comparing the fibril length distributions for diblocks with a fixed length of the amorphous block ($m = 80$), for silk block lengths of $n = 10$ and $n = 14$ (Figure 2.4a,b). For $n = 14$, the weight fraction of protein in short fibrils (1-50 nm) is much lower than for $n = 10$, and the shape of the distribution is relatively flat, leading to weight averaged fibril lengths that are more than twice as large for $n = 14$ than for $n = 10$ ($<L>_w = 185\pm44.8$ nm for $n = 14$, as opposed to $<L>_w = 67\pm9.0$ nm for $n = 10$, see Table 2.2 and Figure 2.4). This shows that at the low concentration of 1.8 μM, when the diblocks form only short fibrils, the self-assembly is very sensitive to the precise length of the silk block.

Next, we examined the effect of the length m of the amorphous elastin block by comparing the fibril length distributions for diblocks with $n = 10$ silk blocks and respectively, $m = 40$ or $m = 80$ elastin-like blocks. The histograms of Figure 2.4 corroborate the qualitative observation that fibrils become shorter as the length of the amorphous block increases from $m = 40$ to $m = 80$: fibrils formed by $m = 80$ diblocks are significantly smaller than fibrils formed by the $m = 40$ diblocks at the same concentration. While for $m = 80$ the distribution peaks at fibril sizes of 1-50 nm, for $m = 40$ the distribution peaks at fibril sizes of 50-100 nm, and additionally the fraction of protein mass in long fibrils is also much larger for $m = 40$. By increasing the length of the elastin-like block from $m = 40$ to $m = 80$, the weight averaged length of the fibrils at 1.8 μM decreases from $<L>_w = 129\pm16.0$ nm to $<L>_w = 67\pm9.0$ nm (Table 2.2 and Figure 2.4). This shows that for $n = 10$, the elastin blocks are able to modulate assembly of the silk blocks in such a way that longer elastin blocks lead to shorter fibrils at low protein concentrations.

The influence of the length m of the amorphous block for a silk block length of $n = 10$ on the fibril lengths at low concentrations is even more evident from Figure 2.5, which show plots of the number averaged fibril lengths $<L>_n$ for $n = 10$ and 14 as a function of m (Figure 2.5). For the $n = 10$ diblocks, fibril lengths clearly decrease when the length m of the amorphous block increases, whereas for $n = 14$, there is no clear trend.

These data collectively indicate that a minimum length of the silk block is required for fibril formation which is between $n = 6$ and $n = 10$. Second, the ability of the amorphous elastin block to influence the fibril length, rapidly diminishes as the silk block length becomes larger, with a strong influence on fibril length for diblocks with a silk block length of $n = 10$, and with no clear influence on fibril length for $n = 14$.

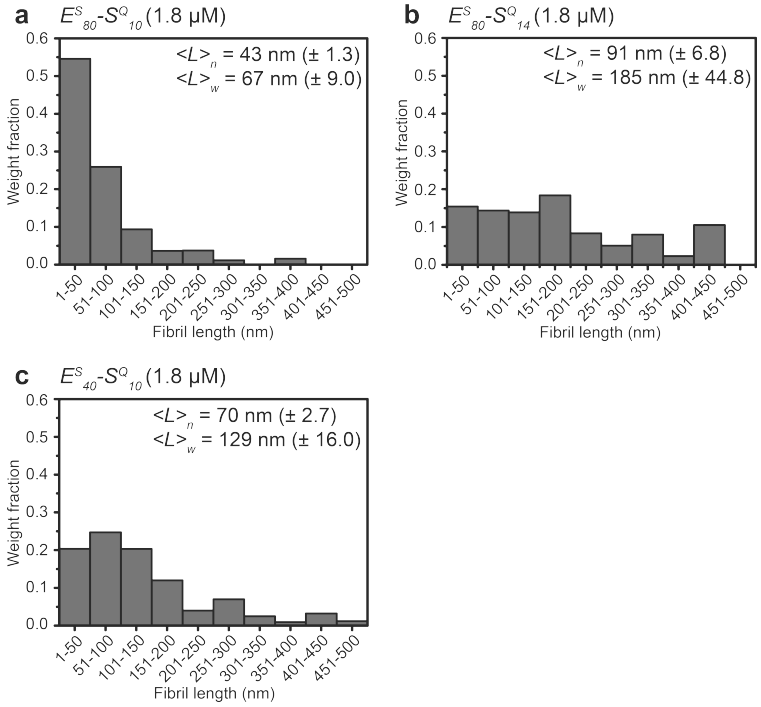


Figure 2.4 – Fibril length distributions of E^S_m-S^Q_n diblocks. Fibril lengths were determined from AFM images of diblocks incubated at 1.8 μM for 24h. Weight fractions of each binned length category were calculated with the assumption that the protein mass of a given fibril is proportional to the fibril length. a) E^S_{80}-S^Q_{10}, b) E^S_{80}-S^Q_{14}, and c) E^S_{40}-S^Q_{10}. For each protein sample, the number ($<L>_n$) and weight ($<L>_w$) averaged fibril lengths (± s.e.m.) are also shown in the plots.

Table 2.2 – Number and weight averaged fibril lengths for the E^S_m-S^Q_n diblocks.

	N	$<L>_n$ (nm)*	$<L>_w$ (nm)*
E^S_{40}-S^Q_{10}	567	70 (± 2.7)	129 (± 16.0)
E^S_{60}-S^Q_{10}	85	54 (± 3.7)	75 (± 13.8)
E^S_{80}-S^Q_{10}	636	43 (± 1.3)	67 (± 9.0)
E^S_{40}-S^Q_{14}	222	81 (± 5.6)	168 (± 41.6)
E^S_{60}-S^Q_{14}	159	57 (± 3.9)	100 (± 30.2)
E^S_{80}-S^Q_{14}	185	91 (± 6.8)	185 (± 44.8)

N is the number of fibrils analysed for each diblock. * Number and weight averaged fibril lengths (± s.e.m.) were calculated using the data obtained from AFM images (such as in Figure 2.3).

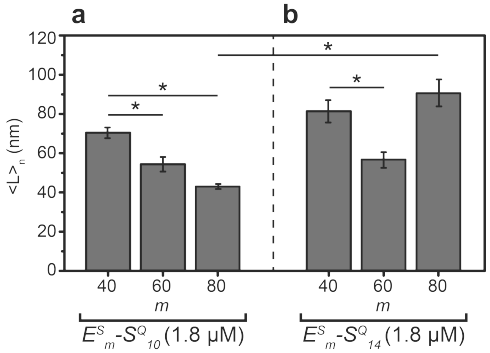


Figure 2.5 – Influence of the elastin block length m on fibril lengths of the E^S_m-S^Q_n diblocks. Number averaged fibril lengths at low protein concentration (1.8 μM) versus the block length m of the amorphous elastin block are plotted for a) E^S_m-S^Q_{10} and b) E^S_m-S^Q_{14}. The error bars represent the standard error of the mean (s.e.m.). * $p < 0.01$ as determined by a Welch's *t*-test for two independent samples with unequal variances.

Discussion

Block lengths and block length ratios are one of the key parameters of protein-based block copolymers governing their self-assembly, co-assembly with a template and other functionalities[11,20,31–33]. The synergy between ordered and amorphous protein domains has similarly been shown to heavily influence the resulting material properties of protein-based materials[34]. In the case of silk-like protein-based polymers, this means that the amorphous blocks influence the molecular interactions of the crystallisable blocks and vice-versa. The diblocks explored here are an excellent model system to study this mutual influence of amorphous and crystallisable blocks in these polymers.

For E^S_m-S^Q_n silk-elastin diblock polypeptides we have shown herein that a minimal length of the crystallisable silk block is required to drive fibril formation. We established that for amorphous block lengths $m \geq 40$, $n = 6$ is insufficient whereas $n = 10$ is sufficient for fibril formation. Next, we have shown that for E^S_m-S^Q_n silk-elastin diblock proteins the fibril length in some cases responds to the length m of the amorphous blocks, while in other cases it does not. More specifically, it appears that the diblocks with block lengths n that are closest to the boundary for fibril formation ($n = 10$) respond strongly to the length of the amorphous block, with longer amorphous blocks leading to shorter fibrils at low concentrations, suggesting that the amorphous blocks in that case oppose fibril growth. For longer silk blocks ($n = 14$), the driving force for fibril growth is apparently already too large to be opposed by the amorphous blocks, since in that case we do not find a significant dependence of fibril growth on the length of the amorphous blocks. Our report of continuous tuning of fibril growth by adjusting the length of the amorphous blocks is new, but also consistent with our previous report on "on/off" switching of fibril formation via amorphous blocks: triblocks $(GXaaYaa)_{67}$-$(GAGAGAGH)_n$-

$(GXaaYaa)_{67}$ with $n = 16$ do not form fibrils in the presence of the amorphous blocks, but when the latter are proteolytically shortened, fibrils are formed[27].

In summary, we examined the interplay between crystallisable and amorphous blocks in $E^S{}_m$-$S^Q{}_n$ silk-elastin diblock polypeptides. We find that for diblocks with lengths of crystallisable blocks just over the threshold for fibril formation, the amorphous blocks act as a break on fibril growth at low concentrations, with longer amorphous blocks leading to shorter fibrils. Since template-inducible silk fibril formation requires that fibril formation only occurs at high concentrations, we conclude that for engineering (template) inducible silk fibril formation of $E^S{}_m$-$S^Q{}_n$ silk-elastin diblock polypeptides, suitable block lengths are most likely those with crystallisable block lengths n as close to the threshold for fibril formation as possible. At such critical silk block lengths n, the amorphous block length m can be used for fine tuning assembly behaviour. In combination with suitable template binding blocks, such diblocks allow for nucleic acid encapsulation in silk fibrils[20], and, for example, may also allow for engineering induced silk fibril formation on the surfaces of biomaterials, which has also been explored for other silks[35–37].

References

[1] J. G. Hardy, L. M. Romer, and T. R. Scheibel, "Polymeric materials based on silk proteins," *Polymer*, vol. 49, no. 20, pp. 4309–4327, 2008.

[2] C. Vepari and D. L. Kaplan, "Silk as a biomaterial," *Progress in Polymer Science (Oxford)*, vol. 32, no. 8-9, pp. 991–1007, 2007.

[3] G. H. Altman, F. Diaz, C. Jakuba, T. Calabro, R. L. Horan, J. Chen, H. Lu, J. Richmond, and D. L. Kaplan, "Silk-based biomaterials," *Biomaterials*, vol. 24, no. 3, pp. 401–416, 2003.

[4] E. S. Gil, B. Panilaitis, E. Bellas, and D. L. Kaplan, "Functionalized Silk Biomaterials for Wound Healing," *Advanced Healthcare Materials*, vol. 2, no. 1, pp. 206–217, 2013.

[5] X. Ding, G. Yang, W. Zhang, G. Li, S. Lin, D. L. Kaplan, and X. Jiang, "Increased stem cells delivered using a silk gel/scaffold complex for enhanced bone regeneration," *Scientific Reports*, vol. 7, no. 1, pp. 1–10, 2017.

[6] X. Wang, J. A. Kluge, G. G. Leisk, and D. L. Kaplan, "Sonication-induced gelation of silk fibroin for cell encapsulation," *Biomaterials*, vol. 29, no. 8, pp. 1054–1064, 2008.

[7] W. Zhang, C. Zhu, D. Ye, L. Xu, X. Zhang, Q. Wu, X. Zhang, D. L. Kaplan, and X. Jiang, "Porous silk scaffolds for delivery of growth factors and stem cells to enhance bone regeneration," *PLoS ONE*, vol. 9, no. 7, pp. 1–9, 2014.

[8] V. E. Bosio, J. Brown, M. J. Rodriguez, and D. L. Kaplan, "Biodegradable porous silk microtubes for tissue vascularization," *J. Mater. Chem. B*, vol. 5, no. 6, pp. 1227–1235, 2017.

[9] X. Wang, T. Yucel, Q. Lu, X. Hu, and D. L. Kaplan, "Silk nanospheres and microspheres from silk/pva blend films for drug delivery," *Biomaterials*, vol. 31, no. 6, pp. 1025–1035, 2010.

[10] A. S. Lammel, X. Hu, S. H. Park, D. L. Kaplan, and T. R. Scheibel, "Controlling silk fibroin particle features for drug delivery," *Biomaterials*, vol. 31, no. 16, pp. 4583–4591, 2010.

[11] X.-x. Xia, Q. Xu, X. Hu, G. Qin, and D. L. Kaplan, "Tunable Self-Assembly of Genetically Engineered Silk Elastin-like Protein Polymers," 2011.

[12] W. Teng, J. Cappello, and X. Wu, "Recombinant silk-elastinlike protein polymer displays elasticity comparable to elastin," *Biomacromolecules*, vol. 10, no. 11, pp. 3028–3036, 2009.

[13] L. H. Beun, X. J. Beaudoux, J. M. Kleijn, F. a. De Wolf, and M. a. Cohen Stuart, "Self-assembly of silk-collagen-like triblock copolymers resembles a supramolecular living polymerization," *ACS Nano*, vol. 6, no. 1, pp. 133–140, 2012.

[14] L. H. Beun, L. Albertazzi, D. Van Der Zwaag, R. De Vries, and M. A. Cohen Stuart, "Unidirectional Living Growth of Self-Assembled Protein Nanofibrils Revealed by Super-resolution Microscopy," *ACS Nano*, vol. 10, no. 5, pp. 4973–4980, 2016.

[15] M. D. Golinska, M. K. Włodarczyk-Biegun, M. W. Werten, M. A. Stuart, F. A. De Wolf, and R. De Vries, "Dilute self-healing hydrogels of silk-collagen-like block copolypeptides at neutral pH," *Biomacromolecules*, vol. 15, no. 3, pp. 699–706, 2014.

[16] M. K. Włodarczyk-Biegun, M. W. Werten, F. A. De Wolf, J. J. Van Den Beucken, S. C. Leeuwenburgh, M. Kamperman, and M. A. Cohen Stuart, "Genetically engineered silk-collagen-like copolymer for biomedical applications: Production, characterization and evaluation of cellular response," *Acta Biomaterialia*, vol. 10, no. 8, pp. 3620–3629, 2014.

[17] M. K. Włodarczyk-Biegun, K. Farbod, M. W. Werten, C. J. Slingerland, F. A. De Wolf, J. J. Van Den Beucken, S. C. Leeuwenburgh, M. A. Cohen Stuart, and M. Kamperman, "Fibrous hydrogels for cell encapsulation: A modular and supramolecular approach," *PLoS ONE*, vol. 11, no. 5, 2016.

[18] N. E. Domeradzka, M. W. Werten, F. A. De Wolf, and R. De Vries, "Cross-Linking and Bundling of Self-Assembled Protein-Based Polymer Fibrils via Heterodimeric Coiled Coils," *Biomacromolecules*, vol. 17, no. 12, pp. 3893–3901, 2016.

[19] M. W. T. Werten, W. H. Wisselink, T. J. Jansen-van den Bosch, E. C. de Bruin, and F. a. de Wolf, "Secreted production of a custom-designed, highly hydrophilic gelatin in Pichia pastoris," *Protein Engineering Design and Selection*, vol. 14, pp. 447–454, jun 2001.

[20] A. Hernandez-Garcia, D. J. Kraft, A. F. J. Janssen, P. H. H. Bomans, N. A. J. M. Sommerdijk, D. M. E. Thies-Weesie, M. E. Favretto, R. Brock, F. a. de Wolf, M. W. T. Werten, P. van der Schoot, M. C. Stuart, and R. de Vries, "Design and self-assembly of simple coat proteins for artificial viruses," *Nature Nanotechnology*, vol. 9, pp. 698–702, aug 2014.

[21] J. R. McDaniel, D. J. Callahan, and A. Chilkoti, "Drug delivery to solid tumors by elastin-like polypeptides.," *Advanced drug delivery reviews*, vol. 62, pp. 1456–1467, dec 2010.

[22] S. R. MacEwan, I. Weitzhandler, I. Hoffmann, J. Genzer, M. Gradzielski, and A. Chilkoti, "Phase Behavior and Self-Assembly of Perfectly Sequence-Defined and Monodisperse Multiblock Copolypeptides," *Biomacromolecules*, vol. 18, no. 2, pp. 599–609, 2017.

[23] W. Hassouneh, T. Christensen, and A. Chilkoti, "Elastin-like polypeptides as a purification tag for recombinant proteins," in *Current Protocols in Protein Science*, pp. 6.11.1–6.11.16, aug 2010.

[24] D. E. Meyer and a. Chilkoti, "Purification of recombinant proteins by fusion with thermally-responsive polypeptides.," *Nature biotechnology*, vol. 17, no. 11, pp. 1112–1115, 1999.

[25] I. Usov and R. Mezzenga, "FiberApp: An Open-Source Software for Tracking and Analyzing Polymers, Filaments, Biomacromolecules, and Fibrous Objects," *Macromolecules*, p. 150210103859007, 2015.

[26] B. Zhao, M. A. Cohen Stuart, and C. K. Hall, "Dock n roll: folding of a silk-inspired polypeptide into an amyloid-like beta solenoid," *Soft Matter*, 2016.

[27] L. H. Beun, I. M. Storm, M. W. T. Werten, F. A. De Wolf, M. A. Cohen Stuart, and R. De Vries, "From micelles to fibers: Balancing self-assembling and random coiling domains in pH-responsive silk-collagen-like protein-based polymers," *Biomacromolecules*, vol. 15, no. 9, pp. 3349–3357, 2014.

[28] M. T. Krejchi, E. D. T. Atkins, A. J. Waddon, M. J. Fournier, T. L. Mason, and D. A. Tirrell, "Chemical sequence control of B-sheet assembly in macromolecular crystals of periodic polypeptides," *Science (New York, N.Y.)*, vol. 265, pp. 1427–1432, 1994.

[29] D. E. Meyer and A. Chilkoti, "Protein Purification by Inverse Transition Cycling," *ProteinProtein Interactions: AMolecular Cloning Manual*, vol. Chapter 18, pp. 329–344, 2002.

[30] P. H. Hirel, M. J. Schmitter, P. Dessen, G. Fayat, and S. Blanquet, "Extent of N-terminal methionine excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 86, no. 21, pp. 8247–51, 1989.

[31] O. S. Rabotyagova, P. Cebe, and D. L. Kaplan, "Protein-based block copolymers," *Biomacromolecules*, vol. 12, pp. 269–289, feb 2011.

[32] S. Lin, S. Ryu, O. Tokareva, G. Gronau, M. M. Jacobsen, W. Huang, D. J. Rizzo, D. Li, C. Staii, N. M. Pugno, J. Y. Wong, D. L. Kaplan, and M. J. Buehler, "Predictive modelling-based design and experiments for synthesis and spinning of bioinspired silk fibres," *Nature Communications*, vol. 6, no. May, pp. 1–12, 2015.

[33] Y. B. Lim, E. Lee, Y. R. Yoon, M. S. Lee, and M. Lee, "Filamentous artificial virus from a self-assembled discrete nanoribbon," *Angewandte Chemie - International Edition*, vol. 47, no. Cd, pp. 4525–4528, 2008.

[34] S. Roberts, T. S. Harmon, J. Schaal, V. Miao, K. Li, A. Hunt, Y. Wen, T. G. Oas, J. H. Collier, R. V. Pappu, and A. Chilkoti, "Injectable tissue integrating networks from recombinant polypeptides with tunable order," *Nature Materials*, 2018.

[35] C. Charbonneau, J. M. Kleijn, and M. A. Cohen Stuart, "Subtle charge balance controls surface-nucleated self-assembly of designed biopolymers," *ACS Nano*, vol. 8, no. 3, pp. 2328–2335, 2014.

[36] T. T. Pham, W. H. Rombouts, R. Fokkink, M. C. Stuart, M. A. Cohen Stuart, and J. M. Kleijn, "Nanoparticle-Templated Formation and Growth Mechanism of Curved Protein Polymer Fibrils," *Biomacromolecules*, vol. 17, no. 7, pp. 2392–2398, 2016.

[37] W. Hwang, B. H. Kim, R. Dandu, J. Cappello, H. Ghandehari, and J. Seog, "Surface induced nanofiber growth by self-assembly of a silk-elastin-like protein polymer," *Langmuir*, vol. 25, no. 21, pp. 12682–12686, 2009.

3

The nature of the amorphous hydrophilic block affects self-assembly of an artificial viral capsid polypeptide

Consensus motifs for sequences of both crystallisable and amorphous blocks in (structural analogues of) silks vary widely. For designing novel silk-like polypeptide materials, an important question is therefore how the nature of these two blocks impacts the self-assembly and the resulting physical properties of the polypeptides. We here address the influence of the amorphous block on the self-assembly of a silk-like polypeptide that was previously designed to encapsulate single DNA molecules into rod-shaped virus-like particles (VLPs). The polypeptide has a triblock architecture, with a long N-terminal amorphous block, a crystallisable midblock, and a C-terminal DNA binding block. We compare a triblock with a very hydrophilic collagen-like amorphous block $(GXaaYaa)_{132}$ to that of a triblock with a slightly less hydrophilic elastin-like amorphous block $(GSGVP)_{80}$. The amorphous blocks have similar lengths and both adopt a random coil structure in solution, but atomic force microscopy revealed significant differences in their self-assembly behaviour, with polypeptides with elastin-like blocks forming much longer polypeptide-only fibrils than polypeptides with collagen-like blocks. We attribute the difference to the more hydrophilic nature of the collagen-like block, which more strongly opposes the growth of polypeptide-only fibrils than the elastin-like blocks. Our work illustrates that rather subtle differences in the chemical nature of amorphous blocks can strongly influence the self-assembly and hence the functionality of engineered silk-like polypeptides.

A modified version of this chapter is submitted as:
L. Willems, L. van Westerveld, S. Roberts, I. Weitzhandler, C. Calcines Cruz, A. Hernandez-Garcia, A. Chilkoti, E. Mastrobattista, J. van der Oost and R.J. de Vries, *The nature of the amorphous hydrophilic block affects self-assembly of an artificial viral coat polypeptide*.

Introduction

Self-assembly of proteins underlies much of the structural integrity of biological systems and it is increasingly exploited for creating, among others, scaffolds for cell growth, nanoparticles for drug delivery, biosensors and nanoreactors[1,2]. A relatively straightforward approach to *de novo* design novel types of self-assembling structural proteins is the use of consensus motifs found in natural structural proteins such as elastin, silk and collagen. These motifs can be repeated and combined in many ways to design sequences for novel self-assembling polypeptides for a wide range of applications[1,3].

An important class of self-assembling polypeptides are silk-like polypeptides, which typically consist of both crystallisable and amorphous blocks. Consensus motifs for sequences of both crystallisable and amorphous blocks in natural silks and structural silk analogues, such as spider silks[4], silkworm silks[5] and squid sucker ring teeth[6], vary widely. An obvious question is therefore how the nature of the crystallisable and amorphous blocks affects the self-assembly and resulting physical properties of these materials. The same question also arises for *de novo* design of silk-like polypeptides, but so far this issue has not yet been addressed in the literature.

We here consider the influence of the nature of the amorphous block on self-assembly for a triblock polypeptide designed previously to encapsulate single DNA molecules into rod-shaped virus like particles (VLPs)[7]. The triblock polypeptide C-$S^Q{}_{10}$-K_{12} features a dodeca-lysine block (K_{12}) for binding to the nucleic acid template, a silk-like midblock $S^Q{}_{10}$ = $(\mathrm{GAGAGAGQ})_{10}$ that self-assembles into the rod-shaped core of the VLP via stacking of β-rolls[8–11], and a collagen-like amorphous block $C = (\mathrm{GXaaYaa})_{132}$ for colloidal stabilisation of the VLP particles. Note that the Xaa and Yaa residues in the collagen-like C block are chosen to be highly hydrophilic and mostly uncharged such that the resulting polypeptide adopts a random coil conformation rather than a triple helix[12].

Apart from a general understanding of the influence of the nature of amorphous blocks on the properties of silk polymers, we also have more specific reasons for exploring other kinds of amorphous blocks for the C-$S^Q{}_{10}$-K_{12} triblock copolymer. First of all, the nature of the amorphous block strongly impacts on the production of the polymer in host organisms: while the C-$S^Q{}_{10}$-K_{12} polymer can be obtained at high yield by secreted expression in the yeast *Pichia pastoris*[7], expression in *Escherichia coli* is problematic (personal communication with A. Hernandez-Garcia). Next, with respect to applications, since the amorphous blocks are on the outside of the VLPs, it is the amorphous block that will, to a large extent, determine the biological response of cells and tissues to the VLPs. Therefore, it is of interest to explore amorphous blocks that have been very well characterised with respect to their biological response, such as elastin-like polymers. Elastin-like polypeptides (ELPs), with the general sequence $E^X{}_n = (\mathrm{GXGVP})_n$ are known to be expressed at high yield in *E. coli*[13], but for *P. pastoris* expression levels are much lower than for the collagen-like C blocks[12], especially for ELPs harbouring a hydrophobic guest residue X[14].

With this in mind, we produced a triblock polypeptide $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} in *E. coli*, and compared its co-assembly with DNA with that of the *C*-$S^Q{}_{10}$-K_{12} polypeptide, produced in *P. pastoris*. To compare with the collagen-like *C* block, we used an elastin-like polypeptide with a length of 80 pentamers to match the length of the approximately 400 amino acid-long *C* block. As a guest residue in the elastin-like polypeptide we used serine (X = S) as this residue is uncharged and hydrophilic.

Materials & Methods

Materials

The *C*-$S^Q{}_{10}$-K_{12} polypeptide was produced and purified as described previously[7]. A pET24a(+) PRe-RDL cloning vector and a cloning vector encoding the $E^S{}_{80}$-$S^Q{}_{10}$ gene were constructed as previously described in ref [15] and Chapter 2. Custom oligonucleotides were synthesised by Integrated DNA Technologies Inc. All restriction enzymes, the calf-intestinal phosphatase (CIP) and the Quick Ligation kit were purchased from New England Biolabs, and the T4 DNA ligase buffer from Invitrogen. DNA miniprep, gel purification and PCR purification kits were from Qiagen Inc. EB5Alpha and BL21 (DE3) chemically competent *E. coli* cells were ordered from EdgeBioSystems. The NoLimits 2000 bp linear dsDNA, PageBlue protein staining solution, SYBR Safe DNA stain, 6x DNA loading dye, GeneRuler 1kb DNA ladder and 50x TAE buffer were all purchased from Thermo Fisher Scientific. The 10% mini-protean TGX precast protein gels and Precision Plus Protein All Blue Prestained Protein Standard were ordered from BioRad, and agarose was bought from Brunschwig Chemie.

Preparation of oligolysine-encoding cloning vector

Two complementary custom oligonucleotides encoding the oligolysine K_{12} were designed: 5'-cAAGAAAAAGAAGAAAAAGAAGAAGAAGAAAAAGAAGgg-3' and 5'-CTTCTTTTTCTTCTTCTTCTTTTTCTTCTTTTTCTTgcc-3'. To anneal the two oligonucleotides, they were diluted in T4 DNA ligase buffer to a concentration of 2 μM, heated to 95°C for 2 min and then slowly cooled to room temperature over 3 hours. Linear pET24a(+) PRe-RDL cloning vector was prepared by the digestion of approximately 2 μg vector for 2 hours at 37°C with 5 U of *BseRI*, the 5' ends of the linearised vector was dephosphorylated using 10 U of CIP for 30 min at 37°C. Next, a PCR purification kit was used to remove the enzymes, and the annealed oligonucleotides were ligated into the linearised vector by incubating a mixture of both in 1x Quick ligase buffer for 5 min at room temperature in presence of Quick ligase. The ligation product was then transformed into EB5Alpha chemically competent cells and the cells were plated on TBdry plates supplemented with 45 μg/mL kanamycin. Colonies containing the plasmid with a copy of the oligolysine K_{12} were selected by colony PCR and confirmed by DNA sequencing.

Gene construction by PRe-RDL

Recursive directional ligation by plasmid reconstruction (PRe-RDL)[15] was used to construct a plasmid encoding for the $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} polypeptide. To obtain the so-called A-fragment the cloning vector coding for $E^S{}_{80}$-$S^Q{}_{10}$ was digested with *BglI* and *AcuI* enzymes for 3 hours at 37°C and the *BglI* x *AcuI* fragment containing the $E^S{}_{80}$-$S^Q{}_{10}$ gene was purified by gel purification. The B-fragment was obtained by digesting the cloning vector coding for K_{12} with *BglI* and *BseRI* for 3 hours at 37°C, followed by gel purification of the *BglI* x *BseRI* fragment containing the K_{12} gene. Ligation of the A- and B-fragments using Quick ligase in a 1x Quick ligase buffer for 5 min at room temperature resulted in the construction of the $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} gene. The ligation product was transformed into EB5Alpha chemically competent cells and the cells were plated on TBdry plates supplemented with 45 μg/mL kanamycin. The sequence was confirmed by DNA sequencing.

Expression and purification

To express the triblock polypeptide $E^S{}_{80}$-$S^Q{}_{10}$-K_{12}, the plasmid as described above was transformed into *E. coli* BL21(DE3) chemically competent cells which were then used to inoculate a 10 mL starter culture of TB medium supplemented with 45 μg/mL kanamycin. After an overnight incubation at 37°C on a shaker at 200 rpm, the starter culture was used to inoculate a culture of 2 L TB medium supplemented with 45 μg/mL kanamycin. Cells were grown for a total time of 24 hours at 37°C on a shaker at 200 rpm. To induce protein expression, isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 1 mM was added to the medium 8 hours after the inoculation. After protein expression, cells were pelleted by centrifugation at 2000 x *g* for 10 min at 10°C and resuspended in 25 mL of PBS or HEPES (20 mM, pH 8.0). The cells were lysed by once pressing through a cooled French press at 1.5 bar. The lysate was centrifuged at 29000 x *g* for 12 min at 4°C to pellet insoluble cellular debris, and the supernatant (soluble lysate) was collected for further purification. To remove nucleic acids, the soluble lysate ($\sim$30 mL) was mixed with 4 mL of 10% polyethylenimine and centrifuged at 29000 x *g* for 12 min at 4°C. The polypeptide in the remaining supernatant (soluble cleared lysate) was further purified exploiting either the thermal properties of the elastin-like block ($E^S{}_{80}$) by using inverse transition cycling (ITC)[13], or by exploiting the strong positive charge of the oligolysine block (K_{12}) in cation-exchange chromatography.

Initial purifications used ITC and a bake-out step. These were performed as follows: (1) 1.5 to 2.0 M of ammonium sulphate or sodium chloride was added to the soluble cleared lysate to induce the phase transition and aggregation of the elastin-like block at room temperature. Maximum transition of the protein was ensured by incubation at 37°C for 15 min. (2) Aggregated proteins were pelleted by centrifugation at 40°C for 12 min at 29000 x *g*. (3) The supernatant was discarded and the protein in the pellet was solubilised in PBS by heating to 95°C for 15 min; this step is termed the bake-out. (4) The samples were cooled on ice and centrifuged at 4°C for 12 min at 29000 x *g* to remove any remaining insoluble

matter. Two rounds of ITC in combination with the bake-out were sufficient to obtain pure $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} polypeptide.

During later purifications, we used cation-exchange chromatography followed by ITC. For the cation-exchange chromatography, the soluble cleared lysate was loaded onto an UNO S6 column (BioRad) connected to a BioLogic DuoFlow chromatography system supplied with a QuadTec detector from BioRad. Unbound molecules were removed from the column by washing with 4 column volumes of 20 mM HEPES buffer (pH 8.0). Bound molecules were eluted by using a salt gradient from 0.0 to 1.0 M of NaCl in a 20 mM HEPES buffer (pH 8.0) applied over 10 column volumes. During elution, 1 mL fractions of the samples were collected from the column. Fractions containing the polypeptide were first dialysed against Milli-Q water and then subjected to one round of ITC using the following protocol: (1) aggregation of the ELP-containing protein was induced at room temperature by the addition of 2.5 to 3.0 M of ammonium sulphate to the dialysed sample, followed by incubation at 37°C for 15 min. (2) Sample was centrifuged at 40°C for 12 min at 29000 x *g* to pellet the aggregated proteins. (3) Proteins in the pellet were solubilised in 50 mM HEPES buffer (pH 8.0). To facilitate solubilisation, samples were vortexed at 1000 rpm for 90 min at room temperature. (4) To remove any remaining insoluble matter, the samples were first cooled on ice and then centrifuged at 4°C for 12 min at 29000 x *g*. Finally, all polypeptides were dialysed against Milli-Q water, lyophilised and stored at room temperature.

Characterisation

Purity of the polypeptides was assessed using SDS-PAGE and matrix-assisted laser desorption ionisation time-of-flight mass spectrometry (MALDI-TOF MS). SDS-PAGE was carried out using 10% mini-protean TGX precast protein gels, 1x Laemmli running buffer and Precision Plus Protein All Blue Prestained Protein Standard, and gels were stained with PageBlue protein staining solution. For MALDI-TOF MS, a matrix solution was prepared of 15 mg/mL 2.5-dihydroxyacetophenone and 4.5 mg/mL diammonium hydrogen citrate in 75% ethanol. Next, 1 volume of this matrix solution was mixed with 1 volume of 1 mg/mL protein and 1 volume of 3% (v/v) trifluoroacetic acid, and 1 μL of this mixture was dried on a 800 μm spot of an MTP AnchorChip 384 target (Bruker). Analysis was then carried out on an UltrafleXtreme mass spectrometer (Bruker).

Complex formation

Polypeptide stock solutions of 100 μM were prepared in a 10 mM phosphate buffer (pH 7.4). To promote solubilisation of the freeze-dried polypeptides, they were incubated at 65°C for 1 hour or 95°C for 10 min, as indicated in the main text. Next, polypeptides were further diluted in 10 mM phosphate buffer (pH 7.4) to a concentration of 1.8 μM or 0.18 μM and, if so indicated, mixed with DNA at a final DNA concentration of 1 μg/mL which corresponds to a protein-to-DNA molar charge ratio N/P of 7. All samples were incubated for 24 hours at room temperature to allow complexes to form.

Atomic force microscopy

For atomic force microscopy (AFM), 5 μL of a sample that was prepared as described in the previous section was deposited on a clean silicon surface and incubated for 2 min. Next the surface was rinsed with 1 mL of Milli-Q water to remove salts and non-absorbed particles and dried slowly under a gentle N_2 stream. Samples were analysed on a NanoScope MultiMode 8 system (Bruker) in the ScanAsyst (PeakForce Tapping) imaging mode, using ScanAsyst-Air cantilevers (Bruker). Areas of 5 x 5 μm were scanned at 512 x 512 pixels and a line rate of 0.977 Hz. All images were subjected to a second-order flattening using NanoScope Analysis 1.40 software. If so indicated, lengths of fibrils were measured using FiberApp software[16]. Settings used for FiberApp are: alpha = 0, beta = 500, gamma = 20, kappa1 = 20, kappa2 = 10, step = 1 pixel, iterations = 100, "Use A* pathfinding algorithm". Weight averaged fibril lengths were calculated assuming the mass of the fibrils was proportional to their length. The standard error of weighted means (s.e.m.$_w$) was approximated using the method of block averages: data was subdivided into 5 subsets, weight average lengths were calculated for the subsets and the standard error of the weighted means was taken to be the standard deviation of the weight average lengths for the subsets.

Gel shift assays

To quantify binding of the $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} triblock polypeptide to DNA, 52.5 ng of linear dsDNA was incubated at room temperature for 2 h with increasing concentrations of the protein. The DNA concentration was always 15 ng/μL. The mixtures were loaded onto 1% agarose gels that were supplemented with 1x SYBR Safe DNA stain. After electrophoresis for 90 min at 60 V, gels were scanned using a Bio-Rad Gel Doc EZ Imager.

Results & Discussion

Cloning, expression and purification

The plasmid encoding the triblock polypeptide $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} was constructed by recursive directional ligation by plasmid reconstruction (PRe-RDL) as described before[15], starting from a PRe-RDL plasmid previously prepared for the diblock $E^S{}_{80}$-$S^Q{}_{10}$ (Chapter 2). Figure 3.1 shows a schematic picture of the final $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} plasmid and a schematic representation of the structure of a single polypeptide when assembled onto DNA.

The $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} polypeptide was expressed in *E. coli* and initially purified according to a protocol developed previously for $E^S{}_m$-$S^Q{}_n$ diblocks (Chapter 2). This protocol combines inverse transition cycling (ITC)[13] with bake-out steps (15 min at 95°C) as for the diblocks it was found that ITC alone does not lead to sufficient purity of the polypeptides. Purity of the $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} triblocks was confirmed by SDS-PAGE analysis (Figure 3.2a). MALDI-TOF MS (Figure 3.2b) was used to verify the molar mass of the polypeptides. The experimentally determined (MALDI-TOF

MS) mass of 39166.2 Da corresponds well to the mass expected for this triblock of 39160.46 Da, assuming cleavage of the N-terminal methionine[17].

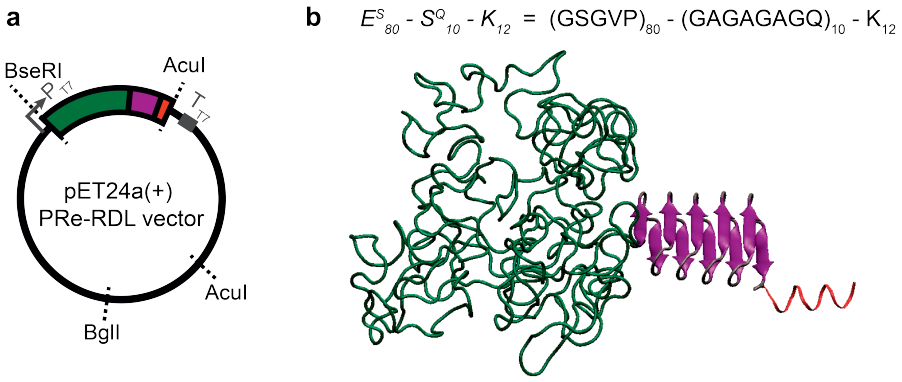


Figure 3.1 – Schematic of expression plasmid and cartoon of the structure of the E^S_{80}-S^Q_{10}-K_{12} triblock polypeptide when assembled onto DNA. a) PRe-RDL plasmid for the expression of the E^S_{80}-S^Q_{10}-K_{12} triblock copolymer. The elastin-like polypeptide block E^S_{80} is shown in green, the silk-like S^Q_{10} midblock in purple and the oligolysine K_{12} binding block in red. b) Cartoon of structure of the triblock polypeptide when assembled onto DNA. Colours as in a). Silk-like midblock is shown in β-solenoid configuration which it supposedly adopts after fibrillar assembly onto DNA[10,11]. Also shown is the full amino acid sequence of the triblock polypeptide.

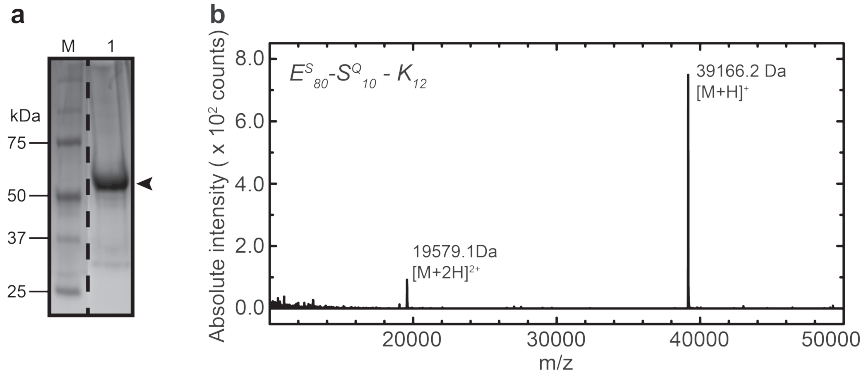


Figure 3.2 – Characterisation of the purified E^S_{80}-S^Q_{10}-K_{12} triblock polypeptide. a) SDS-PAGE analysis of the purified E^S_{80}-S^Q_{10}-K_{12} polypeptide. The arrowhead indicates the protein and lane M shows the molecular weight marker. The gel was stained with PageBlue. b) MALDI-TOF spectrum of purified E^S_{80}-S^Q_{10}-K_{12}.

Heat-treatment precludes proper assembly of virus-like particles

Based on prior experience with the C-$S^Q{}_{10}$-K_{12} triblock polypeptide, the expected assembly behaviour of the $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} polypeptide is that, by itself, it does not assemble into rod-shaped structures except when incubated at high concentrations (on the order of 100 μM). In contrast, in presence of a poly-anionic (DNA) template, it co-assembles with the template already at low polypeptide concentrations (on the order of a few μM)[7,18]. This expected behaviour was first tested for the $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} triblocks purified using a combination of ITC and a bake-out (15 min at 95°C). Indeed, atomic force microscopy (AFM) images show that at low concentrations, the purified triblocks do not assemble into rods in the absence of a DNA template (Figure 3.3a). In presence of DNA, however, the purified triblocks appear to simply coat the DNA templates while only occasionally forming the expected rod-shaped structures (Figure 3.3b). We reasoned that possibly the bake-out step at 95°C during purification had interfered with proper co-assembly of the triblocks with the DNA templates. For the previously developed triblock C-$S^Q{}_{10}$-K_{12}, temperatures of max. 65°C had been used to aid dissolution and such a treatment was shown not to interfere with proper assembly[7]. To test the hypothesis that the 95°C heat-treatment during purification precludes proper co-assembly of artificial viral capsid polypeptides with DNA templates, we incubated purified C-$S^Q{}_{10}$-K_{12} at either 65°C or 95°C for 10 min prior to incubation with DNA. Note that this treatment does not lead to precipitation of the polypeptides. We find that for C-$S^Q{}_{10}$-K_{12} heated at 65°C, rod-shaped VLPs form as expected (Figure 3.3d). In contrast, C-$S^Q{}_{10}$-K_{12} heated at 95°C does seem to coat the DNA templates, but rod-shaped VLPs are formed only occasionally (Figure 3.3e). These results strongly suggest that the heat treatment during purification of the $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} triblocks has precluded proper assembly of the VLPs.

Influence of the nature of the amorphous block on DNA binding properties

The next batch of the $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} triblocks was therefore purified using cation-exchange chromatography followed by ITC without a high temperature bake-out step. Purity and correct molar mass of the newly purified polypeptides was again confirmed using SDS-PAGE and MALDI-TOF MS (Figure 3.4). For this round of purification, the experimental mass is approximately 100 Da lower than expected, which is larger than the experimental uncertainty of the MALDI-TOF MS, suggesting that not only the N-terminal methionine, but also the subsequent glycine (75 Da) has been cleaved off.

To assess the impact of the nature of the amorphous block on the DNA binding properties of the artificial viral capsid polypeptides, an electrophoretic mobility gel shift assay (Figure 3.5) was performed, and compared with the earlier results for the C-$S^Q{}_{10}$-K_{12} triblocks[7]. The gel shift assay shown in Figure 3.5 clearly demonstrate that binding of $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} is highly cooperative, with high-mobility (naked DNA; linear dsDNA, 2000bp) complexes coexisting with low-mobility (encapsulated DNA) complexes. The shift to low-mobility, encapsulated

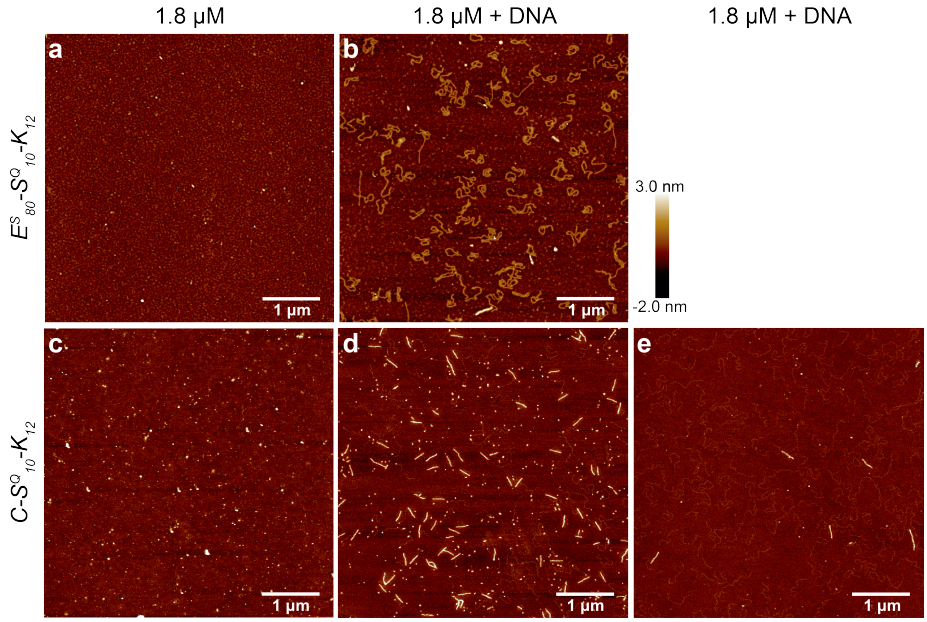


Figure 3.3 – Heat-treatment precludes proper assembly of virus-like particles. AFM of E^S_{80}-S^Q_{10}-K_{12} triblocks (purified by a combination of ITC and bake-out at 95°C) and C-S^Q_{10}-K_{12} incubated at room temperature for 24 h in the absence or presence of linear dsDNA (2000 bp, 1 μg/mL). Polypeptide concentration is 1.8 μM, polypeptide-to-DNA molar charge ratios N/P = 7. a) E^S_{80}-S^Q_{10}-K_{12} without DNA. b) E^S_{80}-S^Q_{10}-K_{12} with DNA. c) C-S^Q_{10}-K_{12} without DNA. d) C-S^Q_{10}-K_{12} with DNA. e) C-S^Q_{10}-K_{12} first heated at 95°C for 15 min, and then incubated with DNA.

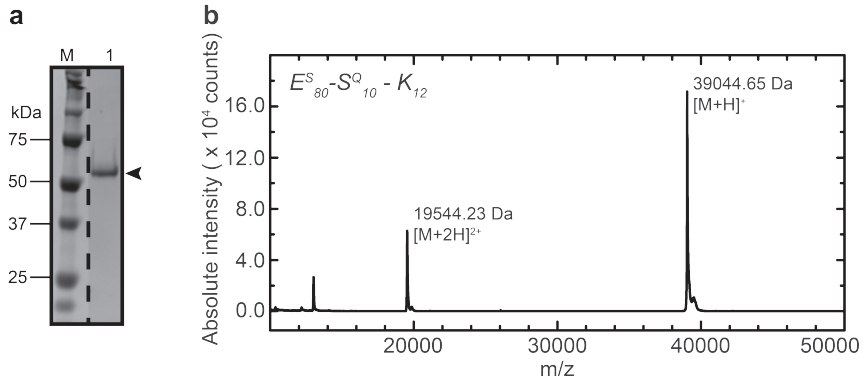


Figure 3.4 – Characterisation of E^S_{80}-S^Q_{10}-K_{12} purified using cation-exchange chromatography followed by ITC. a) SDS-PAGE, arrowhead indicates the protein. Lane M is a molecular weight marker. Gel was stained using PageBlue. b) MALDI-TOF MS analysis.

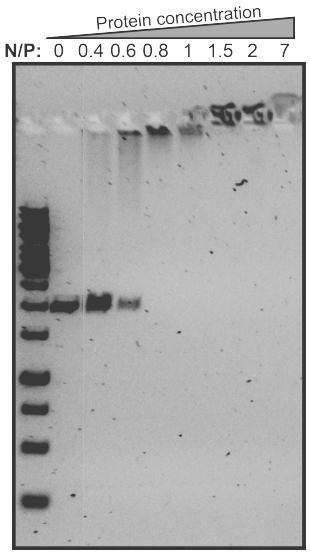


Figure 3.5 – Gel shift assay of E^S_{80}-S^Q_{10}-K_{12} (purified using cation-exchange chromatography followed by ITC). The DNA (linear dsDNA, 2000bp) concentration was kept constant at 15 ng/μL and the DNA was incubated with increasing concentrations of the E^S_{80}-S^Q_{10}-K_{12} triblock. The polypeptide-to-DNA molar charge ratios (N/P) are indicated above each lane. The total amount of DNA loaded into each well is 52.5 ng.

DNA was observed for polypeptide-to-DNA charge ratios (N/P) larger than $\sim$0.6. At high N/P ratio the complexes are observed to run slightly towards the anode, indicating a small net positive zeta-potential for the complexes. While the observation of highly cooperative binding is similar to that for the C-S^Q_{10}-K_{12} triblocks, for the latter triblock polypeptide, encapsulated DNA was only observed for N/P larger than $\sim$3 [7]. Since the amorphous block itself is not directly involved in binding to the DNA, this hints at a significant indirect influence of the nature of the amorphous block on complex formation.

Influence of the nature of the amorphous block on VLP size distributions

We hypothesise that the indirect influence of the nature of the amorphous block on the DNA binding properties of the artificial viral capsid polypeptides, occurs through influencing the assembly of the silk-like midblocks into fibrils: it is the fibril formation of the silk-like midblock that mediates binding cooperativity[7]. Therefore, we next investigated in detail the co-assembly of DNA and E^S_{80}-S^Q_{10}-K_{12} by AFM in order to compare with the corresponding data for C-S^Q_{10}-K_{12}.

Assembly and AFM imaging were performed at the same conditions and using the same procedures as previously used for the C-S^Q_{10}-K_{12} polymer (Figure 3.3d): a polypeptide concentration of 1.8 μM, and an excess of protein over DNA (charge ratio N/P = 7). AFM images of E^S_{80}-S^Q_{10}-K_{12} in the absence and pres-

ence of a 2000 bp linear dsDNA template are shown in Figure 3.6. In the absence of a template, many short rods and an occasional longer rod are formed (Figure 3.6a). In the presence of DNA, the number of longer rods significantly increases (Figure 3.6b). We also performed assembly and AFM imaging for a 10-fold lower concentration of both polypeptide and DNA (protein concentration 0.18 μM). AFM images for assembled structures in the presence and absence of the DNA template are shown in Figure 3.6c and d. Qualitatively, the 10-fold dilution does not seem to lead to major changes in the assembly. We believe that these observations point to complete encapsulation of the DNA templates, since otherwise we should have seen many partially- or non-encapsulated DNA molecules, that can easily be recognised by their smaller width and more flexible appearance.

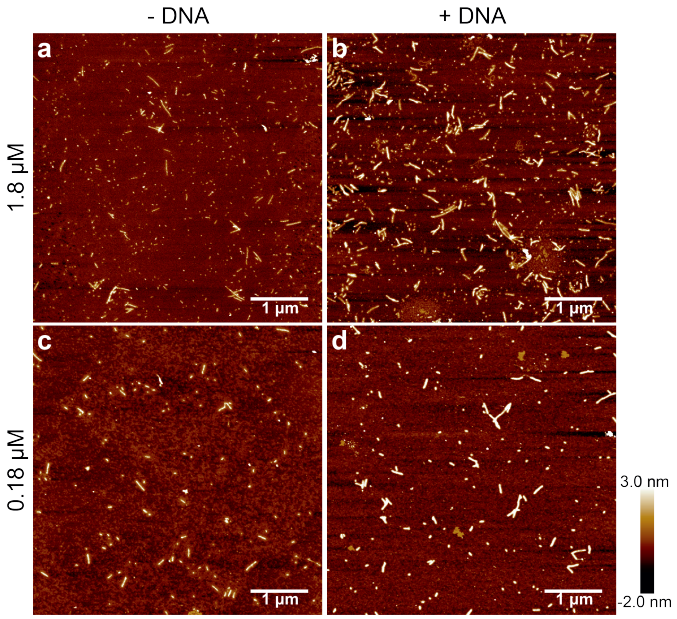


Figure 3.6 – Self-assembly and co-assembly of $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} triblocks at two polypeptide concentrations with and without a DNA template. Triblocks were solubilised at 65°C, diluted to a concentration of 1.8 or 0.18 μM and incubated at room temperature for 24 h in the absence or presence of linear dsDNA of 2000 bp. a) 1.8 μM polypeptide incubated in absence of DNA. b) 1.8 μM polypeptide incubated in presence of DNA. c) 0.18 μM polypeptide incubated in absence of DNA. d) 0.18 μM polypeptide incubated in presence of DNA.

Next, we performed a quantitative analysis of the AFM images by measuring both the lengths and heights of the fibrils from a number of AFM images and for a large number of fibrils. Histograms of fibril lengths and heights are shown in Figure 3.7, average values of fibril lengths and heights, and numbers of fibrils analysed per condition are given in Table 3.1. As observed previously for fibrils self-assembled from silk diblock polypeptides $E^S{}_m$-$S^Q{}_n$ (Chapter 2), length distributions

are very broad, with a very large number of very short fibrils, but a significant fraction of polypeptides being incorporated in the smaller number of much longer fibrils. Therefore, histograms of fibril lengths are reported as the weight fraction of total polypeptide incorporated in fibrils within a certain length range. In view of the broad distribution of the fibril lengths, a single average length is not very meaningful, and we report both the number average lengths $<L>_n$ and the weight average lengths $<L>_w$.

As we used a large excess of polypeptide (N/P = 7) to ensure complete encapsulation of all DNA templates, even in the presence of DNA most of the polypeptide does not co-assemble with this DNA. Hence, DNA encapsulation should show up in the length distributions as an additional peak on top of a background distribution for an excess of polypeptides that have not co-assembled with DNA, but only with themselves. Both the length and the height distributions show clear evidence that co-assembly with DNA occurs. First, in the length distributions for $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} fibrils, a peak appears for 151-200 nm fibrils when DNA is added. For the C-$S^Q{}_{10}$-K_{12} polymer it was established that when it encapsulates DNA, it also condenses DNA. Lengths of the rod-shaped VLPs were found to be reduced with respect to the contour length of the DNA template by a "packing factor" of 3 [7]. For a 2000 bp DNA template, this would imply a VLP length of around 230 nm. The somewhat smaller peak location of 151-200 nm could indicate that for the $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} polymer, the packing factor is somewhat higher than for the C-$S^Q{}_{10}$-K_{12} polymer. Further evidence for successful encapsulation of the DNA templates is provided by the height distributions, which show a clear shift to thicker fibrils for the samples with DNA.

More differences between the assembly and co-assembly behaviour of $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} as compared to that of C-$S^Q{}_{10}$-K_{12} became clear when comparing its distributions with the length and height distributions of C-$S^Q{}_{10}$-K_{12} fibrils assembled in the presence and absence of a 2000 bp DNA template at a polypeptide concentration of 1.8 μM and a polypeptide-to-DNA ratio of N/P = 7. The length and height distributions for C-$S^Q{}_{10}$-K_{12} fibrils are shown in Figures 3.7c and f, and average values are reported in Table 3.1. Very clearly, without a template the $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} fibrils at 1.8 μM have a length distribution that tails off at much larger lengths than the C-$S^Q{}_{10}$-K_{12} fibrils. Indeed, at 1.8 μM, weight average lengths in the absence of a template are $<L>_w = 86 \pm 18.9$ nm for $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} whereas for C-$S^Q{}_{10}$-K_{12}, $<L>_w = 34 \pm 1.2$ nm. As a consequence, when co-assembling C-$S^Q{}_{10}$-K_{12} with DNA, the separation in terms of lengths between fibrils encapsulating DNA and fibrils not encapsulating DNA is much more pronounced (compare Figures 3.7a and c, bottom panels).

In summary, replacing the collagen-like hydrophilic block with the elastin-like block preserves the fundamental features of VLP formation with DNA: DNA binding and self-assembly. However, as (at identical assembly concentrations) in the absence of a DNA template the $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} polypeptide forms much longer fibrils than C-$S^Q{}_{10}$-K_{12} , for the latter polypeptide the separation in length between fibrils encapsulating DNA and fibrils not encapsulating DNA is much greater.

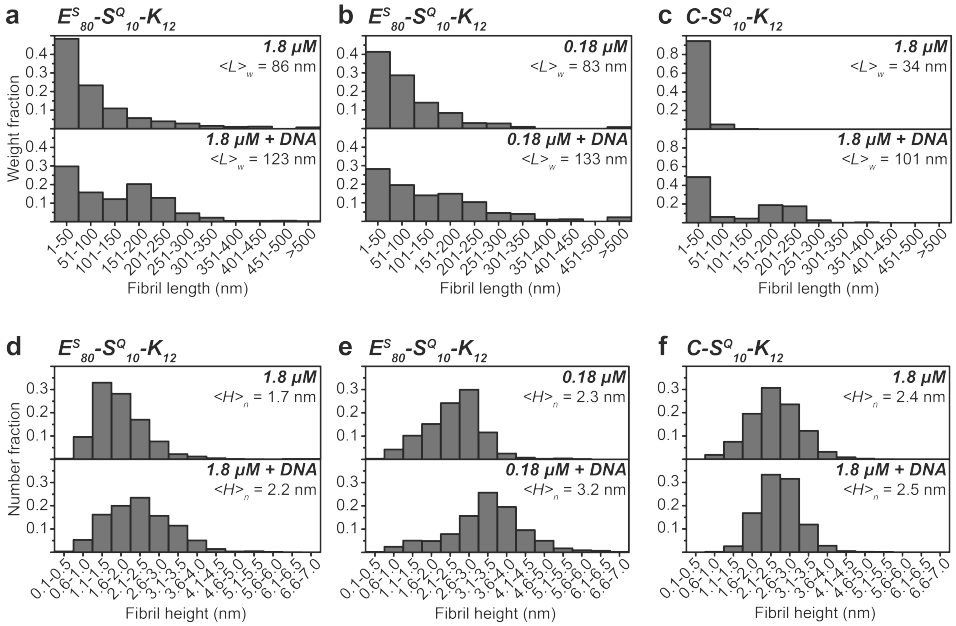


Figure 3.7 – Length and height distributions of C-S^Q_{10}-K_{12} and E^S_{80}-S^Q_{10}-K_{12} fibrils formed in presence and absence of DNA. Length distributions (weight fraction of polypeptide incorporated into fibrils with length in a given interval) and height distributions (number fraction of fibrils with height in a given interval) of fibrils formed by the C-S^Q_{10}-K_{12} and E^S_{80}-S^Q_{10}-K_{12} were derived from AFM images. Representative images for C-S^Q_{10}-K_{12} are shown in Figure 3.3c and d, and for E^S_{80}-S^Q_{10}-K_{12} in Figure 3.6. Length distributions for a) E^S_{80}-S^Q_{10}-K_{12} at 1.8 μM, b) E^S_{80}-S^Q_{10}-K_{12} at 0.18 μM and c) C-S^Q_{10}-K_{12} at 1.8 μM, and height distributions for d) E^S_{80}-S^Q_{10}-K_{12} at 1.8 μM, e) E^S_{80}-S^Q_{10}-K_{12} at 0.18 μM and f) C-S^Q_{10}-K_{12} at 1.8 μM are all split in two graphs with the top graphs showing the distribution in absence of DNA and the bottom graphs in presence of DNA. Note that in c) the y-axis scale differs from that in a) and b).

Table 3.1 – Average fibril lengths and heights for E^S_{80}-S^Q_{10}-K_{12} and C-S^Q_{10}-K_{12} polypeptides. Fibrils were assembled in presence or absence of 2000 bp dsDNA. Fibril lengths and heights were obtained by analysing AFM images (such as in Figures 3.3 and 3.6) and used to calculate the averages (± s.e.m.).

		N	$\langle L\rangle_n$ (nm)	$\langle L\rangle_w$ (nm)	$\langle H\rangle_n$ (nm)
E^S_{80}-S^Q_{10}-K_{12}	1.8 μM	2551	48 (± 0.8)	86 (± 18.9)	1.7 (± 0.01)
	1.8 μM + DNA	2443	63 (± 1.2)	123 (± 9.4)	2.2 (± 0.02)
	0.18 μM	1533	51 (± 1.0)	83 (± 7.1)	2.3 (± 0.02)
	0.18 μM + DNA	1233	66 (± 1.9)	133 (± 17.6)	3.2 (± 0.03)
C-S^Q_{10}-K_{12}	1.8 μM	1478	32 (± 0.2)	34 (± 1.2)	2.4 (± 0.02)
	1.8 μM + DNA	2708	49 (± 1.0)	101 (± 7.9)	2.5 (± 0.01)

Origin of differences in self-assembly is caused by differences in amorphous blocks

As also observed in our previous work on $E^S{}_m$-$S^Q{}_n$ diblock polypeptides (Chapter 2), for co-assembly onto a (DNA) template, the ideal behaviour of the polypeptides is that assembly into fibrils only occurs at a relatively high local protein concentration, that can be induced by accumulating the polypeptides onto the template via a template-binding block. Previously we established for $E^S{}_m$-$S^Q{}_n$ diblocks that this behaviour occurs for silk block lengths n just above a critical length required for fibril formation. At such critical silk block lengths, assembly into fibrils at lower concentrations is significantly reduced for longer amorphous blocks (larger values of m).

Here we observe that this "brake" on fibril formation at low concentrations afforded by the long amorphous blocks critically depends on their chemical nature. Both amorphous blocks adopt a random coil conformation and have a length of about 400 amino acids, but the collagen-like block *C* much more strongly opposes fibril formation at low concentrations than the elastin-like block $E^S{}_{80}$. We speculate that the key difference between the two blocks that gives rise to this different behaviour is their relative hydrophilicity. Increased hydrophilicity translates into larger coil sizes of the amorphous blocks and stronger repulsions between these more hydrated coils that are densely packed along the assembled silk fibrils (also see Figure 1.3b). Indeed, as shown in Table 3.2, the collagen-like and elastin-like blocks have GRAVY (grand average of hydropathy) values[19] that are significantly different from each other, and so are the values for the two corresponding triblock polypeptides.

In fact, the GRAVY value for $E^S{}_{80}$ is 0.200, indicating that this amorphous block has an intermediate hydrophilicity and is close to its theta-point, the transition point between a polymer in good solvent and a polymer in a poor solvent. On the other hand, the collagen-like *C* block is extremely hydrophilic, with a GRAVY value of -1.717.

Can we change the nature of the guest residue X in the elastin-like block consensus motif GXGVP to obtain a block equally hydrophilic as the collagen-like *C* block? This is somewhat difficult, as the most hydrophilic amino acids are charged residues. The amorphous block, however, needs to be net neutral, in order to not interfere with binding to DNA or to compete with DNA for the binding block. Therefore, one would need to use a zwitterionic elastin-like polypeptide, for example with alternating arginine and glutamic acid X residues, which would give a GRAVY value of -0.44. Although this value is not as low as for the collagen-like block, it is already much lower than the value for the $E^S{}_{80}$ block and in combination with increasing its length it may be possible to further reduce fibrillar self-assembly at low concentrations.

Table 3.2 – Grand average of hydropathy (GRAVY) values for the individual blocks and triblocks.

		GRAVY value
Individual blocks	*C*	-1.717
	$E^S{}_{80}$	0.200
	$S^Q{}_{10}$	0.037
	K_{12}	-3.900
Triblocks	$C\text{-}S^Q{}_{10}\text{-}K_{12}$	-1.471
	$E^S{}_{80}\text{-}S^Q{}_{10}\text{-}K_{12}$	0.072

The GRAVY values were obtained by calculating the sum of the hydropathy values of all amino acid residues and dividing this by the total number of residues in the sequence[19].

Concluding remarks

We have shown that self-assembly and co-assembly with DNA of the $C\text{-}S^Q{}_{10}\text{-}K_{12}$ and $E^S{}_{80}\text{-}S^Q{}_{10}\text{-}K_{12}$ polypeptides is significantly different. We expect that differences between the chemical nature of amorphous blocks may similarly influence the assembly of a broad range of natural silk polypeptides and engineered silk-like polypeptides.

Since the much more hydrophilic collagen-like polypeptide *C* is better able to act as a "brake" on self-assembly into fibrils at low concentrations than the much less hydrophilic elastin-like polypeptide $E^S{}_{80}$, we hypothesise that hydrophilicity of the amorphous block is a key variable determining the assembly of natural silk polypeptides and engineered silk-like polypeptides more generally. Indeed, a critical role for hydrophilicity in controlling self-assembly is reported for many other polypeptide block copolymers, with increased hydrophilicity invariably reducing self-assembly[20–22]. Such a role for the hydrophilicity of amorphous blocks in silk-like polypeptides is also consistent with the observation that long hydrophilic blocks surrounding silk-like self-assembly blocks may completely prevent fibril formation[23].

In the present case, a "brake" on self-assembly into fibrils at low concentrations was desired in order to separate concentration ranges and fibril lengths for non-templated and templated assembly of fibrils. An important conclusion is that the chemical nature of the amorphous blocks, in particular their hydrophilicity, plays an important role in determining the assembly of both natural silk-polypeptides and engineered silk-like polypeptides. This also means that it would be very interesting to more systematically study a broader range of amorphous structural polypeptides such as resilin, abductin and ColP[24,25] as amorphous blocks in the context of silk-like polypeptides, in order to better understand the mutual influence of crystallisable and amorphous blocks on the self-assembly of natural and engineered silk polypeptides.

References

[1] D. Chow, M. L. Nunalee, D. W. Lim, A. J. Simnick, and A. Chilkoti, "Peptide-based biopolymers in biomedicine and biotechnology," *Materials Science and Engineering R: Reports*, vol. 62, no. 4, pp. 125–155, 2008.

[2] N. A. Carter, X. Geng, and T. Z. Grove, *Design of self-assembling protein-polymer conjugates*, vol. 940. 2016.

[3] O. S. Rabotyagova, P. Cebe, and D. L. Kaplan, "Protein-based block copolymers," *Biomacromolecules*, vol. 12, pp. 269–289, feb 2011.

[4] J. G. Hardy, L. M. Romer, and T. R. Scheibel, "Polymeric materials based on silk proteins," *Polymer*, vol. 49, no. 20, pp. 4309–4327, 2008.

[5] C. Vepari and D. L. Kaplan, "Silk as a biomaterial," *Progress in Polymer Science (Oxford)*, vol. 32, no. 8-9, pp. 991–1007, 2007.

[6] S. H. Hiew and A. Miserez, "Squid Sucker Ring Teeth: Multiscale Structure-Property Relationships, Sequencing, and Protein Engineering of a Thermoplastic Biopolymer," *ACS Biomaterials Science & Engineering*, vol. 3, no. 5, pp. 680–693, 2017.

[7] A. Hernandez-Garcia, D. J. Kraft, A. F. J. Janssen, P. H. H. Bomans, N. A. J. M. Sommerdijk, D. M. E. Thies-Weesie, M. E. Favretto, R. Brock, F. a. de Wolf, M. W. T. Werten, P. van der Schoot, M. C. Stuart, and R. de Vries, "Design and self-assembly of simple coat proteins for artificial viruses," *Nature Nanotechnology*, vol. 9, pp. 698–702, aug 2014.

[8] M. Schor, A. a. Martens, F. a. DeWolf, M. a. Cohen Stuart, and P. G. Bolhuis, "Prediction of solvent dependent β-roll formation of a self-assembling silk-like protein domain," *Soft Matter*, vol. 5, no. 13, pp. 2658–2665, 2009.

[9] M. Schor and P. G. Bolhuis, "The self-assembly mechanism of fibril-forming silk-based block copolymers.," *Physical chemistry chemical physics : PCCP*, vol. 13, pp. 10457–10467, 2011.

[10] B. Zhao, M. A. Cohen Stuart, and C. K. Hall, "Dock n roll: folding of a silk-inspired polypeptide into an amyloid-like beta solenoid," *Soft Matter*, 2016.

[11] B. Zhao, M. A. Cohen Stuart, and C. K. Hall, "Navigating in foldonia: Using accelerated molecular dynamics to explore stability, unfolding and self-healing of the β-solenoid structure formed by a silk-like polypeptide," *PLOS Computational Biology*, vol. 13, no. 3, p. e1005446, 2017.

[12] M. W. T. Werten, W. H. Wisselink, T. J. Jansen-van den Bosch, E. C. de Bruin, and F. a. de Wolf, "Secreted production of a custom-designed, highly hydrophilic gelatin in Pichia pastoris," *Protein Engineering Design and Selection*, vol. 14, pp. 447–454, jun 2001.

[13] W. Hassouneh, T. Christensen, and A. Chilkoti, "Elastin-like polypeptides as a purification tag for recombinant proteins," in *Current Protocols in Protein Science*, pp. 6.11.1–6.11.16, aug 2010.

[14] R. Schipperus, G. Eggink, and F. A. De Wolf, "Secretion of elastin-like polypeptides with different transition temperatures by Pichia pastoris," *Biotechnology Progress*, vol. 28, no. 1, pp. 242–247, 2012.

[15] J. R. McDaniel, D. J. Callahan, and A. Chilkoti, "Drug delivery to solid tumors by elastin-like polypeptides.," *Advanced drug delivery reviews*, vol. 62, pp. 1456–1467, dec 2010.

[16] I. Usov and R. Mezzenga, "FiberApp: An Open-Source Software for Tracking and Analyzing Polymers, Filaments, Biomacromolecules, and Fibrous Objects," *Macromolecules*, p. 150210103859007, 2015.

[17] P. H. Hirel, M. J. Schmitter, P. Dessen, G. Fayat, and S. Blanquet, "Extent of N-terminal methionine excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 86, no. 21, pp. 8247–51, 1989.

[18] A. Hernandez-Garcia, M. A. Cohen Stuart, and R. de Vries, "Templated co-assembly into nanorods of polyanions and artificial virus capsid proteins," *Soft Matter*, vol. 14, no. 1, pp. 132–139, 2018.

[19] J. Kyte and R. F. Doolittle, "A simple method for displaying the hydropathic character of a protein," *Journal of Molecular Biology*, vol. 157, no. 1, pp. 105–132, 1982.

[20] K. Trabbic-Carlson, D. E. Meyer, L. Liu, R. Piervincenzi, N. Nath, T. LaBean, and A. Chilkoti, "Effect of protein fusion on the transition temperature of an environmentally responsive elastin-like polypeptide: A role for surface hydrophobicity?," *Protein Engineering, Design and Selection*, vol. 17, no. 1, pp. 57–66, 2004.

[21] T. Christensen, W. Hassouneh, K. Trabbic-Carlson, and A. Chilkoti, "Predicting transition temperatures of elastin-like polypeptide fusion proteins," *Biomacromolecules*, vol. 14, pp. 1514–1519, 2013.

[22] I. Weitzhandler, M. Dzuricky, I. Hoffmann, F. Garcia Quiroz, M. Gradzielski, and A. Chilkoti, "Micellar Self-Assembly of Recombinant Resilin-/Elastin-Like Block Copolypeptides," *Biomacromolecules*, vol. 18, no. 8, pp. 2419–2426, 2017.

[23] L. H. Beun, I. M. Storm, M. W. T. Werten, F. A. De Wolf, M. A. Cohen Stuart, and R. De Vries, "From micelles to fibers: Balancing self-assembling and random coiling domains in pH-responsive silk-collagen-like protein-based polymers," *Biomacromolecules*, vol. 15, no. 9, pp. 3349–3357, 2014.

[24] S. Roberts, M. Dzuricky, and A. Chilkoti, "Elastin-like polypeptides as models of intrinsically disordered proteins," *FEBS Lett.*, vol. 589, no. 19, pp. 2477–2486, 2015.

[25] S. Rauscher and R. Pomès, "Structural Disorder and Protein Elasticity," in *Fuzziness: Structural Disorder in Protein Complexes* (M. Fuxreiter and P. Tompa, eds.), vol. 1, pp. 159–183, New York, NY: Springer US, 2012.

4

Design of an artificial viral capsid protein recognising DNA packaging signals

In Chapter 3 we have developed an artificial viral capsid protein E^{S}_{80}-S^{Q}_{10}-B*, with an oligolysine nucleic acid binding block,* $B = K_{12}$*. This protein encapsulates DNA and mRNA irrespective of the sequence of its nucleic acid template. A mechanism natural viruses use to favour encapsulation of their own genome over the encapsulation of other nucleic acid sequences are so-called packing signals on their genomes, which promote capsid nucleation. To better mimic natural viruses, we here describe our work towards the development of a* de novo *designed non-viral capsid protein for which capsid assembly and encapsulation of a specific dsDNA template is induced by the recognition of a packaging signal on the DNA template. To that aim, we replaced the binding block* B *of the* E^{S}_{80}-S^{Q}_{10}-B *artificial viral capsid proteins by a zinc finger (ZnF) binding motif found in the* Drosophila melanogaster *GAGA factor. The* E^{S}_{80}-S^{Q}_{10}-*ZnF protein was produced in* E. coli*, and we observed that bacterial DNA co-purified with the protein in the form of rod-shaped pieces of encapsulated DNA. When using an improved purification protocol it was established that the purified protein was partially degraded at the C-terminus. We argue that the degradation most likely is due to proteolysis, which can possibly be prevented by improving the production and purification protocol such that eventually it can be established whether the* E^{S}_{80}-S^{Q}_{10}-*ZnF protein can indeed sequence-specifically encapsulate dsDNA templates.*

Introduction

Viruses are remarkable nanoparticles that evolved into particles that efficiently encapsulate their genome and deliver it into host cells. Viral genome encapsulation and assembly into viral particles, happens within the complex environment of a host cell. Given this complexity, one may wonder how the viral capsid proteins are able to specifically find and bind to their own genomic nucleic acids. It appears that so-called packaging signals on the viral genomes play a role in the specific recognition and encapsulation of their own genome[1,2]. Such packaging signals are short sequence elements that have been identified in the genome of many DNA and RNA viruses, and which are thought to promote the nucleation of virus particles with high efficiency[3].

Recognition of the packaging signals may be direct, as for tobacco mosaic virus (TMV) where $(NNG)_3$ motifs are very strong promotors of capsid nucleation[4], or more indirect as for baculovirus[5] and simian virus 40 (SV40)[6,7] where packaging signals are recognised by viral proteins that subsequently recruit the capsid proteins. It may also be the case that packaging signals are recognised via a specific secondary structure that they form in the genome, such as GC-rich packaging signals that form G-quadruplexes in the case of herpes viruses[8] and AT-rich sequences that form local DNA bends in the case of adenoviruses[1]. Packaging signals in many RNA viruses, form hairpin structures with at least part of the recognition sequence in the single strand loop[4,9].

As an alternative to viral vectors, it is desired to design safe and efficient nano-delivery devices for a wide range of therapeutic applications. Hence, a number of research groups are developing virus-inspired artificial protein capsids that can encapsulate and deliver DNA or RNA cargoes. If these artificial capsid proteins could, in a host organism, co-assemble with the genome coding for themselves, this would facilitate the development of such artificial capsids towards more efficient delivery agents via directed evolution. In particular, directed evolution approaches in which selection procedures occur outside the production host, such as when selecting for efficient uptake by a target cell, will benefit from this as it provides a way to easily identify the selected capsids on the basis of their encapsulated genome.

The development of a *de novo* (computationally-)designed protein-based artificial capsid in *Escherichia coli* was recently demonstrated by Baker and colleagues[10]. However, encapsulation of RNA by these artificial capsid proteins was completely sequence unspecific, such that the probability of encapsulation of its own capsid RNA genome was low. Inspired by natural viruses, strong packaging signals could potentially increase the probability for the specific encapsulation of the artificial capsid RNA genome. This mechanism was recently demonstrated by Hilvert and colleagues[11], who attached an RNA binding module to a natural, non-viral capsid-forming protein and included the corresponding recognition sequence in the capsid RNA genome. The presence of this packaging signal was shown to increase the encapsulation efficiency of the capsid genome as compared to the many other RNAs present in the *E. coli* cells that were used for the production and assembly of these capsids.

The generation of a *de novo* designed non-viral capsid protein that recognises packaging signals has not been demonstrated yet. We have previously designed and characterised artificial viral capsid proteins composed of simple sequence motifs[12] and here we aim to endow these *de novo* designed proteins with sequence-specific binding modules, such that we can induce the sequence-specific encapsulation of a dsDNA template.

The artificial viral capsid proteins are composed of repeats of simple amino acid motifs found in natural structural proteins such as silk, collagen and elastin. The triblock protein named C-$S^Q{}_{10}$-B was the first version for which we showed co-assembly of the proteins with DNA into rod-shaped virus-like particles (VLPs)[12]. The C-terminal binding block B is the oligolysine K_{12} and gives rise to non-specific binding to nucleic acid templates. Upon template-binding, the local protein concentration increases which triggers the self-assembly of the silk-like midblocks ($S^Q{}_{10}$ = $(\mathrm{GAGAGAGQ})_{10}$) into the rod-shaped cores of the VLPs. Finally, an N-terminal collagen-like block C = $(\mathrm{GXaaYaa})_{132}$ is used to provide colloidal stabilisation of the particles by preventing their aggregation in solution. The Xaa and Yaa residues in the C block are chosen to be mostly uncharged and hydrophilic such that the collagen-like polypeptides adopt random coil configurations over a wide range of solution conditions[13]. In the present thesis, we have shown that for the colloidal stability block, one can also use other hydrophilic polypeptides with random coil configurations. In particular, we have shown in Chapter 3 that one can replace the collagen-like polypeptide C = $(\mathrm{GXaaYaa})_{132}$ with the elastin-like polypeptide $E^S{}_{80}$ = $(\mathrm{GSGVP})_{80}$.

In this chapter we describe our work towards the development of an $E^S{}_{80}$-$S^Q{}_{10}$-B triblock protein that can encapsulate a double-strand (ds)DNA template sequence-specifically. In this design, B is a sequence-specific DNA binding domain so that nucleation of VLPs could be induced by binding to a packaging signal (viz. the recognition sequence of B) in the DNA.

Design considerations and candidate binding domains

Figure 4.1 illustrates the concept of an artificial packaging signal recognised by the artificial viral capsid proteins $E^S{}_{80}$-$S^Q{}_{10}$-B. It assumes that binding block B has a high affinity for its recognition sequence, as well as a much lower (but still significant) affinity for DNA with arbitrary sequences. We consider a mixture of DNA templates, some harbouring a repeat of the recognition sequence named the packaging signal ("specific DNA") and some not ("non-specific DNA"). Already at low protein concentrations in the bulk, the local concentration of proteins on the packaging signal can become high enough to give rise to particle nucleation and growth, and hence the preferential encapsulation of the specific DNA (Figure 4.1a). Only at much higher concentrations in the bulk, the local concentrations can also become large enough on non-specific DNA to allow for nucleation, particle growth and encapsulation (Figure 4.1b).

To select a DNA binding domain that will specifically recognise a packaging signal, we have to consider a few requirements: (1) The domain should have specific affinity for a packaging signal, but should also be able to bind non-sequence

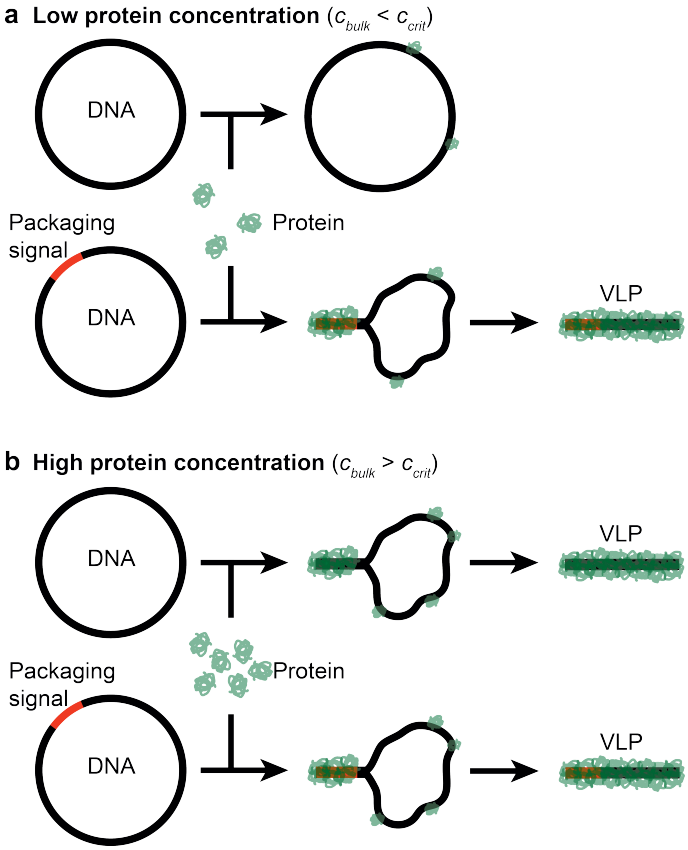


Figure 4.1 – Packaging signal induced encapsulation of a dsDNA template. a) At low concentration, specific DNA (bottom circle, harbouring a packaging signal indicated in red) is preferentially encapsulated over non-specific DNA (top circle). b) At high concentrations, both specific and non-specific DNA are encapsulated.

specifically with a lower affinity to allow particle growth beyond the nucleus. (2) The domain should be small: large binding blocks may sterically hinder the interaction between adjacent silk-like blocks required to form the silk-like core of the rod-shaped VLP. (3) The domain should preferably bind non-cooperatively, since we want to retain modularity in the sense that the cooperativity of the particle assembly is determined only by the central silk block. (4) The domain should be easy to produce in at least *E. coli* as this is the production host that we will use, but ideally it should also be possible to produce it in other hosts such as the yeast *Pichia pastoris* in which earlier versions of the artificial viral capsid proteins were produced. (5) Ideally, it should not be necessary to refold the domain e.g. after storage as a freeze-dried powder.

In addition to above mentioned requirements, it would be helpful if we have freedom in choosing the packaging signal. Three well known classes of "designer"

DNA binding domains[14] that qualify for this criterion are the zinc fingers[15], the DNA binding domains of transcription activator-like (TAL) effectors[16], and the nuclease-deficient CRISPR-associated 9 (dCas9) in association with CRISPR RNA[17,18]. Both the TAL effector proteins and the dCas9/RNA are too large to be packed at the required density along the silk-like core of the virus-like particles and hence they violate our criterion 2. The distance between two adjacent silk-like blocks in the core of the VLP is only 0.5 nm[19], so that we need to look for the smallest possible DNA binding domains that retain some sequence-specificity.

The ubiquitous zinc fingers proteins[7,15] typically consist of multiple "fingers". Each finger is about 30 amino acids long and recognises a specific combination of three base pairs in the DNA[15]. Good recognition and strong binding usually requires the presence of three or more adjacent zinc fingers. However, an apparent exception to this rule is the transcription factor GAGA (GAGA factor) from the fruit fly *Drosophila melanogaster* that specifically recognises the DNA sequence 5'-GAGAG-3' via a single zinc finger only[20]. Two additional basic regions at its N-terminus (Figure 4.3b) contribute to binding[21]. Basic regions 1 and 2 (BR1 and BR2) make specific contacts with two base pairs of the recognition sequence, both in the major and minor groove, thereby wrapping around the DNA (also see the red domain in Figure 4.3c)[20].

A minimal version of the GAGA factor zinc finger, consisting of the single finger of 30 amino acids plus 21 N-terminal amino acids, was shown to have high affinity for its DNA recognition sequence ($K_d \approx 400$ nM)[21,22]. It has been successfully expressed before in *E. coli* and can be stored as a freeze-dried product without loss of activity[20,21]. This binding domain thus checks at least three of the design requirements and is therefore the main candidate binding block that we consider here. Two more small sequence-specific DNA binding domains were identified that might also be suitable for our purpose. For completeness sake these are described in the Appendix. We however, decided to at least initially focus on the GAGA factor zinc finger domain.

Materials & Methods

Materials

The pET24a(+) PRe-RDL cloning vector[23] and the cloning vector encoding $E^S{}_{80}$-$S^Q{}_{10}$ (Chapter 2 of this thesis) were prepared previously. The pUC18 plasmid was purchased from Thermo Fisher Scientific. All oligonucleotides were synthesised by Sigma-Aldrich. For cloning we used restriction enzymes and chemically competent cells (NEB5-alpha, NEB10-beta and BL21(DE3)) from New England Biolabs. The T4 DNA ligase and FastAP thermosensitive alkaline phosphatase were obtained from Thermo Fisher Scientific. The DNA miniprep, gel purification and PCR purification kits used during cloning were purchased from Qiagen Inc. and the GoTaq G2 Green Master Mix from Promega was used for colony PCR. DNase I (RNase-free) was obtained from New England Biolabs. RNase A (DNase- and protease-free) is from Thermo Fischer Scientific. All SDS-PAGE materials (10% mini-protean TGX

precast protein gels, Precision Plus Protein All Blue Prestained Protein Standard and Laemmli running buffer) were ordered from Bio-Rad, except for PageBlue protein staining solution which was obtained from Thermo Fisher Scientific. For agarose electrophoresis, we used agarose from Brunschwig Chemie, SYBR Safe DNA staining from Life Technologies, and 6x DNA loading dye, GeneRuler 1kb DNA ladder or TrackIt 1Kb Plus DNA ladder, and 50x TAE buffer from Thermo Fischer Scientific.

Preparation of target plasmid

To construct a plasmid harbouring the recognition sequence of the zinc finger, two complementary oligonucleotides were designed with 5' overhangs complementary to the *EcoRI* restriction site (Figure 4.2a). They were annealed at a concentration of 2 μM in T4 DNA ligase buffer by heating to 95°C for 2 min, followed by slowly cooling to room temperature over 3 hours. To ligate the annealed oligonucleotides into the pUC18 plasmid, pUC18 was first linearised with 10 U of *EcoRI*. The 5' ends were dephosphorylated using 1 U of FastAP for 30 min at 37°C and the linearised plasmid was purified using a PCR purification kit. The annealed oligonucleotides harbouring the recognition sequence were then ligated into the linearised vector by using T4 DNA ligase for 1 hour at room temperature. The plasmid was transformed into NEB5-alpha chemically competent cells and plated onto TB-Agar plates supplemented with 100 μg/mL ampicillin. Colonies were screened with colony PCR and the presence of the recognition sequence $(GA)_{35}$ was confirmed by DNA sequencing.

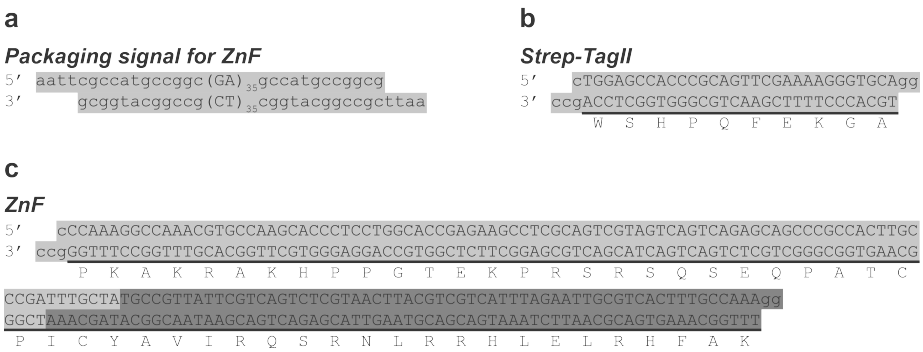


Figure 4.2 – Oligonucleotides designed for cloning and for creating the target plasmid with packaging signal. a) Oligonucleotides used for creating the target plasmid with the packaging signal. b) Oligonucleotides encoding Strep-TagII, with the amino acid sequence written underneath the oligonucleotides. c) Oligonucleotides used for cloning of the ZnF binding block. The set of oligonucleotides encoding the first half of the gene is marked light grey, the set of oligonucleotides encoding the second half of the gene is marked dark grey. The encoded amino acid sequence is written underneath the oligonucleotides.

Preparation of cloning vectors

The pET24a(+) PRe-RDL cloning vector harbouring the $E^S{}_{80}$-$S^Q{}_{10}$ gene was prepared previously (Chapter 2). To allow for convenient purification, we placed a Strep-TagII sequence at the 5' end of the gene to create a gene encoding for the strepII-$E^S{}_{80}$-$S^Q{}_{10}$ protein. To construct the plasmid, two oligonucleotides encoding the Strep-TagII (Figure 4.2b) were annealed in T4 DNA ligase buffer at a concentration of 2 μM by heating to 95°C for 2 min, followed by slowly cooling to room temperature over 3 hours. Next, approximately 2 μg of the cloning vector encoding $E^S{}_{80}$-$S^Q{}_{10}$ was linearised using 5 U of *BseRI* for 2 hours at 37°C. The 5' ends of the linearised vector were dephosphorylated using 1 U of FastAP for 30 min at 37°C and all enzymes were removed by using a PCR purification kit. The annealed Strep-TagII oligonucleotides were then ligated into the linearised vector in presence of T4 DNA ligase for 2 hours at room temperature, and transformed into NEB5-alpha chemically competent cells. The transformed cells were plated on TB-Agar plates supplemented with 45 μg/mL kanamycin and next day colony PCR was performed to screen for the colonies containing the cloning vector with the strepII-$E^S{}_{80}$-$S^Q{}_{10}$ gene. The sequence was confirmed by DNA sequencing.

To construct a gene encoding the minimal version of the GAGA factor zinc finger (ZnF), two sets of two complementary oligonucleotides were ordered (see Figure 4.2c). Both sets were designed to encode approximately half of the gene and have a 7 base pair overhang complementary to each other. To construct a PRe-RDL cloning vector encoding the ZnF gene, the two oligonucleotides of each set were annealed and the pET24a(+) PRe-RDL cloning vector was linearised as described above. Next, a two-step ligation was performed as follows: first the first set of annealed oligonucleotides (light grey in Figure 4.2c) were ligated into the linearised vector in presence of T4 DNA ligase for 4 hours on ice. The ligation was completed by the addition of the second set of annealed oligonucleotides (dark grey in Figure 4.2c) and fresh T4 DNA ligase, followed by an overnight incubation on ice. Next morning, the ligation product was transformed into NEB5-alpha chemically competent cells and the cells were plated on TB-Agar plates supplemented with 45 μg/mL kanamycin. Colonies were screened by colony PCR for the presence of the ZnF gene in the plasmid, which was confirmed by DNA sequencing.

Gene construction

The plasmid encoding for the full strepII-$E^S{}_{80}$-$S^Q{}_{10}$-ZnF protein was constructed by using recursive directional ligation by plasmid reconstruction (PRe-RDL)[23]. The protocol is described in more detail in Chapter 2 (Figure 2.1) with the following changes: the A-fragment was obtained from the cloning vector encoding strepII-$E^S{}_{80}$-$S^Q{}_{10}$ and the B-fragment from the vector encoding ZnF. The two fragments were ligated using T4 DNA ligase for 1 hour at room temperature, transformed in NEB10-beta chemically competent cells and plated on TB-Agar plates supplemented with 45 μg/mL kanamycin. Colonies were screened for the presence of the full gene by colony PCR and the sequence was confirmed by DNA sequencing.

Protein expression and purification

The plasmid encoding strepII-$E^S{}_{80}$-$S^Q{}_{10}$-ZnF was transformed into BL21(DE3) chemically competent cells and expressed as described in Chapter 2. To release the protein, cells were pelleted at 2000 x *g* for 10 min at 10°C, resuspended in 20 mL of binding buffer (100 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, pH 8.0) supplemented with 1 mM of DTT and lysed on ice by sonication (12 pulses (85W) of 10 seconds with 20 seconds intervals; Bandelin, Sonopuls VS70T probe). Next, inclusion bodies were removed by centrifugation at 29000 x *g* for 12 min at 4°C and the supernatant was subjected to another centrifugation step at 4°C for 30 min at 41500 x *g* to obtain a clear lysate containing the protein. The clear lysate was pressed through a 0.2 μm filter and finally purified on a StrepTrap HP column (GE Healthcare Life Sciences) according to manufacturer's protocol: after protein binding, the column was washed with binding buffer and the protein was eluted using elution buffer (100 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 2.5 mM D-desthiobiotin (IBA Lifesciences), pH 8.0). Those fractions collected during column washing that contained pure strepII-$E^S{}_{80}$-$S^Q{}_{10}$-ZnF according to SDS-PAGE analysis and the fractions collected during elution were pooled prior to dialysis against Milli-Q water. At last, the protein was lyophilised and stored at room temperature.

The strepII-$E^S{}_{80}$-$S^Q{}_{10}$-ZnF protein was expressed a second time and purified using a slightly adjusted protocol. This time the pelleted cells were resuspended in 25 mL of PBS binding buffer (20 mM sodium phosphate, 280 mM NaCl, 6 mM KCl, pH 7.4) and lysed by applying a pressure of approximately 16000 psi onto the cells using a cooled French press (Thermo Fischer Scientific). After removal of the inclusion bodies, nucleic acids were pelleted by mixing the lysate with 4 mL of 10% polyethylenimine and centrifuging at 29000 x *g* for 12 min at 4°C. The remaining supernatant was centrifuged again at higher speed (41500 x *g* for 15 min at 4°C) to obtain a clear lysate. Next, the pH was adjusted from pH 10.0 to 7.4 using HCl and the resulting aggregated materials were pelleted through another centrifugation step at 41500 x *g* for 15 min at 4°C. The protein was now further purified on a StrepTrap HP column, using the PBS binding buffer to wash the column and eluting in a PBS elution buffer (20 mM sodium phosphate, 280 mM NaCl, 6 mM KCl, 2.5 mM D-desthiobiotin, pH 7.4). We found that the protein did not bind to the column efficiently, but fractions collected during washing appeared to contain relatively clean protein nevertheless, according to SDS-PAGE analysis. These fractions were pooled and dialysed against Milli-Q water. Finally, the protein was lyophilised and stored at room temperature.

Protein characterisation

Protein purity was assessed using SDS-PAGE and matrix-assisted laser desorption ionisation time-of-flight mass spectrometry (MALDI-TOF MS) as described in Chapter 3 of this thesis.

Protein assembly

For atomic force microscopy (AFM) and gel shift assays, weighed portions of freeze-dried protein were solubilised at a concentration of 100 μM by vortexing in buffer (10 mM sodium phosphate, 1 mM DTT, 110 μM $ZnCl_2$, pH 7.4). Lower concentrations were prepared by immediately diluting the 100 μM protein solution in PB (10 mM sodium phosphate, pH 7.4). The protein solutions prepared for the gel shift assays were incubated at room temperature for 2 hours in presence of 15 ng/μL DNA (pUC18 or pCU18-specific) to allow encapsulation of the dsDNA. Solutions prepared for AFM were allowed to self-assemble by incubation at room temperature for 24 hours in absence of DNA.

For experiments to identify the origin of an unknown hypothetical template for fibril formation, 100 μM protein solutions were prepared by solubilising weighed portions of freeze-dried protein in either DNase I buffer (100 mM Tris-HCl, 2.5 mM $MgCl_2$, 0.5 mM $CaCl_2$, pH 8.0) or, for RNase A treatment, in PB. In the case of the AFM experiments, the 100 μM protein solutions were incubated for 4 days prior to adding 1 U of the enzyme to 50 μL of the protein solution. Solutions were then incubated for 3 hours at 37°C (for DNase I treatment) or at room temperature (for RNase A treatment), followed by a 1 day incubation at room temperature. To analyse the origin of the hypothetical template by a gel shift assay, 300 U DNase I was added to a 15 mL protein solution of 100 μM. To allow complete digestion of all free DNA, the solution was subsequently incubated at room temperature for 90 min, at 37°C for 75 min and again at room temperature overnight.

Atomic force microscopy

Samples for AFM imaging were prepared and analysed as described in Chapter 2 of this thesis.

Gel shift assays

To assess whether the hypothetical unknown template is DNA, 10 μL of the 100 μM protein solution before and after DNase I treatment was loaded on a 0.8% agarose gel which was supplemented with 1x SYBR Safe DNA stain. Samples were run at 120 V for 30 min and imaged with a Bio-Rad Gel Doc EZ imager.

For the gel shift assays to determine efficiency of protein binding to the DNA, 52.5 ng of DNA (pUC18 or pUC18-specific) with increasing concentrations of protein was mixed and incubated for 2 hours at room temperature. The mixtures were loaded on a 1% agarose gel supplemented with 1x SYBR Safe DNA stain. The samples were electrophoresed at 60 V for 90 min and the DNA was visualised with a Bio-Rad Gel Doc EZ imager.

Results

Protein design

Towards the development of an artificial viral capsid protein that, like many natural viruses, is able to specifically encapsulate its own DNA genome, we here designed a new variant of the previously described $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} protein (Chapter 3) in which the non-specific binding block K_{12} was replaced by a small DNA binding block that specifically recognises a short DNA sequence. We focus on the single zinc finger of the GAGA factor (ZnF). For easy purification using affinity chromatography, a strep-tag (strepII) was added to the N-terminus (the $E^S{}_{80}$ side) of the protein during cloning of the construct. Preliminary work for two more binding blocks (the leucine zipper domain of GCN4 (bZIP) and an AT-hook from the HMGA1 proteins) are described in the Appendix. The gene coding for strepII-$E^S{}_{80}$-$S^Q{}_{10}$-ZnF was constructed by PRe-RDL, a cloning strategy developed by Chilkoti and co-workers[23] and described in more detail in Chapter 2. A representation of the PRe-RDL vector with the newly constructed gene, a schematic representation of the ZnF binding block, and a cartoon of the structure of the complete protein when bound to DNA are shown in Figure 4.3a, b and c respectively. Oligonucleotides used for cloning and the encoded protein sequence are shown in Figure 4.2c.

Presence of an unknown template may influence fibril formation

Using atomic force microscopy (AFM) we first explored the self-assembly behaviour of the strepII-$E^S{}_{80}$-$S^Q{}_{10}$-ZnF protein, in the absence of a DNA template. Freshly dissolved protein in presence of a 1.1 equivalent of zinc (required for the ZnF block to adopt its functional conformation) was incubated for 24 h to allow for potential self-assembly of proteins into fibrils. To also analyse the effect of concentration on protein fibril formation, as observed previously for $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} and $E^S{}_{80}$-$S^Q{}_{10}$ proteins (Chapter 2 and 3), solutions were prepared at both a low (1.8 μM) and a high (100 μM) concentration. Occasional long fibrils were observed at low concentrations (Figure 4.4a), while at high concentrations, a large number of long fibrils were observed (Figure 4.4c). While the previously described $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} also forms fibrils in the absence of a template, these appear to be shorter than the fibrils observed here for strepII-$E^S{}_{80}$-$S^Q{}_{10}$-ZnF, especially at low concentrations (see Figures 3.6a and 3.7a in Chapter 3). Higher resolution scans (Figures 4.4b,d) reveal a second unexpected feature at high protein concentrations. As pointed out by the white arrowheads in Figure 4.4d, thin and flexible strands/wires seem to be attached to the sides of the fibrils or sticking out at their ends. Some of these strands make connections between fibrils. The appearance (dimensions and flexibility) of these strands is very similar to that of nucleic acids dried on a silica surface. These observation raise the question whether DNA or RNA may have been inadvertently co-purified together with the strepII-$E^S{}_{80}$-$S^Q{}_{10}$-ZnF protein and may have caused more extensive fibril formation than would have been expected in the absence of a template.

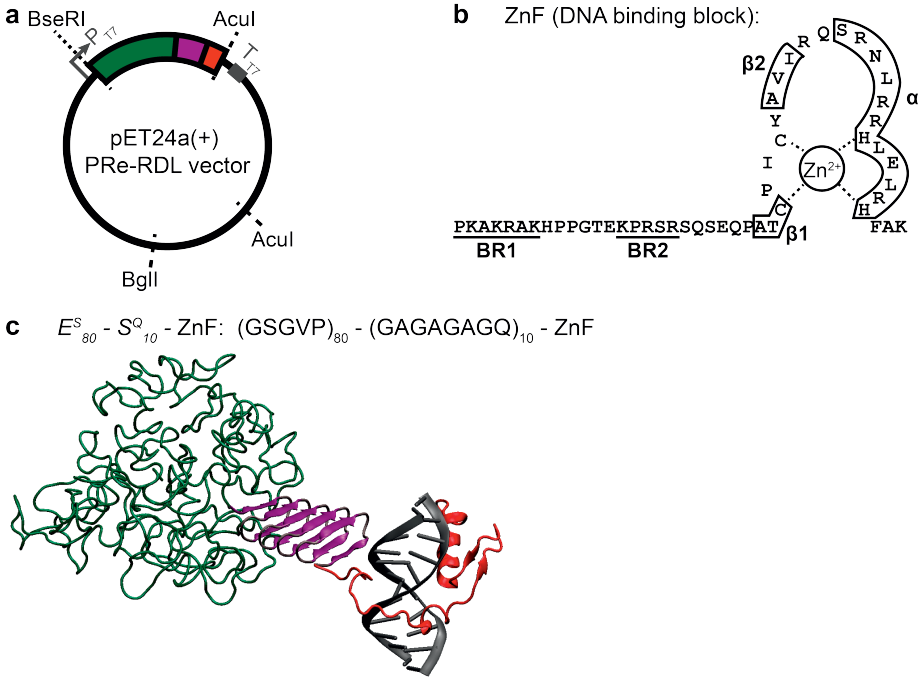


Figure 4.3 – Design of an artificial viral capsid protein sensitive to packaging signals. a) PRe-RDL cloning vector with the strepII-$E^S{}_{80}$-$S^Q{}_{10}$-ZnF protein gene. Green represents the $E^S{}_{80}$ random coil block with the N-terminal strepII tag, purple represents the $S^Q{}_{10}$ silk block, red represents the ZnF DNA binding block. b) Amino acid sequence and representation of solution structure of the ZnF block. BR1 and BR2 indicate basic regions 1 and 2, β1 and β2 indicate two β-sheets, and α indicates an α-helix. c) Cartoon of (hypothetical) solution structure of strepII-$E^S{}_{80}$-$S^Q{}_{10}$-ZnF protein when bound to target DNA. Colours as in a).

The unknown template possibly is bacterial DNA

In an attempt to deduce the origin of the hypothetical unknown template, 100 μM protein solutions were treated with either DNase I or RNase A. No zinc was added to the buffer to limit binding and encapsulation of any potential template by the strepII-$E^S{}_{80}$-$S^Q{}_{10}$-ZnF proteins. After both treatments, fibrils remained visible as seen in AFM (Figure 4.5). Higher resolution scans, however, did not show any flexible strands sticking out or connected to the fibrils upon DNase I treatment (Figure 4.5a,b), whereas we did observe such strands in the RNase A treated sample (Figure 4.5c,d). These findings suggest that the unknown template observed in Figure 4.4 is susceptible to digestion by DNase I, but not by RNase A, and could be bacterial DNA.

To further verify this, agarose gel electrophoresis was performed on 100 μM protein solutions either exposed to DNase I for 20 h or not. SYBR safe was used as nucleic acid stain in the agarose gel electrophoresis and results are shown in

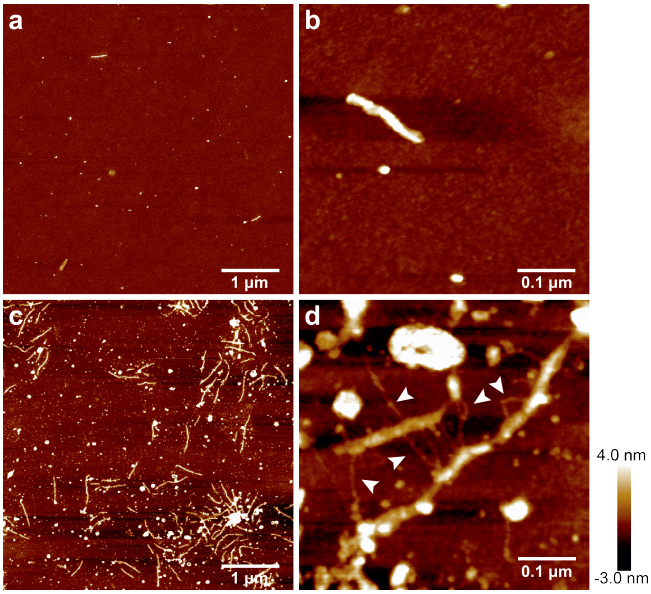


Figure 4.4 – An unknown template possibly causes enhanced fibril formation of purified strepII-$E^S{}_{80}$-$S^Q{}_{10}$-ZnF. AFM images of purified strepII-$E^S{}_{80}$-$S^Q{}_{10}$-ZnF protein, incubated in the absence of a DNA template. a) 1.8 μM protein. b) 1.8 μM protein, high resolution zoom. c) 100 μM protein. d) 100 μM protein, high resolution zoom. White arrowheads indicate thin strands connecting together different fibrils.

Figure 4.5e. For both samples (with and without DNase I treatment) a clear band was visible with very low mobility. This matches the hypothesis that the co-purified templates are nucleic acids and that these nucleic acids have been encapsulated, as such complexes would move with very low mobility. It is, however, not possible from this gel to identify the nucleic acids as DNA, since the bands, both in absence and presence of DNase I treatment, have roughly the same intensities. As previously observed, encapsulated DNA is well protected against DNase I[12], and the enzyme probably only degrades the small portion of DNA that is visible as thin strands on the outside of the VLPs in the AFM images. Consequently, both samples have roughly the same amount of DNA and show similar intensities on the gel.

In summary, is seems very likely that bacterial DNA has inadvertently been co-purified with the strepII-$E^S{}_{80}$-$S^Q{}_{10}$-ZnF protein and has become partially encapsulated by the protein. Indeed, the purification protocol did not involve high salt conditions that might break interactions between the proteins and bacterial DNA, nor was the bacterial lysate treated with nucleases before further purification steps.

Newly purified $E^S{}_{80}$-$S^Q{}_{10}$-ZnF is partly truncated at the C-terminus

Next, we decided to produce the strepII-$E^S{}_{80}$-$S^Q{}_{10}$-ZnF protein again using an additional step to remove any bacterial nucleic acids from the lysate. To this purpose, positively charged polyethylenimine was mixed with the lysate followed

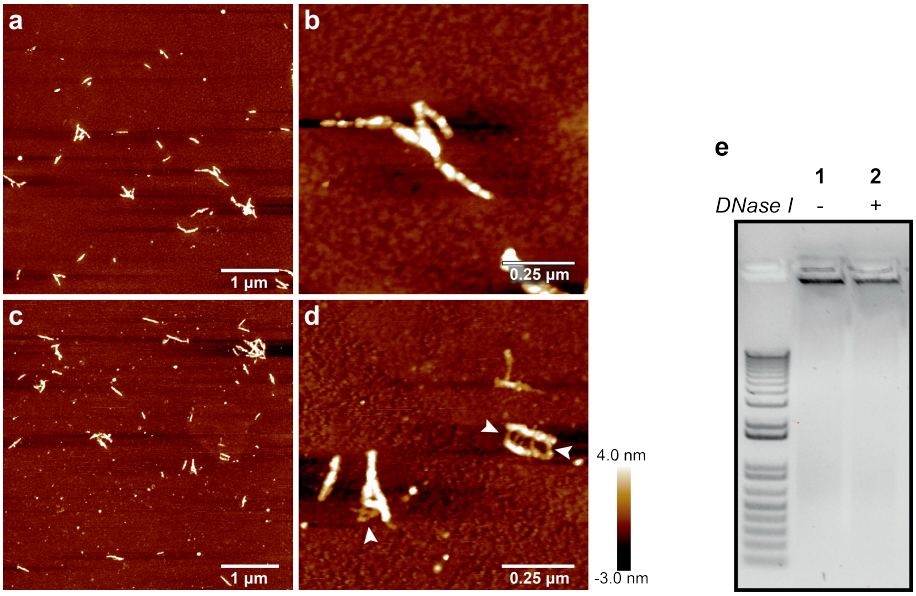


Figure 4.5 – The unknown template possibly is bacterial DNA. AFM images (a-d) and agarose gel electrophoresis (e) of purified strepII-$E^S{}_{80}$-$S^Q{}_{10}$-ZnF protein treated with DNase I or RNase A. a) 100 μM treated with DNase I. b) 100 μM treated with DNase I, high resolution zoom. c) 100 μM treated with RNase A. d) 100 μM treated with RNase A, high resolution zoom. e) Agarose gel electrophoresis. Lane 1: before DNase I treatment, lane 2: after DNase I treatment. Ladder is TrackIt 1kb plus DNA ladder.

by centrifugation to pull-down negatively charged macromolecules including DNA and RNA. Expression and purification of strepII-$E^S{}_{80}$-$S^Q{}_{10}$-ZnF was analysed using SDS-PAGE, showing a single band (Figure 4.6a). As also observed for the previously described $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} (Chapter 3), the protein migrates slower than expected for a 44.7 kDa protein, most likely as a consequence of the poor SDS binding by these type of proteins[12,13,24].

However, MALDI-TOF MS (results shown in Figure 4.6b) does not show a peak at the expected molecular weight of 44745 Da. Instead, three smaller peaks were observed corresponding to molecular weights of 42212.7, 39019.8 and 38230.8 Da. Table 4.1 summarises the expected and experimentally observed masses. The masses observed in MALDI-TOF MS are approximately 2500-6500 Da smaller than expected. Such big differences most likely indicate either that a truncated version of the protein was produced in the bacteria, or that the protein was partially cleaved during or after expression.

Production of a truncated protein is unlikely as previous versions of the artificial viral capsid protein were produced in bacteria without any problems, and literature also describes the bacterial expression of the ZnF sequence[20,21]. Furthermore, sequencing of the full strepII-$E^S{}_{80}$-$S^Q{}_{10}$-ZnF gene did not show any unexpected frameshifts and/or introduced stop codons. We therefore believe that the protein

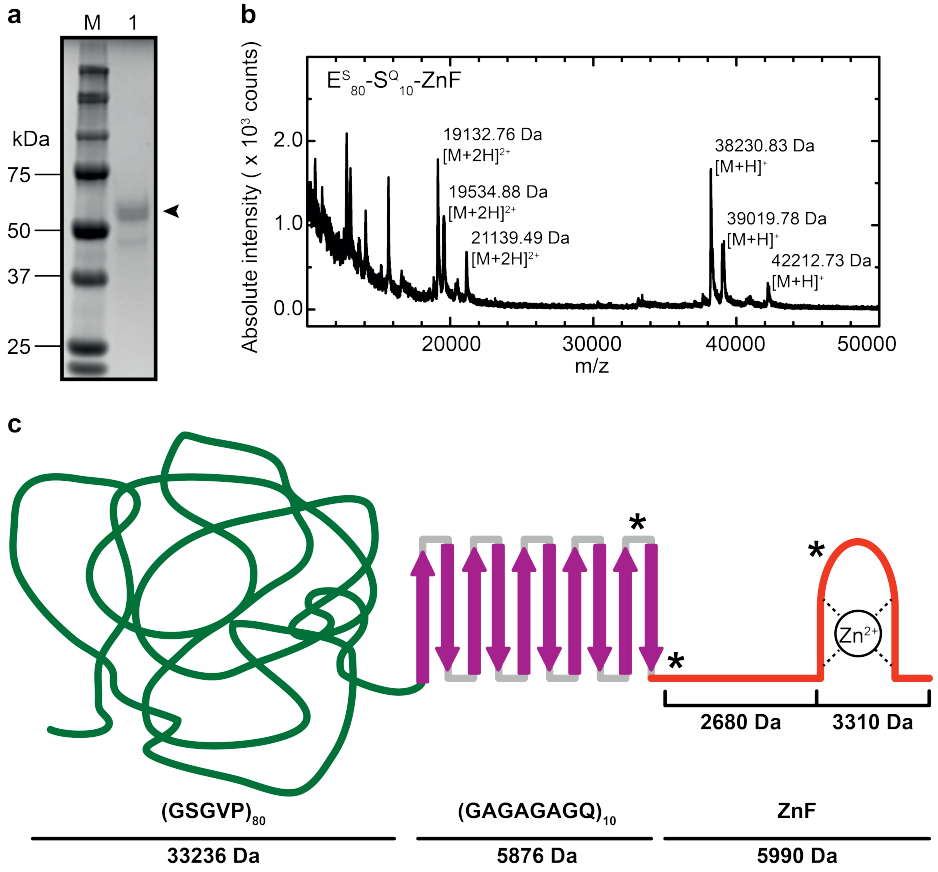


Figure 4.6 – Characterisation of strepII-E^S_{80}-S^Q_{10}-ZnF purified with an additional polyethylenimine precipitation step. a) SDS-PAGE analysis. b) MALDI-TOF spectrum. c) Schematic image of the E^S_{80}-S^Q_{10}-ZnF protein including molar masses of each block and possible cleavage sites as determined from peaks observed in MALDI-TOF MS (indicated with *).

Table 4.1 – Expected and experimental (MALDI-TOF MS) molar masses of strepII-E^S_{80}-S^Q_{10}-ZnF

Expected MW (Da)	Experimental MW (Da)	Mass deviation (Da)
44744.76	42212.7	-2532.06
	39019.8	-5724.96
	38230.8	-6513.96

was proteolytically cleaved during or after expression. Successful purification of the protein using the N-terminal strep-tag suggests that the N-terminal $E^S{}_{80}$ block is still intact, and that cleavage likely occurred at the C-terminal end, which includes the C-terminal ZnF block.

Figure 4.6c shows a schematic image of the $E^S{}_{80}$-$S^Q{}_{10}$-ZnF protein indicating possible positions of cleavage based on the observed peaks in MALDI-TOF MS. A first possible cleavage site, resulting in the biggest protein (42212.7 Da) identified with MALDI-TOF MS, is located around the second β-sheet within the zinc finger (compare with Figure 4.3b). Most of the "finger" structure is thereby lost, but the basic regions are still intact. A second possible cleavage site is located near the N-terminal end of the ZnF block and results in a protein of about 39019.8 Da which lost its full ZnF block. Finally, a third possible cleavage site is also located near the N-terminal end of the ZnF block, but in this case the ZnF block and a few C-terminal amino acids of the silk-like block may have been cleaved off.

For all three cases, the DNA binding ability of the protein via the ZnF block is expected to be compromised, with only a few low-affinity specific interactions with the DNA expected to be formed by the unaffected BR1 and BR2 regions in the biggest protein of 42212.7 Da. Additional experiments were performed in order to further confirm that truncations had occurred at the C-terminus, that should impair DNA binding but not self-assembly per se.

DNA binding of truncated protein

To assess the DNA binding ability of the purified strepII-$E^S{}_{80}$-$S^Q{}_{10}$-ZnF protein, gel shift assays were performed using pUC18 plasmids either containing a $(GA)_{35}$ packaging signal insert or not containing the insert. Increasing concentrations of the protein were mixed with a constant concentration of DNA and incubated for 2 h to allow DNA binding and encapsulation. Protein-to-DNA ratios are given as the number of proteins per base pair ($^{\text{proteins}}/_{\text{bp}}$). For an intact GAGA ZnF binding block the minimal DNA binding site is known to be 9bp[20], hence for an intact binding block we would expect a significant gel shift at a ratio of $^1/_9$ or higher.

Results are shown in Figure 4.7. The protein binds similarly to both types of DNA and the shift to low mobility complexes only occurs at a very high protein-to-DNA ratio of one protein per 2bp. Hence, there seems to be only non-sequence specific, low-affinity binding, consistent with the observation of a C-terminally truncated protein.

Self-assembly in the absence of template is unaffected for the truncated protein

Assuming that the binding block *B* has no (or only a minor) influence on the assembly of triblock proteins $E^S{}_{80}$-$S^Q{}_{10}$-*B* into fibrils, we expect that for the C-terminally truncated strepII-$E^S{}_{80}$-$S^Q{}_{10}$-ZnF protein, the concentration-dependent self-assembly is similar to that of $E^S{}_{80}$-$S^Q{}_{10}$-K_{12}. This is indeed borne out by experiments. AFM analysis reveals that hardly any fibrils are formed at a low protein concentration of

1.8 μM (Figure 4.8a), whereas short fibrils were clearly present at higher concentrations (Figure 4.8b,c). Furthermore, the fibril length increases with the protein concentration from a weight averaged length of 116 nm at 21.24 μM to 238 nm at 100 μM (see histograms in Figure 4.8b,c and Table 4.2). These results are comparable to those we previously found for $E^S{}_{80}$-$S^Q{}_{10}$ diblocks (Chapter 2). This not only (again) confirms that the truncation has occurred C-terminally, but also that the triblock design for the artificial capsid polypeptide is quite modular, with the template binding functionality being largely independent from the self-assembly functionality.

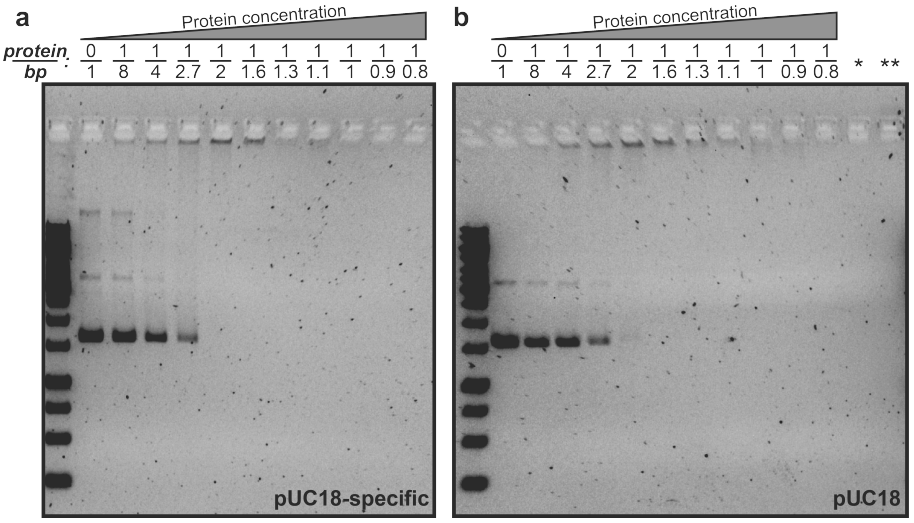


Figure 4.7 – DNA binding of the truncated protein. Gel shift assay: the value indicated above each well is the ratio of the molar concentration of protein to the molar concentration of DNA base pairs. Total DNA concentration was fixed at 24.3 μM. a) pUC18-specific plasmid DNA (with $(GA)_{35}$ packaging signal insert) with increasing concentrations of protein. b) pUC18 plasmid DNA (no insert) with increasing concentrations of protein. Last two lanes: * protein only at 21.24 μM, ** protein only at 100 μM.

Table 4.2 – Number ($<L>_n$) and weight ($<L>_w$) averaged lengths of protein fibrils.

	N	$<L>_n$ (nm)*	$<L>_w$ (nm)*
21.24 μM	483	108 ($\pm$ 29.1)	116 ($\pm$ 2.0)
100 μM	478	197 ($\pm$ 90.3)	238 ($\pm$ 16.1)

* Number and weight averaged particle lengths ($\pm$ s.e.m.) were calculated using the data obtained from AFM images (see Figure 4.8). All particles shorter than 50 nm were omitted from the calculations.

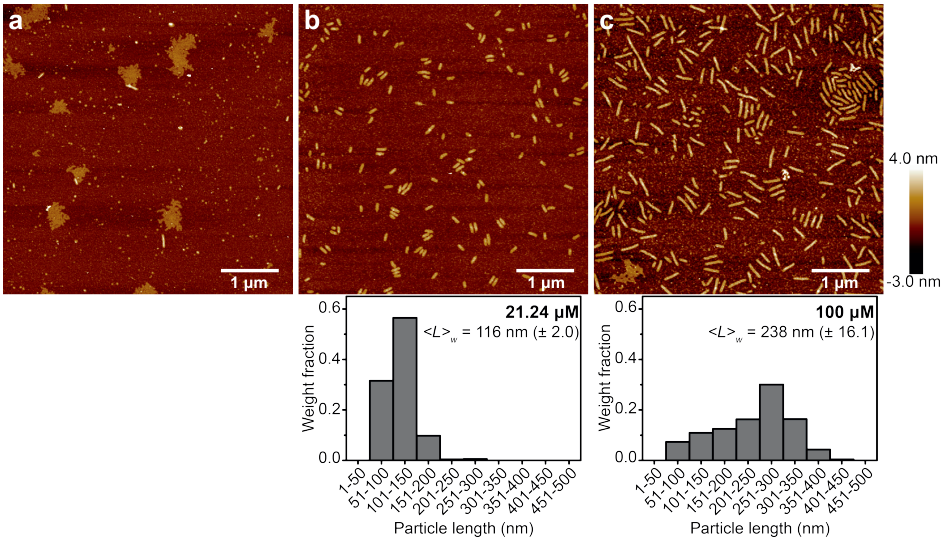


Figure 4.8 – Self-assembly in the absence of template is unaffected for the truncated protein. AFM images (top panels) and fibril length distributions deduced from AFM images (bottom panels; weight fraction of protein incorporated into fibrils as a function of fibril length) for strepII-$E^S{}_{80}$-$S^Q{}_{10}$-ZnF protein in the absence of a template. a) 1.8 μM protein. b) 21.24 μM protein. c) 100 μM protein. Insets of histograms also give the weight averaged lengths $<L>_w$ of the protein fibrils.

Discussion & Outlook

Recently, two research groups reported the development of self-assembling VLPs that encapsulate their own RNA genome inside bacteria[10,11]. Both the computationally-designed capsid proteins described by Baker and colleagues[10] and the modified *Aquifex aeolicus* lumazine synthase (AaLS) proteins from Hilvert and colleagues[11] package RNA while self-assembling inside the *E. coli* cells. Packaging by the computationally-designed proteins, however, is based on non-specific electrostatic interactions and encapsulation is not limited to its own capsid RNA genome. About 25% of the encapsulated RNA originates from the host, and enrichment of the capsid RNA genome was related to its very high overexpression levels which simply increases the chance to randomly encapsulate the capsid genome. In contrast, the AaLS proteins were equipped with a cationic N+ peptide to mediate specific recognition of capsid RNA harbouring BoxBr packaging signals. Consequently, non-specific packaging of host RNA was limited to about 15% of the total amount of encapsulated RNA.

As compared to the complex lumazine synthase capsids, a specific advantage of our simple *de novo* designed artificial viral capsid proteins is that they are highly modular, with template binding being largely independent from self-assembly, as we have also shown in this chapter. Each individual block fulfils a desired function and can be adapted to the intended application, e.g. by introducing single amino

acid changes, by replacing an entire block for another one (Chapter 2 and 3), or by the fusion of additional functional blocks (Chapter 5).

Furthermore, for our proteins the presence of a (DNA or RNA) template itself, is a strong trigger for particle formation (Chapter 3). This is not the case for the lumazine synthase capsids, and has also not been shown to be the case for the computationally-designed capsid proteins described by Baker and colleagues[10]. Therefore, we would expect that with our design it should be possible to truly achieve encapsulation only for templates harbouring the correct packaging signals, and to achieve a higher degree of incorporation of "correct" templates, as compared to the relatively low efficiencies obtained by the Baker[10] and Hilvert[11] groups.

Here, we have not yet been able to show that this is indeed possible with $E^S{}_{80}$-$S^Q{}_{10}$-*B* proteins where *B* is a small sequence-specific binding domain. It should be possible, however, to obtain intact strepII-$E^S{}_{80}$-$S^Q{}_{10}$-ZnF protein by optimising the purification protocol, for example by including protease inhibitors and/or carrying out purification at low temperatures. Also, other sequence-specific binding blocks *B*, such as the AT-hooks and the bZIP leucine zipper described in the Appendix, may allow for the production of intact and functional proteins. In addition, it would be interesting to see if we can extend such specificity to RNA templates.

We have not yet demonstrated that assembly of our artificial viral capsid proteins into VLPs is also possible inside the complex environment of the *E. coli* cells. However, we are encouraged by the observation in this chapter that (encapsulated) bacterial DNA co-purifies with the strepII-$E^S{}_{80}$-$S^Q{}_{10}$-ZnF protein.

In conclusion, despite the fact that here we have not yet obtained this goal, we believe that the $E^S{}_{80}$-$S^Q{}_{10}$-*B* protein design, with suitable choices for the binding block *B*, should definitely allow for capsids that better mimic natural viruses in their ability to specifically encapsulate a genomic template (viz. their own genome). In addition to an increased understanding of viral assembly, this would also pave the way for more advanced directed evolution approaches, with important implications for the development of more efficient nucleic acid delivery vehicles.

Appendix

Other possible sequence-specific binding blocks

As an alternative to the zinc finger (ZnF) motif discussed in the main text of this chapter, two other small DNA binding motifs were identified that could serve as a sequence-specific DNA binding block in our artificial viral capsid protein: the basic region leucine zipper (bZIP) and the AT-hook.

The eukaryotic bZIP motif folds into an α-helix (see the red domain in Figure A4.1a) and consists of a charged basic region that makes specific contacts with nucleotides in the major groove of the DNA and a leucine zipper region for dimerisation of two bZIP-containing proteins[25]. The 56 amino acid long bZIP peptide (see Figure A4.1c for the amino acid sequence) from the yeast transcription factor GCN4 was shown to specifically bind as a dimer to the palindromic sequence 5'-ATGACGTCAT-3' with high affinity ($K_d \approx 50$ nM), and with reasonable affinity ($K_d \approx 1000$ nM) to random DNA sequences[26,27]. Finally, this small peptide has been produced before in *E. coli* indicating that this expression host can produce functional bZIP[27].

AT-hooks are very small DNA binding domains (of only a few amino acids each) found in eukaryotic proteins, and bind in the minor groove of AT-rich DNA (see the red domain in Figure A4.1b)[28,29]. A 10 amino acid long AT-hook (see Figure A4.1d for the amino acid sequence) found in the human HMGA1 proteins was shown to bind 5'-AATT-3' DNA sequences with reasonable affinity ($K_d \approx 6000$ nM)[30]. Although this is not as high as that of the ZnF and bZIP motifs, the AT-hook remains a very interesting candidate due to its small size (10 amino acids only) and its lack of secondary structures which will make it easier to produce a functional version of this DNA binding motif.

Materials & Methods

Only aspects specific to the AT and bZIP binding domains are discussed below, for other aspects, see main text.

Preparation of cloning vectors

For construction of a pET24a(+) PRe-RDL cloning vector encoding for the bZIP binding domain, two sets of oligonucleotides were designed (see Figure A4.1c) and ligated into the cloning vector using the two-step ligation as described for the ZnF domain (see main text). For the AT-hook, a single set of oligonucleotides was sufficient (see Figure A4.1d), which were first annealed and then ligated for 1 hour at room temperature into the linearised cloning vector. All ligation products were then transformed into chemically competent cells, screened and sequenced as described for the ZnF domain in the main text.

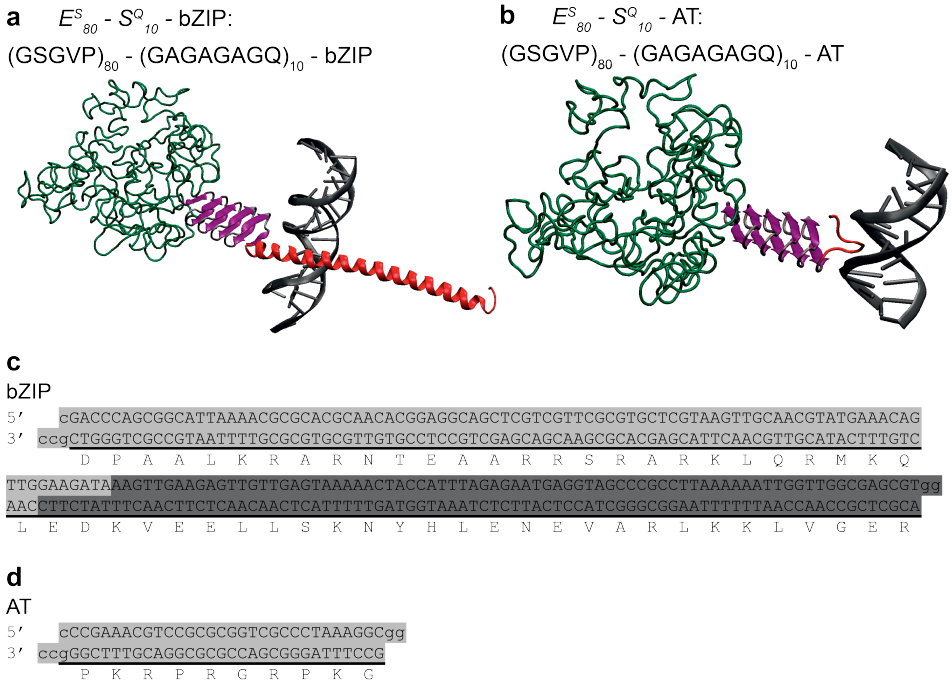


Figure A4.1 – Design of artificial viral capsid proteins with a sequence-specific DNA binding block. a) Cartoon of the hypothetical structure of the E^S_{80}-S^Q_{10}-bZIP protein when bound to DNA. In green the E^S_{80} random coil block with the N-terminal strepII-tag, in purple the S^Q_{10} silk-like block, and in red the DNA binding block. b) The hypothetical structure of E^S_{80}-S^Q_{10}-AT upon binding to the target DNA. Colours as in a). c) Oligonucleotides encoding the gene for the bZIP binding block. The first half of the gene is encoded by the set of oligonucleotide that are marked light, and the oligonucleotides encoding the second half of the gene are marked dark grey. d) Set of oligonucleotides encoding the AT binding block.

Gene construction

The full gene of strepII-E^S_{80}-S^Q_{10}-bZIP is not yet constructed. Construction of the strepII-E^S_{80}-S^Q_{10}-AT gene could not be done using the PRe-RDL cloning strategy that was used to construct strepII-E^S_{80}-S^Q_{10}-ZnF as we unintentionally designed the AT-hook oligonucleotides with a *BglI* site. A requirement for PRe-RDL is that the *BglI* site is unique in the cloning vectors which thus was not the case for this gene. Instead, we treated the strepII-E^S_{80}-S^Q_{10} cloning vector overnight at 37°C with 10 U each of *BseRI* and *AcuI*, and purified the strepII-E^S_{80}-S^Q_{10} fragment from an agarose gel. The cloning vector encoding the AT-hook was linearised overnight at 37°C with 5 U *BseRI*, dephosphorylated at the 5' ends using 1 U of FastAP for 30 min at 37°C and purified using a PCR purification kit. The strepII-E^S_{80}-S^Q_{10} fragment was then ligated into the linearised vector by T4 DNA ligase for 90 min at room temperature. Transformation into chemically competent cells, screening

and DNA sequencing was performed as described in the main text.

Results

Although initial steps were performed in the construction of an expression vector encoding the strepII-$E^S{}_{80}$-$S^Q{}_{10}$-bZIP protein by preparing a pET24a(+) PRe-RDL vector coding for the bZIP domain, we did not yet construct the full strepII-$E^S{}_{80}$-$S^Q{}_{10}$-bZIP gene. For strepII-$E^S{}_{80}$-$S^Q{}_{10}$-AT, on the other hand, an expression plasmid was constructed as described above by ligation of a DNA fragment coding for strepII-$E^S{}_{80}$-$S^Q{}_{10}$ immediately upstream of the AT-coding sequence in a pET24a(+) PRe-RDL vector. Unfortunately, we were not able yet to produce and purify the strepII-$E^S{}_{80}$-$S^Q{}_{10}$-AT protein successfully.

Discussion

The ZnF motif described in the main text of this chapter is a very interesting candidate binding block for our artificial viral capsid proteins because it is relatively small (51 amino acids), has a good affinity for its recognition sequences ($K_d \approx 400$ nM)[21,22], and on top of that it can be genetically engineered to recognise a chosen DNA sequence[14,31–34]. On the other hand, in order to function properly the ZnF domain should adopt a conformation that requires the incorporation of a zinc ion, and hence is slightly more complex and demanding than the two DNA binding candidates discussed in the current appendix.

Indeed, the relatively easy fold of the bZIP peptide (viz. an α-helix) likely facilitates refolding of a bZIP-carrying artificial viral capsid protein into a functional protein after storage as a freeze-dried powder. Issues may however occur at the level of protein self-assembly since the leucine zipper region within the bZIP peptide essentially is a dimerisation domain and may thus interfere with the proteins assembly behaviour instructed by the silk-inspired block. If such problems indeed occur, the design of this binding domain may possibly be adjusted by removing the leucine zipper region from the bZIP coding sequence, thereby leaving only the bZIP basic region that is responsible for the sequence-specific DNA binding.

Concerning the size and folding of the candidate binding blocks, the AT-hook motif is maybe the most interesting. With a length of about 10 amino acids, AT-hooks are about 5x smaller than the ZnF and bZIP motifs and are similar in length to the 12 amino acid counting non-specific binding block K_{12} used in current designs of the artificial viral capsid protein. Hence, it is very likely that they can be easily packed along the silk-like core of the virus-like particles. Additionally, the lack of secondary structures should avoid any issues caused by folding-errors. Despite these two big advantages, AT-hooks have a relatively low affinity for their recognition sequence ($K_d \approx 6000$ nM)[30] if compared to the ZnF and bZIP motifs, so that the typical protein concentrations used in our experiments (1.8 μM) may not be high enough to induce sequence-specific binding. On the other hand, the AT-hooks are so small that it may be possible to increase affinity of the AT binding block by incorporating two or more adjacent AT-hooks instead of only a single one.

Overall, each binding domain may have its own advantages and disadvantages. While we do not know their full implications on the ability of the artificial viral capsid proteins to bind and encapsulate nucleic acids, it would be interesting in future studies to not only produce and characterise strepII-$E^{S}{}_{80}$-$S^{Q}{}_{10}$-ZnF, but also strepII-$E^{S}{}_{80}$-$S^{Q}{}_{10}$-bZIP and strepII-$E^{S}{}_{80}$-$S^{Q}{}_{10}$-AT.

References

[1] H. Fujisawa and P. Hearing, "Structure, function and specificity of the DNA packaging signals in double-stranded DNA viruses," *Seminars in Virology*, vol. 5, no. 1, pp. 5–13, 1994.

[2] P. G. Stockley, R. Twarock, S. E. Bakker, A. M. Barker, A. Borodavka, E. Dykeman, R. J. Ford, A. R. Pearson, S. E. Phillips, N. A. Ranson, and R. Tuma, "Packaging signals in single-stranded RNA viruses: Nature's alternative to a purely electrostatic assembly mechanism," *Journal of Biological Physics*, vol. 39, no. 2, pp. 277–287, 2013.

[3] R. Twarock, R. J. Bingham, E. C. Dykeman, and P. G. Stockley, "A modelling paradigm for RNA virus assembly," *Current Opinion in Virology*, no. Figure 1, pp. 1–8, 2018.

[4] D. R. Turner, L. E. Joyce, and P. J. G. Butler, "The tobacco mosaic virus assembly origin RNA. Functional characteristics defined by directed mutagenesis," *Journal of Molecular Biology*, vol. 203, pp. 531–547, 1988.

[5] M. Marek, C. Romier, L. Galibert, O.-W. Merten, and M. M. van Oers, "Baculovirus VP1054 is an acquired cellular PURα, a nucleic acid-binding protein specific for GGN repeats.," *Journal of virology*, vol. 87, pp. 8465–8480, aug 2013.

[6] A. Oppenheim, Z. Sandalon, A. Peleg, O. Shaul, S. Nicolis, and S. Ottolenghi, "A cis-acting DNA signal for encapsidation of simian virus 40," *Journal of Virology*, vol. 66, no. 9, pp. 5320–5328, 1992.

[7] A. Gordon-Shaag, O. Ben-Nun-Shaul, V. Roitman, Y. Yosef, and A. Oppenheim, "Cellular Transcription Factor Sp1 Recruits Simian Virus 40 Capsid Proteins to the Viral Packaging Signal, ses," *Journal of Virology*, vol. 76, no. 12, pp. 5915–5924, 2002.

[8] B. Biswas, P. Kumari, and P. Vivekanandan, "Pac1 Signals of Human Herpesviruses Contain a Highly Conserved G-Quadruplex Motif," *ACS Infectious Diseases*, vol. 4, no. 5, pp. 744–751, 2018.

[9] S. Shakeel, E. C. Dykeman, S. J. White, A. Ora, J. J. Cockburn, S. J. Butcher, P. G. Stockley, and R. Twarock, "Genomic RNA folding mediates assembly of human parechovirus," *Nature Communications*, vol. 8, no. 1, 2017.

[10] G. L. Butterfield, M. J. Lajoie, H. H. Gustafson, D. L. Sellers, U. Nattermann, D. Ellis, J. B. Bale, S. Ke, G. H. Lenz, A. Yehdego, R. Ravichandran, S. H. Pun, N. P. King, and D. Baker, "Evolution of a designed protein assembly encapsulating its own RNA genome," *Nature*, vol. 552, no. 7685, pp. 415–420, 2017.

[11] N. Terasaka, Y. Azuma, and D. Hilvert, "Laboratory evolution of virus-like nucleocapsids from nonviral protein cages," *Proceedings of the National Academy of Sciences*, vol. 115, no. 21, pp. 5432–5437, 2018.

[12] A. Hernandez-Garcia, D. J. Kraft, A. F. J. Janssen, P. H. H. Bomans, N. A. J. M. Sommerdijk, D. M. E. Thies-Weesie, M. E. Favretto, R. Brock, F. a. de Wolf, M. W. T. Werten, P. van der Schoot, M. C. Stuart, and R. de Vries, "Design and self-assembly of simple coat proteins for artificial viruses," *Nature Nanotechnology*, vol. 9, pp. 698–702, aug 2014.

[13] M. W. T. Werten, W. H. Wisselink, T. J. Jansen-van den Bosch, E. C. de Bruin, and F. a. de Wolf, "Secreted production of a custom-designed, highly hydrophilic gelatin in Pichia pastoris," *Protein Engineering Design and Selection*, vol. 14, pp. 447–454, jun 2001.

[14] Y. Peng, K. J. Clark, J. M. Campbell, M. R. Panetta, Y. Guo, and S. C. Ekker, "Making designer mutants in model organisms," *Development*, vol. 141, pp. 4042–4054, oct 2014.

[15] S. A. Wolfe, L. Nekludova, and C. O. Pabo, "DNA recognition by Cys2His2 zinc finger proteins," *Annual Review of Biophysics and Biomolecular Structure*, vol. 29, pp. 183–212, 2000.

[16] J. Boch, H. Scholze, S. Schornack, A. Landgraf, S. Hahn, S. Kay, T. Lahaye, A. Nickstadt, and U. Bonas, "Breaking the code of DNA binding specificity of TAL-type III effectors.," *Science (New York, N.Y.)*, vol. 326, no. 5959, pp. 1509–1512, 2009.

[17] J. van der Oost, E. R. Westra, R. N. Jackson, and B. Wiedenheft, "Unravelling the structural and mechanistic basis of CRISPR-Cas systems," *Nature Reviews Microbiology*, vol. 12, no. 7, pp. 479–492, 2014.

[18] J. A. Doudna and E. Charpentier, "The new frontier of genome engineering with CRISPR-Cas9," *Science*, vol. 346, no. 6213, 2014.

[19] B. Zhao, M. A. Cohen Stuart, and C. K. Hall, "Dock n roll: folding of a silk-inspired polypeptide into an amyloid-like beta solenoid," *Soft Matter*, 2016.

[20] J. G. Omichinski, P. V. Pedone, G. Felsenfeld, a. M. Gronenborn, and G. M. Clore, "The solution structure of a specific GAGA factor-DNA complex reveals a modular binding mode.," *Nature structural biology*, vol. 4, no. 2, pp. 122–132, 1997.

[21] P. V. Pedone, R. Ghirlando, G. M. Clore, A. M. Gronenborn, G. Felsenfeld, and J. G. Omichinski, "The single Cys2-His2 zinc finger domain of the GAGA protein flanked by basic residues is sufficient for high-affinity specific DNA binding," *Proceedings of the National Academy of Sciences*, vol. 93, pp. 2822–2826, 1996.

[22] S. Negi, M. Dhanasekaran, T. Hirata, H. Urata, and Y. Sugiura, "Biomolecular mirror-image recognition: reciprocal chiral-specific DNA binding of synthetic enantiomers of zinc finger domain from GAGA factor," *Chirality*, vol. 18, no. 4, pp. 254–258, 2006.

[23] J. R. McDaniel, D. J. Callahan, and A. Chilkoti, "Drug delivery to solid tumors by elastin-like polypeptides.," *Advanced drug delivery reviews*, vol. 62, pp. 1456–1467, dec 2010.

[24] D. E. Meyer and A. Chilkoti, "Protein Purification by Inverse Transition Cycling," *ProteinProtein Interactions: AMolecular Cloning Manual*, vol. Chapter 18, pp. 329–344, 2002.

[25] T. E. Ellenberger, C. J. Brandl, K. Struhl, and S. C. Harrison, "The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted α Helices: Crystal structure of the protein-DNA complex," *Cell*, vol. 71, no. 7, pp. 1223–1237, 1992.

[26] Y. Okahata, K. Niikura, Y. Sugiura, M. Sawada, and T. Morii, "Kinetic Studies of Sequence-Specific Binding of GCN4-bZIP Peptides to DNA Strands Immobilized on a 27-MHz Quartz-Crystal Microbalance," *Biochemistry*, vol. 37, no. 16, pp. 5666–5672, 1998.

[27] J. J. Hollenbeck and M. G. Oakley, "GCN4 Binds with High Affinity to DNA Sequences Containing a Single Consensus Half-Site," *Biochemistry*, vol. 39, no. 21, pp. 6380–6389, 2000.

[28] L. Aravind and D. Landsman, "AT-hook motifs identified in a wide variety of DNA-binding proteins," *Nucleic Acids Research*, vol. 26, no. 19, pp. 4413–4421, 1998.

[29] E. Fonfría-Subirós, F. Acosta-Reyes, N. Saperas, J. Pous, J. a. Subirana, and J. L. Campos, "Crystal structure of a complex of DNA with one AT-hook of HMGA1," *PLoS ONE*, vol. 7, no. 5, pp. 1–5, 2012.

[30] A. I. Dragan, J. R. Liggins, C. Crane-Robinson, and P. L. Privalov, "The energetics of specific binding of AT-hooks from HMGA1 to target DNA," *Journal of Molecular Biology*, vol. 327, no. 2, pp. 393–411, 2003.

[31] J. R. Desjarlais and J. M. Berg, "Toward rules relating zinc finger protein sequences and DNA binding site preferences," *Proc. Nat. Acad. Sci. USA*, vol. 89, pp. 7345–7349, 1992.

[32] Y. Choo and A. Klug, "Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions," *Proceedings of the National Academy of Sciences*, vol. 91, no. 11, pp. 11168–11172, 1994.

[33] T. Sera and C. Uranga, "Rational design of artificial zinc-finger proteins using a nondegenerate recognition code table," *Biochemistry*, vol. 41, no. 22, pp. 7074–7081, 2002.

[34] S. A. Wolfe, R. A. Grant, M. Elrod-Erickson, and C. O. Pabo, "Beyond the Recognition Code : structures of two Cys2-His-2 zinc finger/TATA box complexes," *Structure*, vol. 9, pp. 717–723, 2001.

5

Uptake and processing of mRNA encapsulated by artificial viral capsid proteins in HeLa cells

One of the first requirements for successful transfection by non-viral vectors is efficient cellular uptake. Here, we investigate this first crucial step for rod-shaped virus-like particles composed of the C-S^{Q}_{10}-K_{12} *artificial viral capsid protein and a luciferase-encoding mRNA. We also investigate whether cellular uptake can be enhanced by chemically conjugating an RGD cell-binding peptide to the unique N-terminal cysteine on the hydrophilic* C*-block of the* C-S^{Q}_{10}-K_{12} *protein. We established that with 100% RGD-labelled proteins the assembly into VLPs is compromised, but that proper VLP assembly is restored by mixing-in sufficient amounts of unlabelled protein. Next, uptake of fluorescently-labelled VLPs by HeLa cells was studied using confocal laser scanning microscopy. The unique N-terminal cysteine of the proteins was used to fluorescently label a fraction of the proteins in the VLPs, while the mRNA that was used was also fluorescently labelled. The presence of bright fluorescent spots (both protein and mRNA) inside the cells suggests the particles enter HeLa cells via endosomes. The overall fluorescence due to particles inside the cells increases somewhat with an increasing fraction of RGD-labeled protein, but the effect is small, presumably due to shielding of the RGD-peptides by the long and flexible* C *blocks of the proteins. Expression of the luciferase encoded by the mRNA could not be detected, suggesting that the bottleneck to successful transfection is poor endosomal escape or unsuccessful unpacking of the VLPs, but probably not cellular uptake.*

Part of the data in this chapter was obtained in collaboration with:
O. Paulino da Silva Filho and R. Brock (Radboud University Nijmegen)

Introduction

In recent decades, the field of gene therapy has made great strides in moving towards clinically approved therapies. For viral vectors, the first clinically approved therapies have come on the market and many clinical trials are ongoing[1]. Viral vectors still face risks and problems such as in their safe and efficient production[2], their inherent immunogenicity and the risk of insertional mutagenesis[3,4]. Therefore, there has also been continued interest in the further development of non-viral vectors for nucleic acid delivery[5,6].

Recently, mRNA delivery has come to the fore as a method to express therapeutically active proteins in cells in a range of applications such as protein replacement, vaccination and CRISPR-Cas9-mediated gene editing[5–8]. In view of the need for repeated administration and the possible problems of viral delivery vectors with immune reactions, mRNA delivery would benefit from the development of safe and effective non-viral delivery agents.

Good *in vitro* and *in vivo* transfection efficiencies of mRNA, are especially reported for lipid nanoparticles[9]. Development of such efficient nanoparticles, however, may be complex as it requires the precise optimisation of their components, which typically includes cationic lipids, phospholipids and cholesterol[10]. In addition, chemical modification of the lipid nanoparticles with polymers like PEG is in general required to reduce their clearance and immunogenicity[6]. Polycations such as polyethylenimine (PEI) and polylysine and related polymers have also been explored extensively. These suffer from similar rapid clearance issues and additionally from a relatively high toxicity[11–13].

There is sustained interest in the development of novel classes of non-viral delivery agents that can address the above limitations of lipid nanoparticles and polycations. The artificial viral capsid proteins discussed in this thesis could be such a new class of non-viral delivery agents, which is intermediate between polymers and natural virus capsids. Indeed, the triblock protein $C\text{-}S^{Q}{}_{10}\text{-}K_{12}$ that was developed earlier[14] has been shown to not only encapsulate DNA, but also mRNA[15]. The mRNA was shown to be encapsulated into rod-shaped virus-like particles (VLPs) and to be protected against degradation by RNases. The VLPs are non-toxic to HeLa cells and red blood cells, but transfection could not be detected and only a weak transfection efficiency was found for HEK293T cells[15].

Successful protein production from mRNA, only occurs if each of the many steps involved in delivery proceeds successfully. At the cellular level, some of the key steps are uptake of the delivery agents by the cells, escape from the endosomes in which they usually end up, and unpacking from the delivery vehicle[5,16,17]. Only if intact mRNA ends up in the cytoplasm, it will be transcribed into protein and can have its intended therapeutic effect.

Coming back to mRNA encapsulated into rod-shaped VLPs by the $C\text{-}S^{Q}{}_{10}\text{-}K_{12}$ protein, its low transfection efficiency may be caused by a bottleneck in any of the above steps. Here we address the first basic step, that of entry and uptake of the VLPs by cells, at the *in vitro* level. Due to the lack of positively charged amino acids at the surface of the VLPs, it might very well be that they are not taken up by cells, for which their membranes are negatively charged, as efficiently as some cationic

lipid- and polymer-based vehicles. This problem might be resolved by the addition of cell-targeting peptides that facilitate binding of the particles to cell membranes and subsequent entry[16].

In this chapter we therefore explore the cellular uptake and processing in HeLa cells of mRNA encapsulated by *C*-S^Q_{10}-K_{12} proteins into rod-shaped VLPs. We compare unmodified VLPs to VLPs modified with cell-binding RGD peptides[18,19]. The latter are chemically coupled to the N-terminal cysteine of the *C*-S^Q_{10}-K_{12} proteins, adjacent to the *C* block. In order to track the fate of the mRNA and protein components of the VLPs once inside the cells, we use fluorescently labelled *C*-S^Q_{10}-K_{12} proteins as well as fluorescently labelled mRNA. We find that VLPs are taken up by HeLa cells, presumably in endosomes, but that uptake is only somewhat enhanced by the presence of the RGD peptides.

Materials & Methods

Materials

The *C*-S^Q_{10}-K_{12} protein was produced and purified previously as described by Hernandez-Garcia *et al.*[14]. CleanCap Cyanine 5 (Cy5) Firefly Luciferase (FLuc) mRNA of 1921nt was purchased from TriLink Biotechnologies. With a 5' Cap 1 structure and a 3' poly-A-tail, the mRNA mimics a fully processed mature mRNA. The mRNA is labelled with the Cy5 dye, which is coupled to approximately 25% of the uridines in this mRNA. pUC18 DNA was obtained from Thermo Fisher Scientific. Also the Alexa Fluor 594 C_5-maleimide dye was purchased from Thermo Fisher Scientific, while the RGD-maleimide peptide with the sequence GRGDSPGK(Maleimide), was synthesised by PepScan. Reducing agents Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) and DL-Dithiothreitol (DTT) were purchased from Sigma-Aldrich. The 10% mini-protean TGX precast protein gels, Precision Plus Protein All Blue Prestained Protein Standard and Laemmli running buffer were obtained from Bio-Rad. The PageBlue protein staining solution was bought from Thermo Fisher Scientific, same as the Slide-A-Lyzer dialysis cassettes with a molecular weight cutoff of either 3.5 kDa or 10 kDa and a capacity of 0.5-3 mL. The Amicon Ultra-0.5 centrifugal filters with a cutoff of 3 kDa, and the Ultrafree-MC PVDF centrifugal filters with pore size 0.22 μm were purchased from Merck. RPMI 1640 and DMEM cell culture media with and without phenol red, Glutamax I, fetal calf serum (FCS), Opti-MEM medium, and HEPES buffered saline (HBS) were all obtained from Gibco, Thermo Fisher Scientific. Also Lipofectamine Messenger Max was purchased from Thermo Fischer Scientific. Trypsin in EDTA was from PAN Biotech, HeLa and HEK293T cells were obtained from the German collection of microorganisms and cell cultures (DSMZ). SKOV3 and HFL1 cells were purchased from the American type culture collection. μ-slide (chambered coverslip) 8-well plates for confocal laser scanning microscopy were purchased from Ibidi. For immunocytochemistry, bovine serum albumin (BSA) was purchased from Sigma-Aldrich. Anti-$\beta 1$ (anti-human, mouse 4B4 clone), anti-αv (anti-human, mouse 17E6 clone) and anti-$\beta 3$ (anti human, Y2/51 clone),

and isotype-matched non-specific mouse or rat IgGs were acquired from Beckman Coulter and secondary goat anti-mouse antibody conjugated to Alexa Fluor 488, purchased from Thermo Fischer Scientific.

Conjugation of cell-binding peptide and fluorophore to the protein via thiol-maleimide coupling

The cell-targeting peptide RGD and the Alexa Fluor 594 dye were conjugated to the N-terminal unique cysteine of the *C*-S^{Q}_{10}-K_{12} protein via thiol-maleimide coupling. For coupling of the RGD peptide, the *C*-S^{Q}_{10}-K_{12} protein was solubilised in 50 mM sodium phosphate buffer (PB; pH7.4) at a concentration of 50 μM by heating to 65°C for 10 min while vortexing. A 10x molar excess of the reducing agent DTT or TCEP was added to the protein and incubated for 1 hour at RT to reduce any disulphide bridges formed by the cysteines. In the case of DTT, the reducing agent was removed after incubation by using 3 kDa centrifugal filters. Removal of TCEP was not necessary as this compound does not reduce thiol-maleimide bonds. Next, a 20x molar excess of the RGD peptide in DMSO was added dropwise, while stirring the protein under argon atmosphere. The reaction was allowed to proceed for 2 hours at room temperature. After the reaction, the protein was purified from the excess peptide and, if applicable, the TCEP either by using 3 kDa centrifugal filters, or by using 3.5 kDa Slide-A-Lyzer dialysis cassettes in MilliQ water. Finally, the protein was lyophilised and stored at room temperature.

Conjugation of the Alexa Fluor 594 (AF594) dye was performed as described above using TCEP as the reducing agent. The AF594-modified *C*-S^{Q}_{10}-K_{12} was purified by using a 3.5 kDa Slide-A-Lyzer dialysis cassettes in MilliQ water and lyophilised. Size exclusion chromatography however showed that not all unbound dye was removed. Therefore, the freeze-dried product ($\sim$3.13 mg) was solubilised in 2.0 mL MilliQ by vortexing. The solution was dialysed against MilliQ water using a 10 kDa Slide-A-Lyzer dialysis cassette for a total of 45 hours at 4°C. The dialysis buffer was refreshed 4x during these hours. Unbound dye was further removed by 3 kDa centrifugal filters for 13x. The remaining product was lyophilised and stored at room temperature.

Characterisation of modified *C*-S^{Q}_{10}-K_{12}

The modified *C*-S^{Q}_{10}-K_{12} proteins were analysed by size exclusion chromatography (SEC) to evaluate purity and for AF594 also successful conjugation. The freeze-dried proteins were solubilised in 50 mM PB with 150 mM NaCl (pH7.4) by heating to 65°C and simultaneously vortexing for 15 min, and filtered through a 0.22 μm centrifugal filter. The AF594-conjugated protein was freeze-dried twice with an intervening step in which the unbound dye was further removed by dialysis. For SEC, we used the initial freeze-dried product that was not yet further dialysed. The resulting solutions had a protein concentration of 20 μM. Per protein solution, 100 μL was loaded onto a Superdex 75 10/300 GL column (GE Healthcare Life Sciences) equilibrated in 50 mM PB with 150 mM NaCl (pH7.4). The used flow rate was 0.5 mL/min. Protein peptide bonds were detected at a wavelength of

214 nm and the Alexa Fluor 594 dye was detected at 588 nm which is the maximum excitation wavelength for this dye.

In addition to SEC, SDS-PAGE analysis was performed to evaluate protein purity and for AF594 also successful conjugation. After electrophoresis, the dye was visualised in unstained gels using ultraviolet illumination. Next, proteins were stained with PageBlue and imaged with white light. Finally, conjugation of the RGD peptide was confirmed by matrix-assisted laser desorption ionisation time-of-flight mass spectrometry (MALDI-TOF MS). The protocols used for SDS-PAGE and MALDI-TOF MS are described in more detail in Chapter 3 of this thesis.

Preparation of VLPs

For experiments to optimise VLP formulations, stock solutions of 100 μM RGD-*C*-$S^Q{}_{10}$-K_{12} and 50 μM unmodified *C*-$S^Q{}_{10}$-K_{12} were prepared by solubilising freeze-dried protein in 10 mM PB (pH7.4) by heating to 65°C for 10 min while vortexing. RGD-*C*-$S^Q{}_{10}$-K_{12} and unmodified *C*-$S^Q{}_{10}$-K_{12} were mixed at various ratios, containing 100, 33.3, 9.1 or 4.8% RGD-*C*-$S^Q{}_{10}$-K_{12}. These protein mixtures were then mixed with pUC18 DNA at an N/P ratio of 7 in 10 mM PB. The final (total) protein and DNA concentrations per sample were always 1.8 μM and 1 ng/μL respectively. As a control, 100 μM RGD-*C*-$S^Q{}_{10}$-K_{12} was diluted to 1.8 μM in 10 mM PB in absence of DNA. All sample conditions were incubated for 24 hours at room temperature.

For cell entry experiments, stock solutions of 1 mg/mL ($\sim$22 μM) of RGD-*C*-$S^Q{}_{10}$-K_{12}, AF594-*C*-$S^Q{}_{10}$-K_{12} and unmodified *C*-$S^Q{}_{10}$-K_{12} were prepared in 10 mM PB (pH7.5) supplemented with 0.1 mM DTT (VLP buffer). DTT was added to reduce any disulphide bridges formed between the cysteines of unmodified *C*-$S^Q{}_{10}$-K_{12}. The freeze-dried proteins were solubilised by heating to 65°C for 30 min while vortexing. Modified proteins were mixed with unmodified protein at the three different ratios to form the protein formulations listed in Table 5.1. The protein mixtures were then split into two equal aliquots and supplemented with either Cy5-labeled FLuc mRNA, or with VLP buffer only. Proteins and mRNA were mixed at an N/P ratio of 6, with a final (total) protein concentration of 14.14 μM and a final mRNA concentration of 10 ng/μL. Finally, to form VLPs all mixtures were incubated for 24 hours at room temperature.

Table 5.1 – Protein formulations used for cell entry experiments.

	RGD-*C*-$S^Q{}_{10}$-K_{12} (%)*	**AF594-*C*-$S^Q{}_{10}$-K_{12} (%)***	**Unmodified *C*-$S^Q{}_{10}$-K_{12} (%)***
VLP	0	9.1	90.9
RGD-VLP	9.1	9.1	81.8
6xRGD-VLP	54.5	9.1	36.4

* Percentage modified *C*-$S^Q{}_{10}$-K_{12} of the total amount of protein within a sample.

Preparation of Lipofectamine-mRNA control

Immediately prior to transfection experiments, a positive control was prepared for the transfection of Cy5-labelled mRNA using Lipofectamine Messenger Max according to manufacturer's protocol. In addition, it serves as a negative control for the AF594 signal since it does not contain any AF594-*C*-$S^{Q}{}_{10}$-K_{12}. First, a Lipofectamine solution was prepared by diluting 0.3 μL of Lipofectamine in 10 μL of Optim-MEM medium, and incubated for 10 min at room temperature. Meanwhile, the Cy5-labeled FLuc mRNA stock solution was diluted 50 times in Opti-MEM medium to obtain an mRNA concentration of 20 ng/μL. Lipofectamine and mRNA were then mixed in equal volumes and incubated for 5 min to allow complexes to form. The final mRNA concentration in the mRNA-Lipofectamine solution was 10 ng/μL.

Atomic force microscopy

After 24 hours of incubation, samples were prepared for AFM imaging and analysed as described in Chapter 2. Lengths of the VLPs were measured using FiberApp as also described in Chapter 2, with the exception that particles below 50 nm were excluded from the calculations of the number and weight averaged lengths.

Cell culture

HeLa and HFL1 cells were maintained in full cell culture medium, composed of RPMI 1640 medium supplemented with Glutamax I and 10% FCS whereas HEK293T and SKOV3 cells were maintained in (high glucose) DMEM containing 10% FCS. They were grown at 37°C in a T-75 culture flask in a humidified incubator with 7.5% CO_2. Every two to three days, cells were passaged when a confluency of ~80% was reached. To passage the cells, $^{8}/_{10}{}^{\text{th}}$ of the culture medium was removed and cells were washed with 5 mL of HBS, followed by detachment of the cells by incubating with 0.05% of trypsin in EDTA for 5 min. Cells were then spun down at 150 x *g* for 3 min and the cell pellet was resuspended in 10 mL of full cell culture medium. At last, $^{1}/_{3}{}^{\text{rd}}$ of the resuspended cells was diluted four times in full cell culture medium and plated in a new T-75 culture flask.

Immunocytochemistry for integrin expression

Integrin expression was evaluated in four different cell lines (HeLa, HEK293T, HFL1 and SKOV3) by immunocytochemistry. The following mouse monoclonal IgG antibodies were used to bind to the human αv, $\beta 1$ and $\beta 3$ integrin subunits: anti-integrin αv (17E6), anti-integrin $\beta 1$ (4B4) and anti-integrin $\beta 3$ (Y2/51). For detection, a goat anti-mouse IgG secondary antibody conjugated to an Alexa Fluor 488 dye was used in combination with flow cytometry. To prepare the cells for immunocytochemistry, the following protocol was used. First, cells were detached from culture flasks using 1x phosphate buffered saline pH7.4 (1x PBS) with 5 mM EDTA for 20-30 min. The cell suspension was then neutralised by adding an equal volume of 1x PBS containing 10% FCS and centrifuged at 300 x *g* for 5 min.

The supernatant was discarded and the pelleted cells were resuspended in cold 1x PBS supplemented with 3% BSA to yield a concentration of $1*10^6$ cells/mL. The resuspended cells were distributed over polypropylene tubes to prepare 50 μL aliquots. To each aliquot, an equal volume was added of a 20 μg/mL solution of one of the mouse anti-integrin antibody, and incubated for 60 min at 4°C on a roller mixer. To remove any unbound antibodies, cells were pelleted for 5 min at 300 x *g*, washed 3x with 500 μL of cold 1x PBS with 3% BSA, and pelleted once more at 300 x *g* for 5 min. The cell pellet was resuspended in 50 μL of 1x PBS with 3% BSA which was supplemented with a 1:100 dilution of the Alexa Fluor 488-conjugated goat anti-mouse IgG secondary antibody. Cells were incubated with the secondary antibody in the dark for 60 min at 4°C on a roller mixer, followed by two rounds of pelleting the cells at 300 x *g* for 5 min and washing with cold 1x PBS. The final pellet was resuspended in 100 μL of cold 1x PBS. Finally, the fluorescence was measured by flow cytometry using a FACSCalibur (BD Biosciences). Cells were excited at 488 nm using a 15 mW argon laser, and emitted green fluorescence was collected by a photomultiplier tube with a 530/30 nm bandpass filter. The data was then analysed using FlowJo software to determine the relative mean fluorescence of each sample.

Confocal laser scanning microscopy of labelled mRNA and proteins

One day before the cell entry experiment, 40000 HeLa cells per well were seeded on 8-well μ-slides. They were maintained in full cell culture medium and incubated at 37°C in a humidified incubator with 7.5% CO_2. To start the experiment, cells were washed and incubated in full cell culture medium with naked mRNA, mRNA-Lipofectamine complexes, or nanoparticles with or without mRNA. For each condition, 20 μL of the sample with or without 10 ng/μL mRNA was mixed with 180 μL of full cell culture medium prior to adding it to the cells. Hence, the amount of mRNA added per well was 0 or 200 ng. After 2 hours, cells were washed and full cell culture medium without phenol red but supplemented with 20 mM HEPES was added. Confocal laser scanning microscopy was then performed using a Leica SP5 microscope equipped with an HCX PL APO 63x N.A. 1.2 water immersion lens from Leica. During image acquisition, cells were kept at 37°C. Alexa Fluor 594 and Cy5 dyes were excited using HeNe 594 and HeNe 633 lasers respectively. For Alexa Fluor 594, emission was collected between 620 and 654 nm, and for Cy5 between 707 and 786 nm. After imaging, fresh full cell culture medium was added to the cells and cells were kept in culture. After 48 hours, full cell culture medium without phenol red and with 20 mM HEPES was added to the cells again and imaging was repeated as described above.

Quantitative analysis of confocal images was conducted with Fiji[20]. For each image, average fluorescence intensities of the Alexa Fluor 594 and Cy5 dyes were calculated. To do this, images were first smoothed and duplicated. One copy of the image was turned into a binary mask of pixels above background, with value = 0 for background pixels and value = 1 for any pixel above background (a valid pixel). Using the option “Image Calculator ” in Fiji, the binary image was multiplied with

the original image, resulting in a final image that only contains information about the intensity of valid pixels as the intensity of the background pixels are all set to zero. Next, the sum of the intensities of all valid pixels was obtained from this final image, and the total number of valid pixels was obtained from the binary image. Finally, the average intensity of a dye in an image was calculated by dividing the sum of the valid pixel intensities by the total number of valid pixels.

Results

Assembly of virus-like particles containing modified *C-S*$^{Q}_{10}$-*K*$_{12}$ proteins

First, we describe the chemical modifications of the *C-S*$^{Q}_{10}$-*K*$_{12}$ proteins that we use to probe and possibly improve cellular uptake of encapsulated mRNA. The question that we address here is to what extent these chemical modifications affect the formation of VLPs. An integrin-binding RGD peptide with the sequence GRGDSPGK was coupled to the unique N-terminal cysteine of the *C-S*$^{Q}_{10}$-*K*$_{12}$ protein (for the amino acid sequence of this protein, see ref[14]). Bioconjugation was achieved via thiol-maleimide coupling, using a maleimide group attached to the lysine of the peptide. The same coupling chemistry was used to fluorescently label the proteins via the attachment of a maleimide derivative of Alexa Fluor 594 (AF594).

Reaction products were analysed using size exclusion chromatography (SEC), SDS-PAGE and MALDI-TOF MS (Figure 5.1). Peaks observed in the MALDI-TOF MS are presented in Table 5.2 and compared to the expected molecular weights. MALDI-TOF MS spectra for the unmodified *C-S*$^{Q}_{10}$-*K*$_{12}$ and the RGD-modified *C-S*$^{Q}_{10}$-*K*$_{12}$ (Figure 5.1e) demonstrate that conjugation of the RGD peptide was successful. SEC and SDS-PAGE for the RGD-modified and unmodified *C-S*$^{Q}_{10}$-*K*$_{12}$ (Figure 5.1a,b) are not significantly different, as expected given the small mass of the peptide. These results also demonstrate that dialysis has successfully removed the excess free RGD peptides.

No signal could be obtained for the AF594-modified *C-S*$^{Q}_{10}$-*K*$_{12}$ in MALDI-TOF MS, but both the SEC and the SDS-PAGE clearly demonstrates that labelling was successful. However, both the SEC and the SDS-PAGE also demonstrate that, unfortunately, despite extensive dialysis of the reaction products, we have not been able to remove all of the excess AF594 dye.

For AF594-labelling of the VLPs, only a small fraction of the incorporated proteins needs to be labelled in order to obtain a sufficient signal. Therefore, we do not expect a major influence on protein co-assembly into VLPs. The required amount of RGD labelling to achieve optimal cell-binding of the VLPs however, is unknown. For this case we therefore consider the entire range of VLPs composed of 0-100% RGD-*C-S*$^{Q}_{10}$-*K*$_{12}$, and study the impact of the RGD modification on co-assembly.

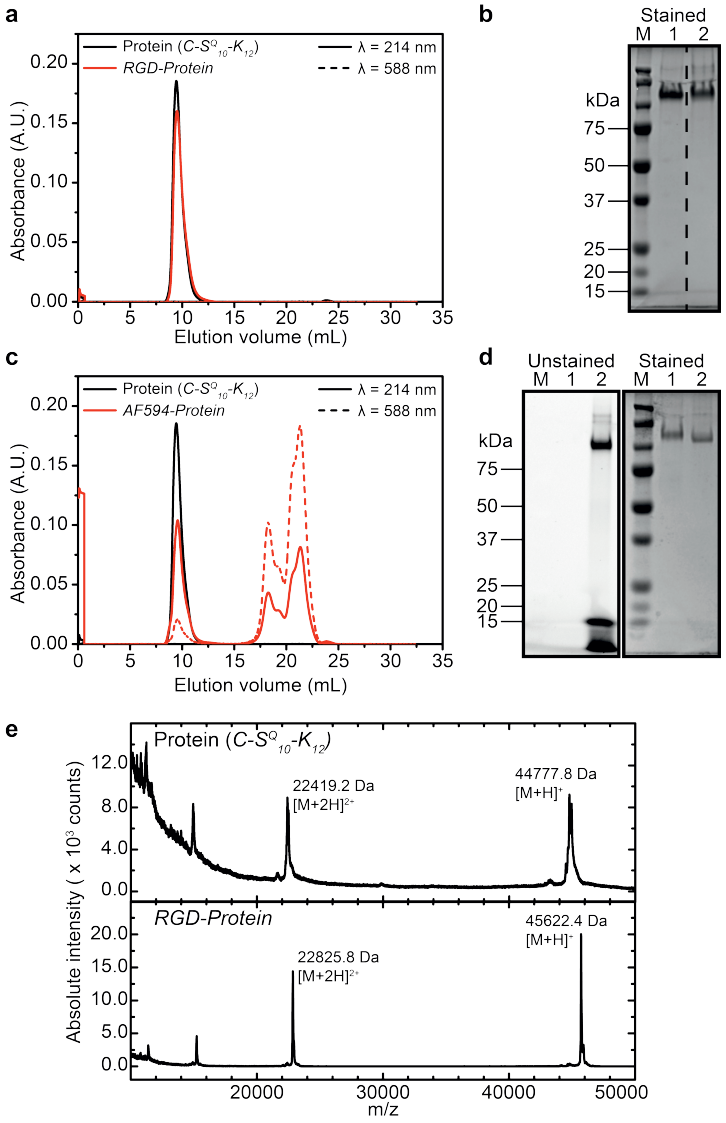


Figure 5.1 – Chemical modifications of the artificial viral capsid protein C-S^Q_{10}-K_{12} and characterisation of the reaction products. a) SEC chromatograms for C-S^Q_{10}-K_{12} and RGD-C-S^Q_{10}-K_{12}. Protein UV absorption signal is measured at 214 nm. b) SDS-PAGE of C-S^Q_{10}-K_{12} (lane 1) and RGD-C-S^Q_{10}-K_{12} (lane 2). Marker lane is labelled M. As explained elsewhere[14,21], C-S^Q_{10}-K_{12} proteins move at anomalously low mobilities due to low SDS binding of C block. c) SEC chromatograms for C-S^Q_{10}-K_{12} and AF594-C-S^Q_{10}-K_{12}. Protein UV absorption signal is measured at 214 nm and AF594 fluorophore UV absorption signal is measured at 588 nm. d) SDS-PAGE of C-S^Q_{10}-K_{12} (lane 1) and AF594-C-S^Q_{10}-K_{12} (lane 2). Marker lane is labelled M. Left gel: unstained and UV illuminated (AF595 signal), right gel: PageBlue-stained and white light illuminated (protein signal). e) MALDI-TOF MS spectra of C-S^Q_{10}-K_{12} and RGD-C-S^Q_{10}-K_{12} .

Table 5.2 – Molar masses of the conjugated proteins deduced from peak positions in MALDI-TOF MS spectra of Figure 5.1e.

	Expected MW (Da)	Experimental MW (Da)	Mass deviation (Da)
$C\text{-}S^{Q}{}_{10}\text{-}K_{12}$	44749.25	44778.0	28.8
RGD-$C\text{-}S^{Q}{}_{10}\text{-}K_{12}$	45601.10	45622.4	21.3
AF594-$C\text{-}S^{Q}{}_{10}\text{-}K_{12}$	45658.22	*	*

* Data could not be determined.

To do so we use pUC18 plasmid DNA as a template, since particle formation with $C\text{-}S^{Q}{}_{10}\text{-}K_{12}$ has previously been especially well characterised for DNA rather than for mRNA. Mixed RGD-$C\text{-}S^{Q}{}_{10}\text{-}K_{12}$ / $C\text{-}S^{Q}{}_{10}\text{-}K_{12}$ protein solutions (labelling percentages of 4.8%, 9.1%, 33.3% and 100%; total protein concentration 1.8 μM) were incubated for 24h in the absence or presence of pUC18 DNA at an N/P ratio of 7, and the resulting fibril structures were imaged using AFM. From the AFM images, length distributions of fibrils were obtained. Both the AFM images and the resulting fibril length distributions are shown in Figure 5.2. Values for the average fibril length and the spread in the fibril length distributions are presented in Table 5.3.

For pure RGD-$C\text{-}S^{Q}{}_{10}\text{-}K_{12}$ (i.e. 100% RGD), we find distinctly different self-assembly behaviour as compared to that reported before for the unmodified protein $C\text{-}S^{Q}{}_{10}\text{-}K_{12}$[14]. At a concentration of 1.8 μM, the pure RGD-$C\text{-}S^{Q}{}_{10}\text{-}K_{12}$ only assemble into fibrils in the presence of the pUC18 DNA (compare the AFM images in Figure 5.2a and b). The length of these fibrils is approximately 120 nm (Table 5.3), which is much shorter than expected: from previous work we know the unmodified $C\text{-}S^{Q}{}_{10}\text{-}K_{12}$ condenses dsDNA a factor of $\sim$3 in length such that VLPs with a length of about 300 nm would have been expected[14]. Higher resolution AFM images (Figure 5.2b, right image) reveal details that point to explanations for the unexpectedly short VLPs. It appears that not all DNA is encapsulated into fibrils. Non-encapsulated DNA can be seen to connect and/or interrupt fibrils. This suggests that for some reason the labelling with the RGD peptide prevents fibril formation to go to completion. Alternatively, the RGD labelling may give rise to a much higher nucleation frequency with multiple nuclei arising even on relatively short templates such as the pUC18 DNA.

Next we investigated whether the expected VLP assembly behaviour is restored at lower fractions of RGD-labelled protein (Figure 5.2c-e). This does indeed seem to be the case, with mixtures of 33.3% of RGD-labelled proteins already giving VLP fibril length distributions that peak at a much larger value of 230 nm (also see Table 5.3), and AFM images showing uninterrupted straight fibrils. Values of VLP lengths around 230 nm are still somewhat smaller than the calculated 300 nm assuming a packing factor $\sim$3. This may be due to the circularly supercoiled nature of the pUC18 plasmid DNA we use here as compared to the linear DNA used in previous experiments.

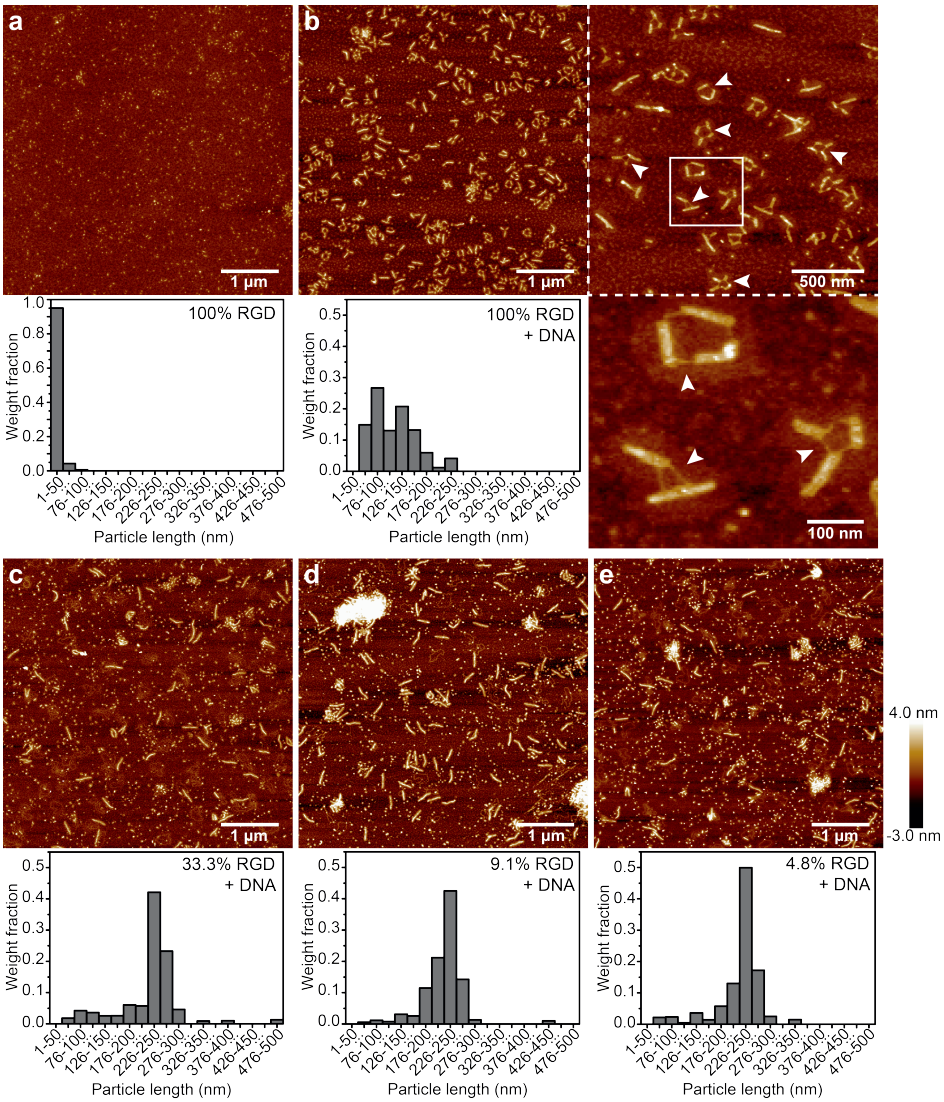


Figure 5.2 – Assembly of VLPs with varying fractions of RGD-labelled *C*-S^{Q}_{10}-K_{12} and pUC18 plasmid DNA at N/P = 7. AFM images and fibril length distributions derived from them (weight fraction of protein incorporated into fibrils versus fibril length). a) pure RGD-*C*-S^{Q}_{10}-K_{12}, no DNA. b) pure (100%) RGD-*C*-S^{Q}_{10}-K_{12}, with DNA. Right panel is high resolution zoom, part of which is still further magnified digitally (below) to highlight non-encapsulated DNA. c) 33.3% RGD-*C*-S^{Q}_{10}-K_{12}, d) 9.1% RGD-*C*-S^{Q}_{10}-K_{12}. e) 4.8% RGD-*C*-S^{Q}_{10}-K_{12}.

Table 5.3 – Weight averaged lengths of fibrils formed by mixed RGD-*C*-$S^Q{}_{10}$-K_{12}/*C*-$S^Q{}_{10}$-K_{12}protein solutions (total protein concentration is 1.8 μM) incubated for 24h in the absence or presence of pUC18 DNA at an N/P ratio of 7.

	N	$<L>_w$ (nm)	σ (nm) **
100% RGD	340	39	7
*100% RGD + DNA**	326	121	40
*33.3% RGD + DNA **	184	228	72
*9.1% RGD + DNA **	207	224	48
*4.8% RGD + DNA **	110	226	61

* Particles smaller than 50nm were excluded. ** The standard deviation σ is calculated over the length of all particles in each condition and is a measure for the width of the fibril length distribution.

Having concluded that RGD-modified *C*-$S^Q{}_{10}$-K_{12} proteins form VLPs as expected, provided that the degree of labelling with RGD peptides is not too large, we continue by assembling VLPs with mRNA to be used for the cell entry experiments. Successful encapsulation and protection of mRNA by the *C*-$S^Q{}_{10}$-K_{12} proteins was shown previously[15]. Here we check whether at the labelling conditions used for the cell entry experiments the VLPs have indeed assembled as expected.

VLPs were formed at an N/P ratio of 6, using 10 ng/μL mRNA. Table 5.1 summarises the composition of the different VLPs that were formed. Degrees of labelling with RGD peptides used were 0% (*VLP*), 9.1% RGD-labelled protein (*RGD-VLP*), and 54.5% RGD-labelled protein (*6xRGD-VLP*). Samples were prepared both with *(+)* and without *(-)* mRNA. The degree of protein labelling with the Alexa Fluor 594 dye (AF594), was always 9.1%. Finally, the mRNA that was used was a 1921nt-long luciferase reporter mRNA, fluorescently labelled with Cyanine 5 (Cy5).

AFM imaging confirms that VLPs form properly for all the conditions we intend to use in the VLP uptake experiments. As all conditions (*VLP*, *RGD-VLP* and *6xRGD-VLP*) show similar results, Figure 5.3 only shows the data for *RGD-VLP(-)* and *RGD-VLP(+)*. Data for the other conditions can be found in the Appendix, and is summarised in Table 5.4.

In the absence of mRNA (*RGD-VLP(-)*, Figure 5.3a; *VLP(-)* and *6xRGD-VLP(-)*, Figure A5.1a,c in the Appendix), at a protein concentration of 14.14 μM, the proteins do not form fibrils: only small and mostly globular aggregates are observed. At the same protein concentrations but in presence of the mRNA, we observe short fibrils (*RGD-VLP(+),* Figure 5.3b; *VLP(+)* and *6xRGD-VLP(+)*, Figure A5.1b,d in the Appendix). Higher resolution images of mRNA-containing VLPs show that in addition to a majority of straight rod-shaped VLPs, there are also occasional thinner branched structures (see white arrowhead in Figure 5.3b), presumably indicating a small fraction of non-encapsulated mRNA.

The lengths and heights of all particles were measured from the AFM images to obtain histograms of the weight fraction of protein incorporated into fibrils at a given length and height. These are shown in Figures 5.3c,d and in Figure A5.1e-h

in the Appendix.

The length distributions of the mRNA VLPs show a clear peak for lengths of 76-100 nm (Figure 5.3c, bottom panel) on top of the length distribution expected for proteins only (Figure 5.3c, top panel). This background distribution of protein-only fibrils is expected, since we use a 6-fold excess of protein with respect to mRNA. Average lengths for the whole distribution are reported in Table 5.4. Since most fibrils do not contain mRNA and are much shorter than the mRNA-containing fibrils, the peak location is in fact a better estimator for the length of mRNA-containing VLPs than the overall averages reported in Table 5.4.

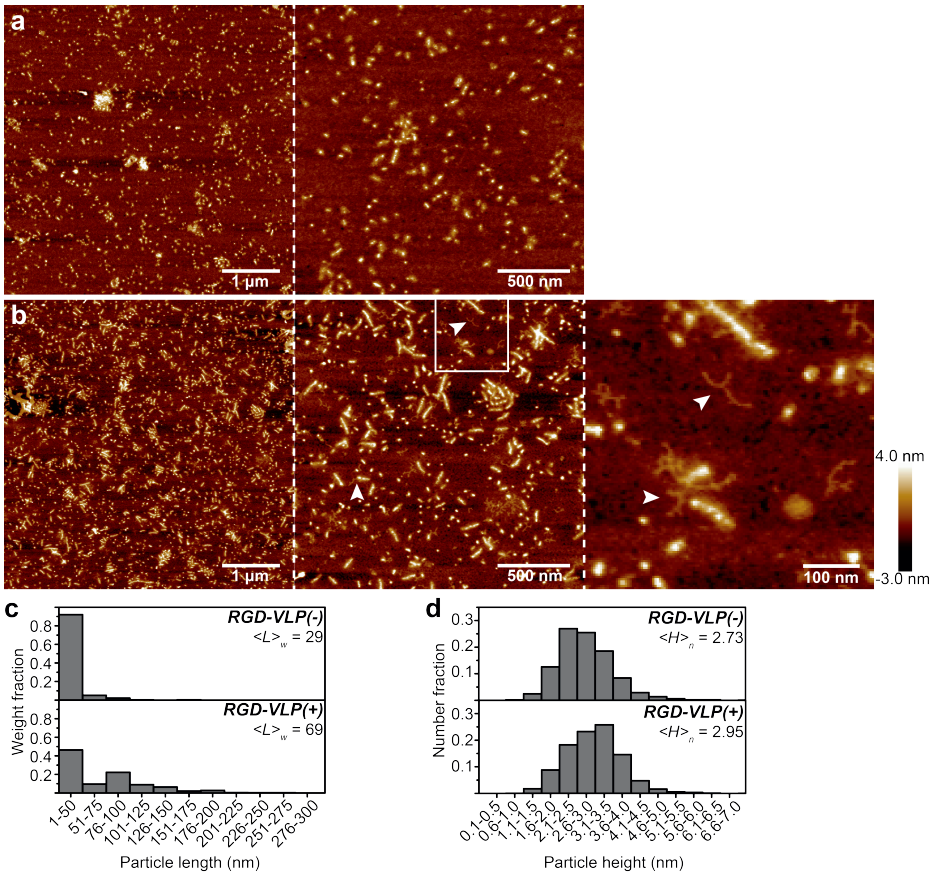


Figure 5.3 – AFM imaging of mRNA-VLPs used to study cellular uptake. a) *RGD-VLP*(-): 9.1% RGD-*C*-S^{Q}_{10}-K_{12} + 9.1% AF594-*C*-S^{Q}_{10}-K_{12} + 81.8% *C*-S^{Q}_{10}-K_{12}. b) *RGD-VLP*(+): 9.1% RGD-*C*-S^{Q}_{10}-K_{12} + 9.1% AF594-*C*-S^{Q}_{10}-K_{12} + 81.8% *C*-S^{Q}_{10}-K_{12} + mRNA (N/P6). c) Length distribution of fibrils (weight fraction of protein incorporated in fibrils of a given length) in a) and b). d) Height distribution of fibrils (number fraction of fibrils with a given length) in a) and b).

Table 5.4 – Weight averaged lengths $<L>_w$ and number averaged heights $<H>_n$ of protein and protein-mRNA fibrils.

	N	$<L>_w$ (nm)	σ (nm) *	$<H>_n$ (nm)	σ (nm) *
VLP(-)	2042	35	12	2.78	0.72
VLP(+)	3667	71	33	3.11	0.70
RGD-VLP(-)	1569	29	9	2.73	0.77
RGD-VLP(+)	3421	69	32	2.95	0.76
6xRGD-VLP(-)	1402	31	11	2.26	0.77
6xRGD-VLP(+)	2189	69	32	2.66	0.81

* The standard deviation σ is calculated over the length or height of all particles in each condition and is a measure for the spread in the fibril length/height distributions.

The mRNA has a length of 1921nt such that its contour length should be around 650 nm (assuming a contour length of 0.34 nm per nt). Previously it was established that for double-stranded linear DNA the VLP length was approximately 1/3rd of the DNA contour length, whereas for single-stranded linear DNA it was about 1/6th of the contour length[14]. Assuming the same packing factor of 6 for the single-stranded mRNA, we would expect VLP lengths of $\sim$100 nm, in agreement with the peak at 76-100 nm in the length distribution for the mRNA VLPs. This argues for encapsulation of a single mRNA molecule in each VLP, which is different from the earlier study on mRNA VLPs which used a shorter mRNA of 996nt, and for which it was found that most likely multiple mRNAs were encapsulated in a single VLP[15].

We have also measured height distributions for the fibrils and established that the average fibril height is significantly higher in presence of mRNA (Table 5.4), which is another indication for mRNA encapsulation.

Uptake of mRNA-VLPs by HeLa cells

RGD peptides are known to bind to integrins expressed on the cell membrane of many cell types[22,23]. The RGD peptide used here (GRGDSPGK) is expected to have affinity for $\alpha 5\beta 1$ integrins (composed of the α-subunit $\alpha 5$ and β-subunit $\beta 1$) and to a lesser extent for $\alpha v\beta 3$ and $\alpha IIb\beta 3$ integrins[18]. Immunocytochemistry was used to quantify αv, $\beta 1$ and $\beta 3$ subunit expression for a set of model human cells lines: HEK293T (embryonic kidney), HeLa (cervical cancer), HFL1 (lung fibroblast) and SKOV3 (ovarian cancer). Results are shown in Figure 5.4. All cell lines express a significant amount of the integrin $\beta 1$ subunit, so that we assume that they then also express the $\alpha 5\beta 1$ integrins that most strongly bind to our RGD peptides. Since HeLa cells are widely used in transfection experiments and as they were also used in a previous study by us, HeLa cells were here chosen to study entry and uptake of the mRNA VLPs.

HeLa cells were incubated with the mRNA VLPs and various control samples (i.e. naked mRNA and Lipofectamine-mRNA complexes). After 2 h of incubation, cells were washed to remove any unbound VLPs, and incubated again for another

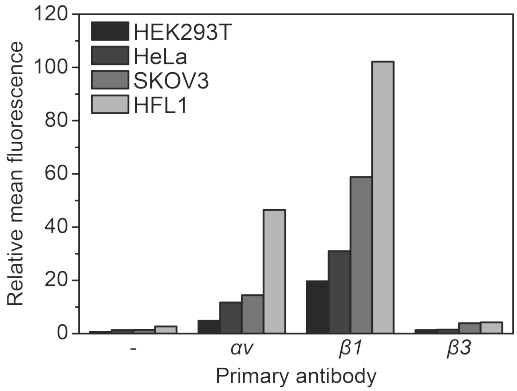


Figure 5.4 – Integrin expression in various cell lines. Immunocytochemistry to quantify the expression of integrin subunits αv, $\beta 1$ and $\beta 3$ for four human cell lines: HEK293T, HeLa, SKOV3 and HFL1. Primary antibodies are anti-integrin αv (17E6), anti-integrin $\beta 1$ (4B4) and anti-integrin $\beta 3$ (Y2/51). For detection with flow cytometry, a goat anti-mouse IgG secondary antibody conjugated to an Alexa Fluor 488 dye was used. The negative control is indicated with "-", where cells were only incubated with the secondary antibody without prior incubation with a primary antibody. All values are relative to untreated cells for which the mean fluorescence was set to 1.

2 h. Confocal microscopy was used to visualise cellular uptake by imaging either the AF594 dyes on the proteins or the Cy5 dyes on the mRNA (Figure 5.5). Lipofectamine-mRNA complexes were used as a negative control for the AF594 signal, and at the same time as a positive control for the Cy5 signal. We find that for the mRNA-containing VLPs, both the AF594 and Cy5 signal is mostly concentrated in vesicle-like structures, presumably endosomes (Figure 5.5, *VLP(+)*, *RGD-VLP(+)* and *6xRGD-VLP(+)*; also see Figure 5.6 for enlarged images of *6xRGD-VLP(+)*). In addition, we find a weak and non-homogeneous signal for both dyes throughout the cells, which likely indicates that VLPs and protein-only fibrils have adhered to the cell membranes but did not get internalised yet. It should be reminded that part of the AF594 signal may also be due to the fraction of free dye.

Nevertheless, we do also observe vesicle-like structures inside the cells for all *VLP* conditions, indicating that at least some VLPs were internalised via an endocytic pathway. For the *VLP* and *RGD-VLP* conditions, the number of vesicle-like structures seems similar. For the *6xRGD-VLPs* their number appears to be larger, although this has not yet been confirmed quantitatively. A higher number of vesicles could possibly indicate that for that case, the larger number of RGD peptides on the surface of the VLPs has promoted cell entry via endocytosis.

Finally, for all conditions we find large structures with very high AF594 (protein) signals. These appear to be structures that reside outside the cells. They might be large aggregates of mRNA VLPs and protein-only fibrils, possibly formed during incubation of the samples with the cells.

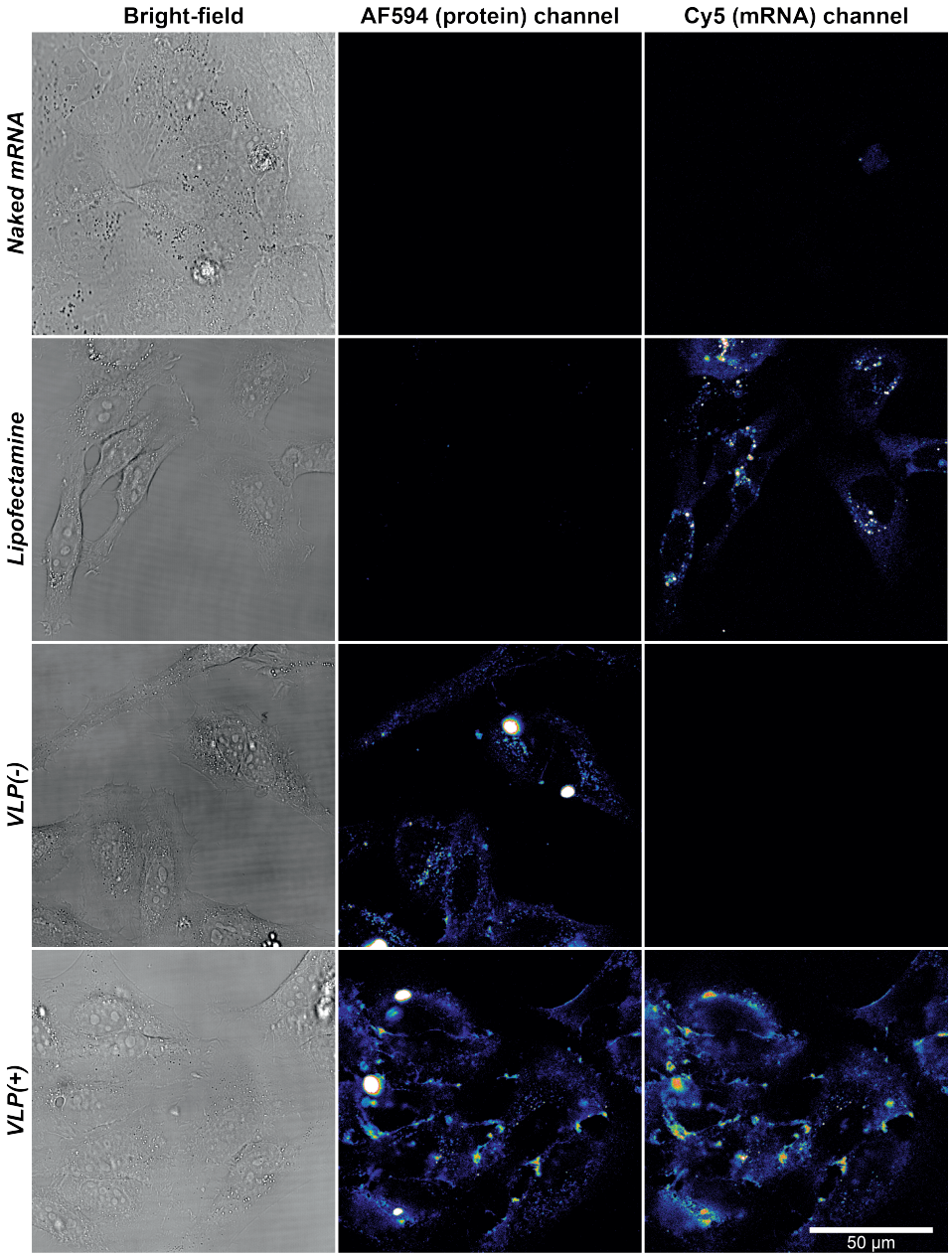
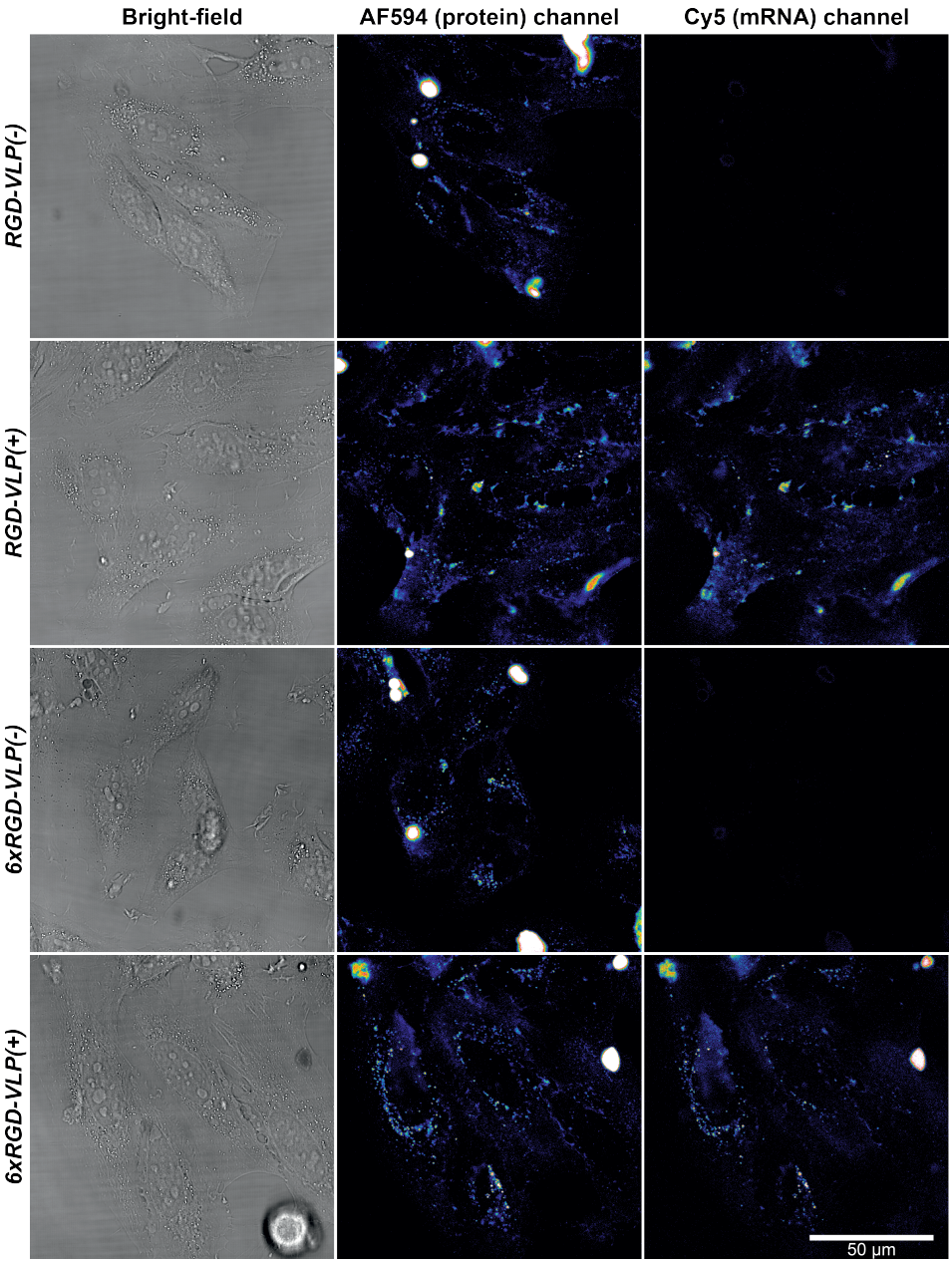


Figure 5.5 – VLP uptake by HeLa cells. HeLa cells were incubated for 2h with the indicated condition. Next, cell were washed and incubated for another 2h, so that confocal images were taken 4h post-transfection. Left column is bright-field image, middle column is the AF595 (protein) signal, right column is the Cy5 (mRNA) signal. Scale bar is 50 μm and is the same for all images. This figure continues on the right page.


Bright-field
AF594 (protein) channel
Cy5 (mRNA) channel
RGD-VLP(-)
RGD-VLP(+)
6xRGD-VLP(-)
6xRGD-VLP(+)
50 µm

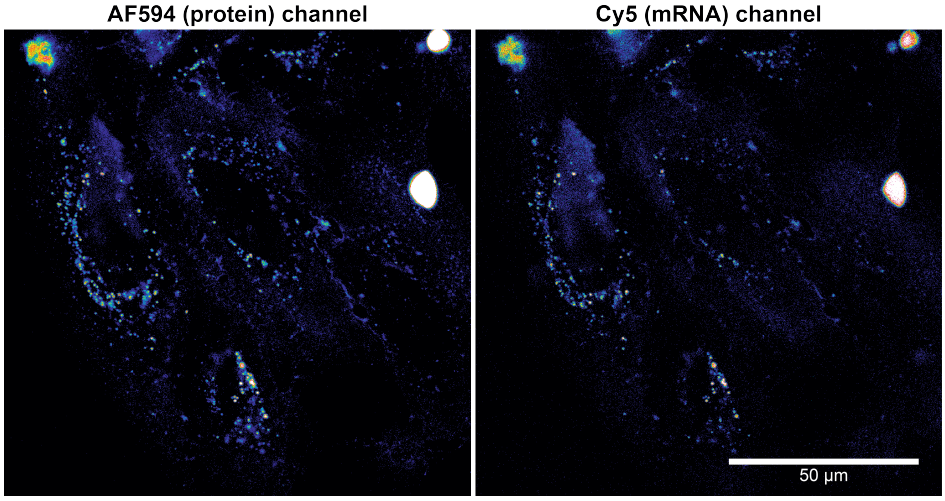


Figure 5.6 – *6xRGD-VLP(+)* uptake by HeLa cells. Confocal images of *6xRGD-VLP(+)* (identical to the images in Figure 5.5) are enlarged in order to better observe the presumed endosomes. Scale bar is 50 μm and is the same for all images.

Quantification of time-dependent VLP uptake and transfection

Average fluorescence intensities of the AF594 and Cy5 dyes were calculated for confocal fluorescence images of the HeLa cells 4h and 48h after the start of transfection, only excluding the large structures with very high AF594 signal which presumably are aggregates located on the outside of the cells. Results are shown in Figure 5.7.

First, we focus on the 4h time-point for which the fluorescent signals purely represent cell binding and uptake of the particles and are not yet affected by degradation. Results for the AF594 (protein) signal (Figure 5.7a) are in agreement with our visual observation that high degrees of RGD-labelling lead to higher uptake of both the mRNA VLPs and the short protein-only fibrils, but the effect is not very strong. Surprisingly, we do not see a similar trend for the Cy5 (mRNA) signal for which the signals are approximately equal between the mRNA-containing VLPs at all labelling degrees (Figure 5.7b). We however note that for the Lipofectamine-mRNA control the signal is unexpectedly low, and on the other hand, some of the mRNA(-) samples have an unexpectedly high Cy5 signal. Possibly, bleed-through from the AF594 dye to the Cy5 channel occurred so that we unfortunately cannot draw any significant conclusions from the Cy5 data.

Returning to the AF594 data, we now focus on the 48h time-point (Figure 5.7a, right graph). A sustained fluorescent signal is observed for the mRNA VLPs and protein-only fibrils with the highest degree of RGD-labelling (*6xRGD-VLP(+)* and *(-)*), but not with the lower degrees of labelling (*VLP* and *RGD-VLP* conditions). Upon endosomal uptake it may be expected that mRNA VLPs and protein-only fibrils are digested within 48h, thus leading to reduced AF594 (protein) signals.

The sustained signal for the *6xRGD-VLP* conditions however, suggests that these particles were not degraded. As we do not expect particles with higher amounts of RGD-labelling to be more stable from degradation, a possible explanation could be that a significant amount of mRNA VLPs and protein-only fibrils is in not (yet) internalised and instead is still bound to the cell membrane.

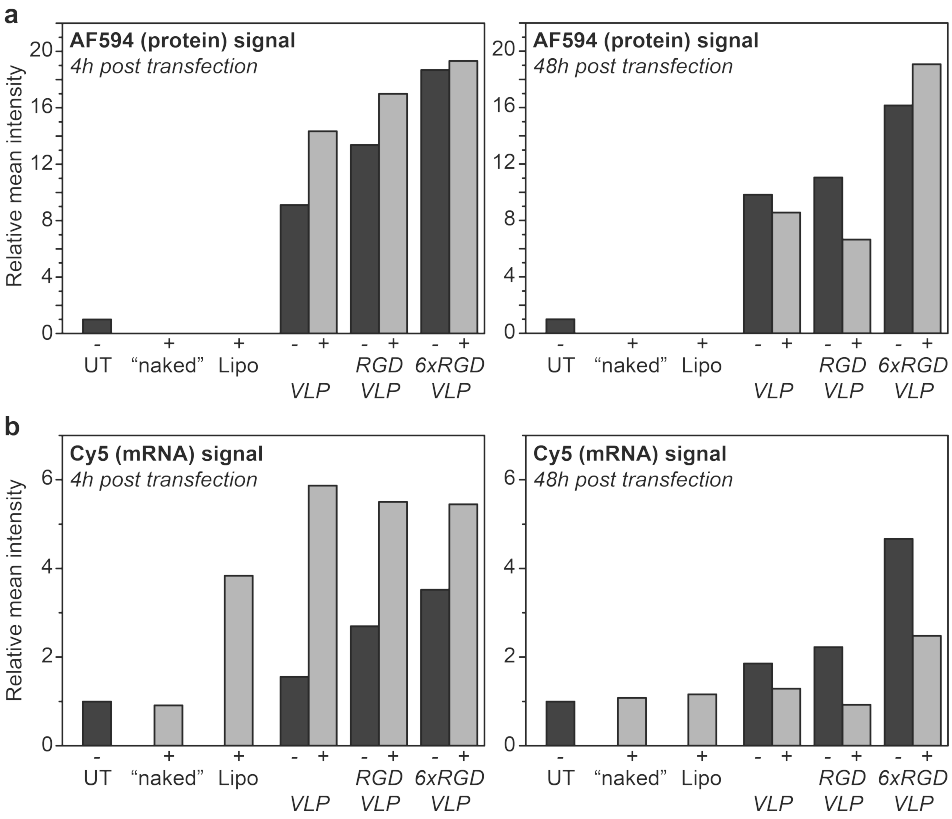


Figure 5.7 – Quantification of VLP uptake. Relative mean (fluorescence) intensities of the AF594 (protein) signal and the Cy5 (mRNA) signal, 4h (left graphs) and 48h (right graphs) post-transfection. Samples are indicated in the figures, with UT = untreated cells, "naked" = negative control with naked mRNA, "Lipo" = control with Lipofectamine-mRNA complexes, *VLP/RGD-VLP/6xRGD-VLP* = the samples, and "-" and "+" indicating the absence or presence of mRNA. a) AF594 protein signal, b) Cy5 mRNA signal. Fluorescence intensities are relative to those of the untreated cells (UT) for which the value is set to 1.

Finally, previously we have shown[15] that with HEK293T cells, mRNA VLPs without RGD labelling resulted in only a small degree of transfection (as reflected in the expression of a reporter gene), prompting us to investigate here whether this might be caused by a bottleneck in the cellular uptake of the VLPs. The mRNA used in the present study also encodes for a reporter gene, viz. firefly luciferase (FLuc). To establish whether mRNA VLPs with or without RGD labelling lead to transfection

of HeLa cells, we therefore finally carried out a luciferase assay at days 2 to 5 after transfection. Whereas the Lipofectamine-mRNA positive control showed the expected strong luminescence signal, none of the mRNA VLPs led to a luminescence signal much stronger than the background.

Discussion

We have here equipped the *C* block of the *C*-$S^Q{}_{10}$-K_{12} protein with an integrin-binding RGD peptide via its unique C-terminal cysteine. Although uptake did seem to increase with increasing labelling degrees of the proteins with the RGD peptide, the effect was not very large (Figures 5.5 and 5.7). Possibly, the large amorphous *C* block partly shields the peptide from binding to integrins on the cell membrane, as suggested in another study where these peptides were connected to long flexible linkers[24]. The same study concludes that long, stiff linkers result in improved binding of the RGD peptides to integrins. Therefore, in further work it might be advantageous to project the RGD peptide further outwards of the corona around the VLPs, for example by attaching it onto a long and stiff linker.

Most nano-sized delivery vehicles are internalised via endosomes, and our mRNA VLPs appear to be no exception. Although both the AF594-labelled proteins and the Cy5-labelled mRNA appear to be internalised in vesicle-like structures that are presumably endosomes, no expression was observed in a luciferase assay for the HeLa cells. Based on our results we have no reason to believe that this is due to poor uptake, but co-localisation studies with for example endosomal markers may provide some further insights in the entry pathway.

Poor endosomal escape is a frequently observed bottleneck for non-viral vectors, and this may also be the case for the VLPs. Presumably, the absence of luciferase expression was not the result of mRNA degradation, since previous work has shown that DNA and RNA templates are quite well protected from nuclease activity when having been encapsulated by the *C*-$S^Q{}_{10}$-K_{12} proteins[14,15]. This thus further increases the probability that poor endosomal escape is the bottleneck for our VLPs.

Various strategies have been proposed for enhancing endosomal escape of non-viral vectors and other nanomedicines[25–27], which might also be applied to the mRNA VLPs. For example, the *C*-$S^Q{}_{10}$-K_{12} proteins could be modified by (genetically) attaching peptides that disrupt the endosomal membrane upon acidification[28]. Furthermore, as opposed to silk domains containing glutamines as turn residues (S^Q = GAGAGAGQ), we might employ silks with histidines as turn residues (S^H = GAGAGAGH). At the low pH values occurring in the endosomes, this might promote both disassembly of the VLPs and give rise to the so-called proton sponge effect[29,30], both of which might promote the escape of intact mRNA to the cytoplasm as required for successful transfection. Indeed, before it has been shown that triblock polypeptides containing S^H domains are soluble at low pH and self-assemble into fibrils at neutral pH[31–33].

Besides the potential design modification that we could make to improve transfection efficiency of the *C*-$S^Q{}_{10}$-K_{12} VLPs, we should also address some experimental

issues in future experiments, in particular the presence of free (uncoupled) AF594 dye and the high amount of protein-only fibrils. Both may result in false-positive observations for the entry of mRNA-containing VLPs into cells and should therefore be removed prior to the transfection experiments, e.g. by using size-exclusion chromatography.

A final reason for not yet giving up on the possibility to develop the artificial viral capsid proteins as delivery vehicles is that it is one of the few systems that have a precisely controlled rod-shaped nanoarchitecture. Large differences exist between delivery *in vivo*, and *in vitro* in 2D cell cultures. Recently it was shown that a key property influencing the ability of nanomedicines to penetrate deeply into three dimensional tissue[34,35] is the shape of the nanoparticles, with rod-shaped particles penetrating deeper into tissues than spherical particles[36,37].

Appendix

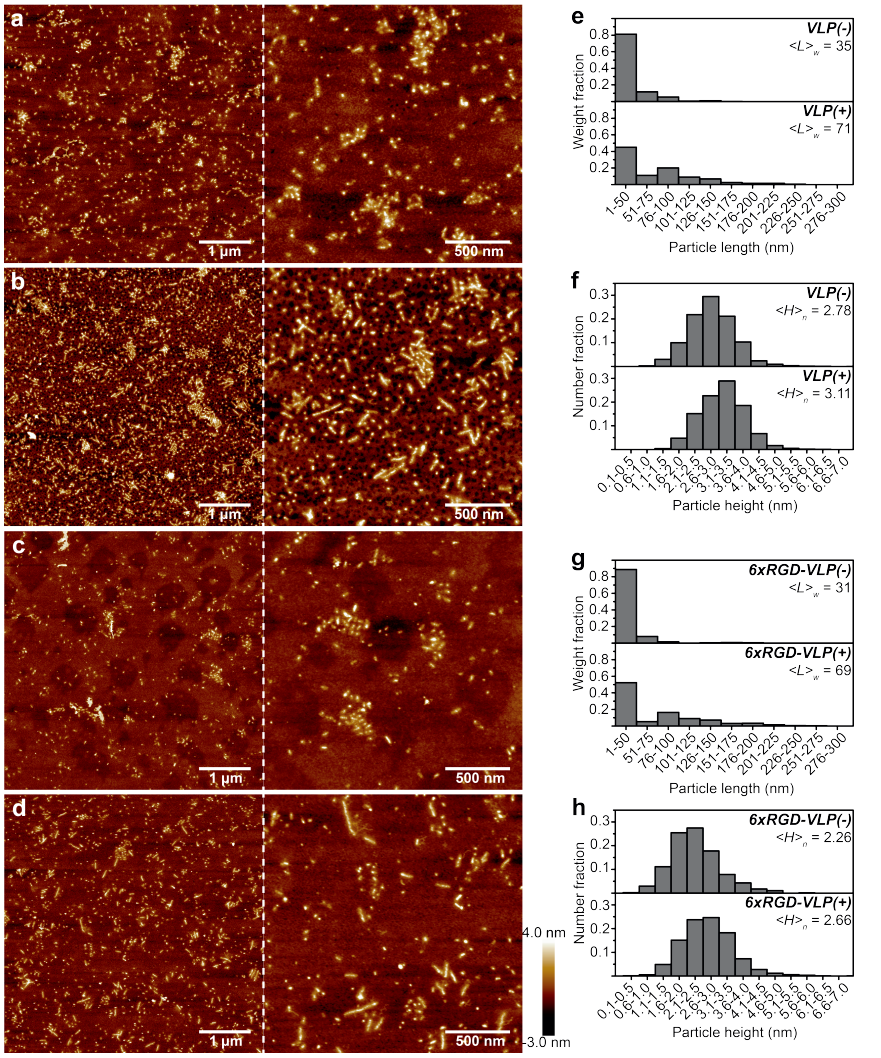


Figure A5.1 – Additional AFM imaging of mRNA VLPs used to study cellular uptake. a) *VLP*(-): 9.1% AF594-*C*-S^{Q}_{10}-K_{12} + 90.9% *C*-S^{Q}_{10}-K_{12}. b) *VLP*(+): 9.1% AF594-*C*-S^{Q}_{10}-K_{12} + 90.9% *C*-S^{Q}_{10}-K_{12} + mRNA (N/P6). c) 6x*RGD-VLP*(-): 54.6% RGD-*C*-S^{Q}_{10}-K_{12} + 9.1% AF594-*C*-S^{Q}_{10}-K_{12} + 36.3% *C*-S^{Q}_{10}-K_{12}. d) 6x*RGD-VLP*(+): 54.6% RGD-*C*-S^{Q}_{10}-K_{12} + 9.1% AF594-*C*-S^{Q}_{10}-K_{12} + 36.3% *C*-S^{Q}_{10}-K_{12} + mRNA (N/P6). e) Fibril length distribution (weight fraction of protein incorporated in fibrils of a given length) for particles in a) and b). f) Height distribution (number fraction of fibrils with a given height) for particles in a) and b). g) Fibril length distribution (weight fraction of protein incorporated in fibrils of a given length) for particles in c) and d). h) Height distribution (number fraction of fibrils with a given height) for particles in c) and d).

References

[1] S. L. Ginn, A. K. Amaya, I. E. Alexander, M. Edelstein, and M. R. Abedi, "Gene Therapy Clinical Trials Worldwide To 2017 - an Update," *The Journal of Gene Medicine*, no. January, p. e3015, 2018.

[2] J. C. M. van der Loo and J. F. Wright, "Progress and challenges in viral vector manufacturing," *Human molecular genetics*, vol. 25, no. R1, pp. R42–R52, 2016.

[3] C. E. Thomas, A. Ehrhardt, and M. A. Kay, "Progress and problems with the use of viral vectors for gene therapy," *Nature Reviews Genetics*, vol. 4, no. 5, pp. 346–358, 2003.

[4] D. Bouard, N. Alazard-Dany, and F. L. Cosset, "Viral vectors: From virology to transgene expression," *British Journal of Pharmacology*, vol. 157, no. 2, pp. 153–165, 2009.

[5] H. Yin, R. L. Kanasty, A. A. Eltoukhy, A. J. Vegas, J. R. Dorkin, and D. G. Anderson, "Non-viral vectors for gene-based therapy," *Nature Reviews Genetics*, vol. 15, no. 8, pp. 541–555, 2014.

[6] K. A. Hajj and K. A. Whitehead, "Tools for translation: Non-viral materials for therapeutic mRNA delivery," *Nature Reviews Materials*, vol. 2, pp. 1–17, 2017.

[7] N. Pardi, M. J. Hogan, F. W. Porter, and D. Weissman, "mRNA vaccines a new era in vaccinology," *Nature Reviews Drug Discovery*, vol. 17, no. 4, pp. 261–279, 2018.

[8] W. Y. Wu, J. H. G. Lebbink, R. Kanaar, N. Geijsen, and J. van der Oost, "Genome editing by natural and engineered CRISPR-associated nucleases," *Nature Chemical Biology*, vol. 14, no. 7, pp. 642–651, 2018.

[9] J. B. Miller, S. Zhang, P. Kos, H. Xiong, K. Zhou, S. S. Perelman, H. Zhu, and D. J. Siegwart, "Non-Viral CRISPR/Cas Gene Editing In Vitro and In Vivo Enabled by Synthetic Nanoparticle Co-Delivery of Cas9 mRNA and sgRNA," *Angewandte Chemie - International Edition*, vol. 56, no. 4, pp. 1059–1063, 2017.

[10] K. J. Kauffman, J. R. Dorkin, J. H. Yang, M. W. Heartlein, F. Derosa, F. F. Mir, O. S. Fenton, and D. G. Anderson, "Optimization of Lipid Nanoparticle Formulations for mRNA Delivery in Vivo with Fractional Factorial and Definitive Screening Designs," *Nano Letters*, vol. 15, no. 11, pp. 7300–7306, 2015.

[11] H. Lv, S. Zhang, B. Wang, S. Cui, and J. Yan, "Toxicity of cationic lipids and cationic polymers in gene delivery," *Journal of Controlled Release*, vol. 114, no. 1, pp. 100–109, 2006.

[12] F. Alexis, E. Pridgen, L. K. Molnar, and O. C. Farokhzad, "Factors affecting the clearance and biodistribution of polymeric nanoparticles," *Molecular Pharmaceutics*, vol. 5, no. 4, pp. 505–515, 2008.

[13] E. Moreau, M. Domurado, P. Chapon, M. Vert, and D. Domurado, "Biocompatibility of polycations: In vitro agglutination and lysis of red blood cells and in vivo toxicity," *Journal of Drug Targeting*, vol. 10, no. 2, pp. 161–173, 2002.

[14] A. Hernandez-Garcia, D. J. Kraft, A. F. J. Janssen, P. H. H. Bomans, N. A. J. M. Sommerdijk, D. M. E. Thies-Weesie, M. E. Favretto, R. Brock, F. a. de Wolf, M. W. T. Werten, P. van der Schoot, M. C. Stuart, and R. de Vries, "Design and self-assembly of simple coat proteins for artificial viruses," *Nature Nanotechnology*, vol. 9, pp. 698–702, aug 2014.

[15] S. Jekhmane, R. de Haas, O. Paulino da Silva Filho, A. H. van Asbeck, M. E. Favretto, A. Hernandez Garcia, R. Brock, and R. de Vries, "Virus-Like Particles of mRNA with Artificial Minimal Coat Proteins: Particle Formation, Stability, and Transfection Efficiency," *Nucleic Acid Therapeutics*, vol. 27, no. 3, pp. 159–167, 2017.

[16] E. Mastrobattista, M. A. E. M. van der Aa, W. E. Hennink, and D. J. A. Crommelin, "Artificial viruses: a nanotechnological approach to gene delivery.," *Nature reviews. Drug discovery*, vol. 5, pp. 115–121, 2006.

[17] U. Lächelt and E. Wagner, "Nucleic Acid Therapeutics Using Polyplexes : A Journey of 50 Years (and Beyond)," *Chemical Reviews*, vol. 115, no. 19, pp. 11043–11078, 2015.

[18] U. Hersel, C. Dahmen, and H. Kessler, "RGD modified polymers: Biomaterials for stimulated cell adhesion and beyond," *Biomaterials*, vol. 24, no. 24, pp. 4385–4415, 2003.

[19] J. Park, K. Singha, S. Son, J. Kim, R. Namgung, C. O. Yun, and W. J. Kim, "A review of RGD-functionalized nonviral gene delivery vectors for cancer therapy," *Cancer Gene Therapy*, vol. 19, no. 11, pp. 741–748, 2012.

[20] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona, "Fiji: An open-source platform for biological-image analysis," *Nature Methods*, vol. 9, no. 7, pp. 676–682, 2012.

[21] M. W. T. Werten, W. H. Wisselink, T. J. Jansen-van den Bosch, E. C. de Bruin, and F. a. de Wolf, "Secreted production of a custom-designed, highly hydrophilic gelatin in Pichia pastoris," *Protein Engineering Design and Selection*, vol. 14, pp. 447–454, jun 2001.

[22] M. Barczyk, S. Carracedo, and D. Gullberg, "Integrins," *Cell and Tissue Research*, vol. 339, no. 1, pp. 269–280, 2010.

[23] H. T. Ahmedah, L. H. Patterson, S. D. Shnyder, and H. M. Sheldrake, "RGD-binding integrins in head and neck cancers," *Cancers*, vol. 9, no. 6, pp. 1–17, 2017.

[24] D. Pallarola, A. Bochen, H. Boehm, F. Rechenmacher, T. R. Sobahi, J. P. Spatz, and H. Kessler, "Interface immobilization chemistry of cRGD-based peptides regulates integrin mediated cell adhesion," *Advanced Functional Materials*, vol. 24, no. 7, pp. 943–956, 2014.

[25] M. Durymanov and J. Reineke, "Non-viral delivery of nucleic acids: Insight into mechanisms of overcoming intracellular barriers," *Frontiers in Pharmacology*, vol. 9, pp. 1–15, 2018.

[26] A. K. Varkouhi, M. Scholte, G. Storm, and H. J. Haisma, "Endosomal escape pathways for delivery of biologicals," *Journal of Controlled Release*, vol. 151, no. 3, pp. 220–228, 2011.

[27] L. Johannes and M. Lucchino, "Current Challenges in Delivery and Cytosolic Translocation of Therapeutic RNAs," *Nucleic Acid Therapeutics*, vol. 28, no. 3, pp. 178–193, 2018.

[28] W. Li, F. Nicol, and F. C. Szoka, "GALA: A designed synthetic pH-responsive amphipathic peptide with applications in drug and gene delivery," *Advanced Drug Delivery Reviews*, vol. 56, no. 7, pp. 967–985, 2004.

[29] O. Boussif, F. Lezoualc'h, M. A. Zanta, M. D. Mergny, D. Scherman, B. Demeneix, and J.-P. Behr, "A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: Polyethylenimine," *Proc. Nat. Acad. Sci. USA*, vol. 92, no. 16, pp. 7297–7301, 1995.

[30] J.-P. Behr, "The Proton Sponge: a Trick to Enter Cells the Viruses Did Not Exploit," *Chimia*, vol. 51, pp. 34–36, 1997.

[31] L. H. Beun, I. M. Storm, M. W. T. Werten, F. A. De Wolf, M. A. Cohen Stuart, and R. De Vries, "From micelles to fibers: Balancing self-assembling and random coiling domains in pH-responsive silk-collagen-like protein-based polymers," *Biomacromolecules*, vol. 15, no. 9, pp. 3349–3357, 2014.

[32] M. D. Golinska, M. K. Włodarczyk-Biegun, M. W. Werten, M. A. Stuart, F. A. De Wolf, and R. De Vries, "Dilute self-healing hydrogels of silk-collagen-like block copolypeptides at neutral pH," *Biomacromolecules*, vol. 15, no. 3, pp. 699–706, 2014.

[33] B. Zhao, M. A. Cohen Stuart, and C. K. Hall, "Navigating in foldonia: Using accelerated molecular dynamics to explore stability, unfolding and self-healing of the β-solenoid structure formed by a silk-like polypeptide," *PLOS Computational Biology*, vol. 13, no. 3, p. e1005446, 2017.

[34] D. Van Den Brand, L. F. Massuger, R. Brock, and W. P. Verdurmen, "Mimicking Tumors: Toward More Predictive in Vitro Models for Peptide- and Protein-Conjugated Drugs," *Bioconjugate Chemistry*, vol. 28, no. 3, pp. 846–856, 2017.

[35] D. van den Brand, C. Veelken, L. Massuger, and R. Brock, "Penetration in 3D tumor spheroids and explants: Adding a further dimension to the structure-activity relationship of cell-penetrating peptides," *Biochimica et Biophysica Acta - Biomembranes*, vol. 1860, no. 6, pp. 1342–1349, 2018.

[36] V. P. Chauhan, Z. Popović, O. Chen, J. Cui, D. Fukumura, M. G. Bawendi, and R. K. Jain, "Fluorescent nanorods and nanospheres for real-time in vivo probing of nanoparticle shape-dependent tumor penetration," *Angewandte Chemie - International Edition*, vol. 50, no. 48, pp. 11417–11420, 2011.

[37] R. Fernandes, N. R. Smyth, O. L. Muskens, S. Nitti, A. Heuer-Jungemann, M. R. Ardern-Jones, and A. G. Kanaras, "Interactions of skin with gold nanoparticles of different surface charge, shape, and functionality," *Small*, vol. 11, no. 6, pp. 713–721, 2015.

6
General Discussion

Synthetic viruses, i.e. nanoparticles that mimic natural viruses in key properties such as protection and delivery of cargo molecules, are promising nano-containers and nano-delivery devices for a variety of applications since they are easier to produce and are associated with fewer safety risks than natural viruses[1–4]. Protein-based synthetic viruses have the additional attractive features that they might be biodegradable and that their protein components can be precisely genetically engineered[5,6].

Controlled assembly of protein-based block copolymers

As discussed in the General Introduction of this thesis, for co-assembly with a template such as DNA or RNA, a protein-based block copolymer requires a binding block for interactions with the template, a crystallisable block for assembly into fibrils, and an amorphous block for ensuring solubility of the fully assembled virus-like particles (VLPs). The previously published artificial viral capsid protein C-$S^Q{}_{10}$-K_{12} comprises a collagen-inspired amorphous block (C), a silk-inspired crystallisable block ($S^Q{}_{10}$) and an oligolysine for electrostatic-induced template binding (K_{12})[7]. In this thesis, we successfully substituted the collagen-inspired block for an amorphous elastin-inspired block (E^S) and studied the physical rules for the design of such a protein-based triblock copolymer in more detail by studying $E^S{}_m$-$S^Q{}_n$ and $E^S{}_m$-$S^Q{}_n$-B proteins, where m and n indicate the number of repeats of the elastin-inspired and silk-inspired sequences respectively and B is the template-binding block.

Additional design rules concerning protein self-assembly

From our studies on the $E^S{}_m$-$S^Q{}_n$ and $E^S{}_m$-$S^Q{}_n$-B proteins, we can extract some additional design rules for artificial viral capsid proteins with a triblock structure:

(1) The length ratio of the crystallisable to amorphous blocks should be very precisely optimised to allow for optimal co-assembly behaviour. As shown in Chapter 2, the optimal silk-like block length appears to be one just larger than the minimal length that allows for silk-like block crystallisation: only in this case the amorphous block length can be increased to oppose fibril formation in the absence of templates.

(2) As shown in Chapter 3, self-assembly of the artificial viral capsid proteins is rather sensitive to the chemical details of the amorphous block, in particular to their hydrophilicity. Increasing hydrophilicity of the amorphous block counteracts self-assembly of the crystallisable block and hence is another handle that can be used to oppose fibril formation in the absence of templates.

Recently, a similar self-assembly behaviour was observed for diblock copolymers of hydrophobic resilin-like polypeptides (RLP) and hydrophilic ELPs, where longer hydrophobic RLP blocks promote and longer hydrophilic ELP blocks reduce self-assembly into fibrils (such as described in design rule (1))[8]. In this case, the hydrophobic sequences likely act similar to the crystallisable block in our artificial viral capsid proteins while the hydrophilic sequences act similar to the amorphous block. Additionally, the same study also shows that further increasing hydrophilicity of the "amorphous" ELP block opposes self-assembly into fibrils, which is in agreement with design rule (2). Above mentioned design rules (1) and (2) are thus not unique to our artificial viral capsid proteins and hence should also be considered when designing other self-assembling protein-based block copolymers composed of both hydrophobic and hydrophilic sequence blocks.

Design recommendations for co-assembly of the artificial viral capsid proteins: increasing length of amorphous $E^S{}_m$ block

If template-induced assembly is the aim, as is the case for the artificial viral capsid proteins presented here, it is essential to prevent the non-template-induced self-assembly of the proteins into fibrils that do not contain any template. To achieve this, according to above mentioned design rules (1) and (2), the crystallisible silk block length should be kept as short as possible (Chapter 2) and for the amorphous block its length (Chapter 2) and hydrophilicity (Chapter 3) should be increased enough to reduce the self-assembly behaviour of the silk block while retaining its co-assembly behaviour.

The *C*-$S^Q{}_{10}$-K_{12} artificial viral capsid proteins were shown to co-assemble with polyanionic templates into rod-shaped virus-like particles (VLPs) that are uniform in length, while assembled structures are completely absent without such templates[7,9]. The length of crystallisable silk block and the length and hydrophilicity of the amorphous collagen-like block thus seem to be quite optimal for the purpose of co-assembly. The design of the $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} artificial viral capsid proteins, which is identical to *C*-$S^Q{}_{10}$-K_{12} except for the amorphous block, does not seem to be completely optimal yet (Chapter 3). These proteins assemble into rod-shaped particles even in the absence of polyanionic templates, indicating that the non-template induced self-assembly is too strong and should be reduced. While the length of the amorphous block is similar to that of the collagen-like block in *C*-$S^Q{}_{10}$-K_{12}, the $E^S{}_{80}$ block is much less hydrophilic and therefore likely less efficient in reducing the self-assembly behaviour of the $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} proteins. To remedy this, we could apply the above mentioned design rules and increase either the length or the hydrophilicity of the amorphous elastin-like block. As discussed in Chapter 3, hydrophilicity of the elastin-like block cannot be increased to levels similar to that of the collagen-inspired block, but increasing the number of repeats m of the $E^S{}_m$ block is quite straightforward using the PRe-RDL cloning technique[10] used in this thesis. Initial work to produce these proteins was carried out recently ($E^S{}_{100}$-$S^Q{}_{10}$-K_{12} and $E^S{}_{120}$-$S^Q{}_{10}$-K_{12}; also see the Appendix), but we were not able yet to characterise their self- and co-assembly behaviour. Completely replacing the elastin-like block for another hydrophilic amorphous block would also be an

interesting strategy. It will, however, be more difficult to determine whether observed differences between the assembly behaviour of $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} and such a new version of the protein is the result of e.g. increased hydrophilicity or due to a change in yet other chemical details of this new amorphous block.

Design recommendations for co-assembly of the artificial viral capsid proteins: sequence-specific template binding

Natural viruses are capable of specifically encapsulating their own genomic nucleic acids within the complex environment of host cells. They do this via the recognition of packaging signals in their genomes[11–13]. In contrast, our initial artificial viral capsid proteins bind electrostatically to polyanionic templates via the K_{12} binding block, and hence encapsulate nucleic acids non-specifically[7]. By choosing a binding block that specifically binds to a certain sequence, we can use this sequence as a packaging signal, potentially allowing for the specific packaging of templates containing one or more of these packaging signals (Chapter 4). In an attempt to develop such better mimics of natural viruses, we designed an $E^S{}_{80}$-$S^Q{}_{10}$-B protein with a binding block B based on the GAGA factor zinc finger found in *D. melanogaster*. Although here we have not yet been able to show the specific encapsulation of DNA harbouring a packaging signal due to various technical problems, we have no doubts about the eventual feasibility of this goal. Indeed, applying this same strategy of a sequence-specific binding module and using the recognition sequence as an artificial packaging signal, has allowed for the sequence-specific encapsulation of mRNA molecules in a non-viral protein capsid[14].

Packaging of the nucleic acid template

In the General Introduction we discussed how the artificial viral capsid proteins self-assemble into fibrils and what these fibrils look like at a molecular level (Figure 1.3b in the General Introduction). We did however not discuss the folding of the nucleic acid template within the rod-shaped VLPs. Unfortunately, as of yet, this is still an outstanding issue. Possibly, during encapsulation, the DNA forms a (super)helix around the core of the VLPs that is formed by the silk-inspired blocks (Figure 6.1). How could such a hypothesis be tested?

A first option would be to use FRET (fluorescence resonance energy transfer[15], using DNA strands labelled with pairs of fluorophores at well-chosen distances. If these pairs come close to each other during DNA encapsulation and condensation, they will give a FRET signal. Alternatively, using single-molecule force-spectroscopy[16,17], the kinetics of encapsulation of a single DNA chain may be probed in great detail and this may also provide clues on how the DNA is condensed and packed inside the artificial virus capsids. Both types of experiments are currently being done in labs of our collaborators. High resolution cryo-electron microscopy, solid-state NMR (nuclear magnetic resonance) or X-ray scattering may yield further clues such that eventually we may know how the DNA is packed inside the VLPs.

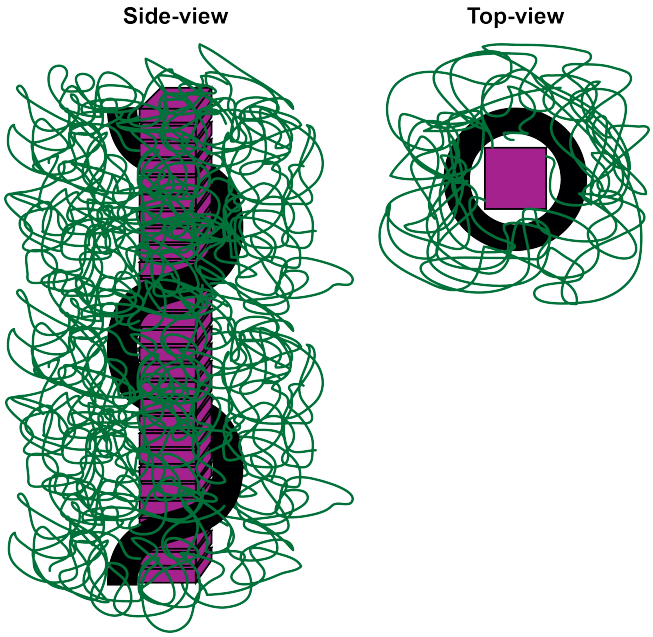


Figure 6.1 – Hypothetical appearance of a nucleic acid template encapsulated by the artificial viral capsid proteins. The exact folding of a nucleic acid template upon encapsulation in a VLP formed by the artificial viral capsid proteins is yet unknown. As depicted here, a possible conformation could be that of a (super)helix (nucleic acid template in black) around the rod-shaped silk-like core (in purple) of the VLPs. The template is thereby protected from degradation by nucleases due to the shielding effect of the collagen- or elastin-like blocks that form a corona (in green) around the silk-like/nucleic acid core of the VLPs.

Cellular delivery of nucleic acid molecules by protein block copolymers

VLPs based on our *C*-$S^{Q}{}_{10}$-K_{12} artificial viral capsid protein that encapsulate mRNA, show limited transfection efficiency[18]. Indeed, many biological barriers exist that each need to be overcome to allow for successful translation of the encapsulated genes into protein products[1,2] (also see the General Introduction of this thesis).

Design recommendations to improve cell entry

Ignoring for the moment the problem of targeted delivery at the organism-level, and focusing on uptake of the VLPs by cells, the first barrier is the cell membrane. Crossing this membrane, usually via an active cellular uptake process, is a first key step for successful delivery and transfection. For many nanomedicines this is found to require their conjugation to receptor-binding or cell-penetrating peptides[19–21].

In Chapter 5, we focused on this first step. An integrin-binding RGD peptide was conjugated to the amorphous *C* blocks that are forming a corona around the VLP core consisting of the silk-inspired blocks, the oligolysine blocks and the nucleic acid template. We found only a weak effect of this conjugation on cell entry and hypothesised that the receptor-binding peptide may have been shielded from binding to the receptors by the long flexible, amorphous *C* blocks. Such a shielding effect should be considered as a general issue when presenting cell-binding or cell-penetrating peptides on the surface of synthetic virus particles. Indeed, similar observations have also been discussed in literature for the attachment of RGD peptides to gold nanoparticles using long, flexible PEG linkers[22]. Long and stiff linkers, on the other hand, improved binding, likely due to projecting the peptide further away from the surface[22]. In the same way, for our VLPs it may be advisable to include long and stiff linkers when attaching cell-binding peptides to the amorphous block, such that the peptides are projected away from the shielding effect of this block.

Many lipid- and polymer-based synthetic viruses with good transfection efficiencies, also present positive charges on their surface. These are thought to facilitate binding to the negatively charged membranes of cells via non-specific electrostatic interactions. The surface of *C*-$S^{Q}{}_{10}$-K_{12} VLPs is formed by the collagen-inspired *C* blocks. These have a low net negative charge at physiological conditions[7], as the sequence of the *C* block contains more negatively charged glutamic acid residues (viz. 16) than positively charged lysines (viz. 12). This weak negative charge may be unfavourable for cell binding and hence transfection. Therefore, it would also be interesting to study cell binding and entry of $E^{S}{}_{m}$-$S^{Q}{}_{10}$-K_{12} VLPs, since these do not carry such a negative surface charge but instead have uncharged elastin-inspired amorphous blocks on their surface. Alternatively, artificial viral capsid proteins could even be designed with elastin-like amorphous blocks that possess a small net positive charge.

Design recommendations to improve endosomal escape

Often synthetic virus particles enter cells via so-called endosomes, membrane bound compartments that are actively internalised by eukaryotic cells. For the delivery of nucleic acids to the cytoplasm it is crucial to escape from such endosomes in time, since in principle they are on a pathway that ultimately leads to the breakdown of their contents. If the synthetic virus particles have successfully escaped from the endosomes into the cytoplasm, a next step is the release of the nucleic acids from the artificial viral capsids[1,2]. In Chapter 5 (Figure 5.5 and 5.6), we obtained microscopic indications that, like most nanoparticles, mRNA containing *C*-$S^{Q}{}_{10}$-K_{12} VLPs do indeed enter via endosomes. We, however, did not observe any gene expression (as measured with a luciferase assay), indicating that the VLPs failed to escape from the endosomes or otherwise did not release intact mRNA into the cytoplasm.

In order to improve the design of the artificial viral capsid proteins with respect to transfection, it will be important to obtain more detailed information on the uptake and processing of the VLPs by the cells. For example, it would be interesting

to perform co-localisation studies by not only using fluorescently labelled proteins and mRNA, but also fluorescent markers for subcellular compartments such as endosomes, and test the effects of various inhibitors for specific steps during cell entry. This way, it should be possible to more precisely identify the main bottlenecks to transfection by our artificial viral particles.

Upon identification of the bottlenecks, precise adjustments can be made to the design of the artificial viral capsid proteins. Endosomal escape, for example, could be triggered by attaching peptides such as GALA to the surface of the VLPs that disrupt the endosomal membrane upon acidification[23]. Similar to presenting the RGD cell-binding peptide on the VLP surface, the endosomal escape peptides likely have to be placed on top of stiff linkers prior to attachment to the amorphous *C* block. Attachment can be done via bioconjugation, as for the RGD peptides in Chapter 5, or via genetic engineering.

A strategy that may possibly promote both endosomal escape of the VLPs and the release of the mRNA from the VLPs is to design artificial viral capsid proteins with a pH-responsive silk-inspired block $S^H{}_n$ = $(GAGAGAGH)_n$. Upon decreasing the pH, the histidine residues within this silk-like motif become protonated. The liberation of these protons may result in disruption of the endosome through the so-called proton sponge effect, as was also suggested for PEI-based particles[24–26]. In addition, for triblock proteins containing an S^H block it has been shown that these are soluble at low pH and self-assemble into fibrils only at physiological pH[27,28]. If this $S^H{}_n$ block is used as the midbock in our artificial viral capsid protein, acidification in the endosome may thus also promote the disassembly of the VLPs and consequently the release of the mRNA. While we have already designed and genetically engineered a set of $E^S{}_{80}$-$S^H{}_n$-K_{12} proteins (n = 10, 16, 20 or 24; see the Appendix) that could be used to study these effects, these proteins have not yet been produced and characterised.

Applications for the artificial viral capsid proteins

Advantages of the artificial viral capsid proteins

As compared to other synthetic viruses intended for the delivery of nucleic acids (or other cargo) into cells, we believe that synthetic virus particles based on proteins have a number of important advantages (also discussed in the introduction of this thesis) such as their biodegradability, and the possibility to precisely genetically engineering them[5,6,29]. In addition, the specific triblock design of the artificial viral capsid proteins discussed in this thesis may also have advantages over other protein-based synthetic virus particles.

Firstly, they are highly modular with respect to their design: each physical "function" is carried out by a unique block within the protein[7,30] and is largely independent from the actions of the other blocks. Consequently, the functionalities of the protein can be adapted relatively easily depending on the intended application, for example by replacing individual protein blocks for other ones or by the fusion of additional functional blocks. Indeed, in Chapter 3 of this thesis it was shown that

replacing the amorphous collagen-inspired *C* block by an amorphous elastin-like $E^S{}_m$ block preserves the co-assembly behaviour of the protein with nucleic acid templates, although the lower hydrophilicity of the $E^S{}_m$ does lead to some changes in the assembly of the protein in the absence of a template. In a similar way, one could replace the amorphous block by yet another type of block, for example, one exhibiting an additional property, e.g. a thermal transition (either LCST or UCST), or a reduced immunogenicity.

Fusion with a domain that adds an extra functionality was demonstrated in Chapter 5 of this thesis, where we attached an RGD integrin-binding peptide to the amorphous block of the protein. Although we used bioconjugation in that chapter, it should also be possible to fuse the extra domain to the amorphous block via genetic engineering. In a similar way, many other types of blocks could be fused to the amorphous block in order to impart additional functionalities to the VLPs, for example receptor-binding domains that are much more cell-type specific than the integrin binding RGD peptides[31].

The data in this thesis, in addition, shows that versions of the artificial viral capsid proteins can be produced in different expression hosts. The new version of the protein ($E^S{}_{80}$-$S^Q{}_{10}$-K_{12}) is produced in the bacterium *E. coli*, in contrast to the original *C*-$S^Q{}_{10}$-K_{12} protein that is produced in the yeast *P. pastoris*[7]. The flexibility to produce this protein in a variety of expression hosts is an important feature as every host has its own advantages and disadvantages with respect to protein production at different scales and for different purposes[32]. Bacteria, for example, are highly suitable for the fast and easy production of proteins used in laboratory experiments. Production in yeast and insect cells, on the other hand, can be scaled-up more easily and in addition may support post-translational modifications that are required for the proper folding of some proteins. Flexibility regarding the production host thus facilitates the development of a synthetic virus into a tool that can be used in a variety of applications and eases the transition to large-scale production.

A unique feature of the present VLPs is that they are rod-shaped, while most other synthetic viruses have spherical shapes. The shape of nanoparticles has recently been shown to have a significant impact on its penetration into 3D tissues, with rod-shaped particles being able to travel further into the tissues than spherical particles[33,34]. The development of rod-shaped or filamentous synthetic viruses may therefore be crucial to eventually obtain synthetic virus particles that not only transfect cells, but also effectively penetrate tissues.

More than in other synthetic virus particle systems, assembly of the particles in our case is strongly dependent on the presence of a template. If the amorphous and crystallisable blocks are properly tuned, essentially no self-assembly occurs at protein concentrations below a precisely determined critical value[7]. We believe that template-induced assembly occurs due to binding of the artificial viral capsid proteins to the template, so that the local protein concentration at some stage exceeds this critical concentration for self-assembly, even though the protein concentration in the bulk solution is still very low. Such a template-controlled assembly should allow for the production of pure batches with only template-carrying VLPs.

Model system to study assembly and disassembly of natural virus particles

The modular and simple design of the artificial viral capsid proteins facilitates their use in a broad range of applications. First of all, they are a useful model system for the (*in vitro*) assembly and disassembly of capsids composed of natural viral proteins. The original design already reveals three elementary physical properties of viral capsid proteins: nucleic acid binding, self-assembly and colloidal stability[7]. In this thesis, we describe initial steps to allow for the preferential encapsulation of nucleic acids harbouring specific sequences (packaging signals), such as seen in many natural viruses (Chapter 4). As assembly of our artificial viral particles is highly template-induced, we believe that such specificity for a packaging signal would result in the encapsulation of only the specific template while leaving non-specific templates unpacked. This is in contrast with the recently published synthetic virus particles encapsulating their own genomes[14,35] for which the capsids still package significant amounts of other (cellular) nucleic acids. Further development of the artificial viral capsid proteins, discussed in this thesis, into capsid proteins that encapsulate their own genome inside a model bacterium such as *E. coli*, may ultimately lead to a self-replicating artificial virus that could serve as a model system to study even more aspects of natural viruses.

The *de novo* or bottom-up design of synthetic materials that mimic specific biological entities is not only being considered for viruses. Since 2017, 17 laboratories in the Netherlands collaborate to build a so-called synthetic cell from scratch and for sure more of such attempts will follow in Europe[36] and worldwide. Indeed, the bottom-up mimicry of biological systems is a very appealing strategy to find out precisely which elements of a biological entity are absolutely essential and which are not. Attempting to build biological entities from scratch will greatly improve our understanding of how life works.

Applications in nucleic acid delivery

Using the artificial viral capsid proteins, both DNA[7] and RNA[18] templates can be encapsulated. Due to the shielding effect of the amorphous block, these templates are protected from degradation by nucleases, so that the assembled VLPs are potentially very interesting nucleic acid delivery vehicles. Many applications exist within the field of nucleic acid delivery[2,37,38] and the modular design of the artificial viral capsid proteins, in theory, allows for their adaptation to any of these applications. Examples are gene therapy (including protein replacement, gene silencing and gene editing) and DNA and RNA vaccination.

Gene therapy includes various strategies, such as protein replacement, gene silencing and gene editing[39]. Protein replacement can be applied in genetic diseases characterised by the insufficient expression of a protein or the expression of a defective protein (e.g. in lipoprotein lipase deficiency[40]). In these cases, DNA or mRNA molecules coding for the functional protein can be delivered into cells using synthetic viruses to induce the production of the functional protein. While mRNA is degraded quickly in the cell, repeated treatment may be necessary. DNA,

on the other hand, has the advantage that it can reside as an episome in the cell nucleus for months or even years[41], so that it has a continued therapeutic effect and a single treatment may be sufficient.

Diseases caused by the overexpression of a protein, can be treated using gene silencing[42]. Using synthetic viruses, small interfering RNAs (siRNAs) could be delivered to the target cells in which they bind to and cleave complementary mRNA in order to downregulate the expression of the mRNA-encoded protein. This strategy is for example used in the treatment of viral infections by silencing viral genes[43], or in the treatment of cancer for example by downregulating genes involved in proliferation, migration and anti-apoptosis[44]. The effects of siRNA treatments are usually short as RNAs are degraded quickly. If sustained gene silencing is desired, it may be more interesting to use short hairpin RNAs (shRNAs)[45]. shRNAs downregulate protein expression in a similar fashion as siRNA. However, shRNAs can be expressed from a DNA plasmid that is delivered to the cell nucleus. Since DNA in the nucleus can be relatively stable, the effect of shRNAs lasts much longer.

Long-lasting effects can also be obtained by gene editing, which involves the introduction of breaks at specific locations in the genome of a cell using endonucleases such as zinc finger (ZnF) nucleases, transcription activator-like effector nucleases (TALENs) and the more recently discovered CRISPR RNA-guided Cas9[46,47]. This strategy relies on the cellular DNA repair machinery and could result in the deletion or replacement of a genomic sequence or in the insertion of a "donor" sequence. Delivery of the nucleases to the target cells could be achieved by delivery of DNA or RNA coding for these nucleases and the guide RNA, with the use of synthetic viruses[48].

DNA and RNA vaccination is another interesting application for which synthetic viruses may be useful[37,38,49]. DNA or RNA molecules coding for pathogenic antigens are delivered into antigen presenting cells using synthetic viruses. The antigen presenting cells then express these pathogenic antigens and present them on their cell membrane, which elicits an immune response against the antigen. Similar to the more traditional (viral) vaccines that are based on whole viruses or viral subunits, the DNA and RNA vaccines elicit a protective immunity against the pathogens to assure a fast response upon an actual infection. In addition, DNA and RNA vaccines are now also studied in the context of cancer treatment by using DNA or RNA coding for tumour antigens.

It should be mentioned that there might also be some limitations in the use of the artificial viral capsid proteins for nucleic acid delivery. Presumably the biggest issue will be the size of the encapsulated nucleic acids. Encapsulated nucleic acids are condensed with a packing factor of $\sim$3 for double-stranded molecules and $\sim$6 for single-stranded molecules[7]. Based on the experiments published before[7,18] and the experiments discussed in this thesis, we know that this packing factor is independent of the template length, so that the length of the rod-shaped VLPs increases proportionally with the length of the template. It remains to be elucidated what the limits are. For long mRNA or long DNA it may therefore be interesting to develop artificial viral capsid proteins with a larger packing factor. This could possibly be achieved by using longer oligolysine binding blocks, since this would allow each individual protein in a VLP to bind to a larger number of base pairs and

hence force the nucleic acid template to be condensed to a greater extend.

Other applications

Although, herein we mainly focused on the delivery of nucleic acids, the artificial viral capsid proteins can potentially also be used in many other applications in medicine and biotechnology. Examples are drug delivery and nanoreactors. These applications, and more, have been extensively reviewed by Steinmetz and colleagues in the context of virus-based nanomaterials[50]. Due to the modular design of our artificial viral capsid proteins, they can presumably be deployed for many of these applications. We will, however, shortly discuss only a selection of them here.

The rod-shaped assemblies of the artificial viral capsid proteins may be ideal platforms for the conjugation of drugs or enzymes, similarly to the loading and coating of TMV particles with such substances. That way, TMV particles have for example been explored for use in cancer therapy to deliver toxic drugs such as doxorubicin[51,52]. In general, drugs delivered with TMV particles are loaded inside the hollow tube of the particles to prevent off-target effects of the drug. The space inside the VLPs formed by our artificial viral capsid proteins is limited though and likely does not allow for drug-loading, although a DNA-binding drug such as doxorubicin could be loaded into a DNA-containing VLP.

Instead the VLPs may be used as scaffolds to which drugs or enzymes can be attached. The flexible amorphous blocks may (partially) shield the drugs from acting on off-targets as was presumably also seen in Chapter 5 for conjugated RGD peptides. Release of the drugs inside the target cells could then potentially be facilitated by connection via pH sensitive linkers, or the amorphous domain itself might be made cleavable by endosomal proteases in order to release the drugs.

Coating of our VLPs with enzymes may be an interesting strategy in biotechnology as several enzymes can be conjugated onto a single VLP scaffold and made to be in close proximity of each other. This way, subsequent steps of an enzyme cascade can be coupled to facilitate the efficient conversion of a substrate into a desired product. This strategy is similar to the idea of nanoreactors that are composed of hollow, spherical particles encapsulating a selection of enzymes[53,54].

Concluding remarks

We have argued that protein-based synthetic viruses are promising nanoparticles for many applications in nanomedicine and biotechnology. The artificial viral capsid proteins presented in this thesis have specific advantages over other designs that we believe make them useful in specific applications. To realise their full potential however, significant steps still have to be taken. Key steps are especially (1) the elucidation of why the current artificial viral capsid particles have low transfection efficiency, such that we can remedy this by elaborating on our basic design, and (2) the successful development of a capsid protein that can encapsulate its own genome inside a cellular host, such that we can use directed evolution to rapidly generate new versions with improved properties.

Appendix

Overview of constructs encoding artificial viral capsid proteins

A number of constructs were created that encode variants of the artificial viral capsid protein and are compatible with protein expression in *Escherichia coli*. This appendix gives an overview of those protein variants, of which some were also discussed in this thesis.

All genes encoding for the artificial viral capsid proteins were constructed by a cloning technique named recursive directional ligation by plasmid reconstruction (PRe-RDL)[10]. The vector used for cloning is the pET24a(+) PRe-RDL vector (Figure A6.1) and is also the expression vector for protein expression in *E. coli*. Figure A6.1 depicts such a vector containing the gene for a triblock variant of the artificial viral capsid protein. Green represents an amorphous elastin-inspired ${E^S}_m$ block where ${E^S}_m = (\mathrm{GSGVP})_m$ with m number of repeats. The purple block is a silk-inspired self-assembly block with n number of repeats of the sequence ${S^X}_n = (\mathrm{GAGAGAGX})_n$, where X is either glutamine (Q) or histidine (H). At last, the red block represents the nucleic acid binding block B, which could either be a non-specific oligolysine K_{12} or a sequence-specific DNA binding block.

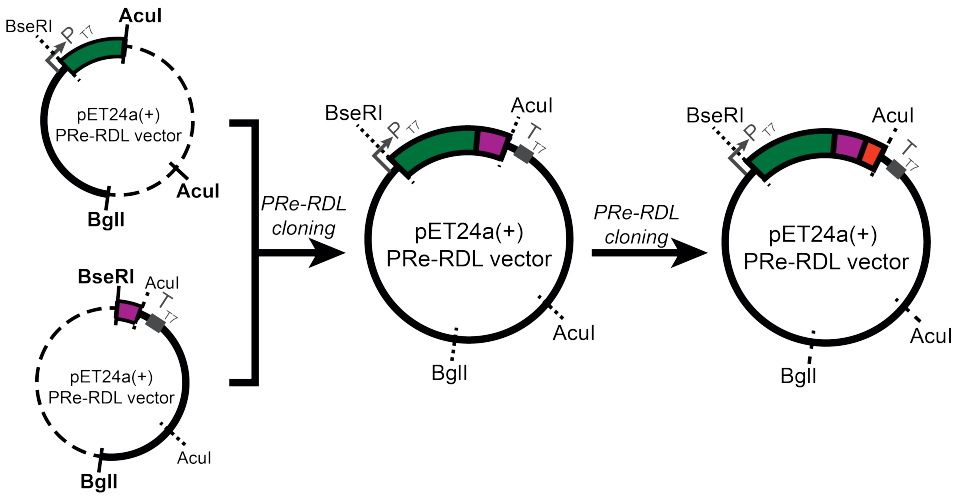


Figure A6.1 – PRe-RDL vector with a gene encoding an artificial viral capsid protein. The pET24a(+) PRe-RDL vector is used for both cloning and expression of the artificial viral capsid proteins. The cleavage sites of the three restriction enzymes (*BglI*, *BseRI* and *AcuI*) used for PRe-RDL cloning are shown. Also the T7 promoter and terminator are indicated. The gene encoding for the protein is depicted as a triblock, with green representing the elastin-inspired ${E^S}_m$ block, purple the silk-inspired ${S^X}_n$ block, and red the nucleic acid binding block.

Proteins discussed in this thesis:

$\boldsymbol{E^S{}_m\text{-}S^Q{}_n}$ diblocks, with m = 40, 60 or 80 and n = 6, 10 or 14

Discussed in Chapter 2. These proteins were used to study the effect of changing the block lengths on protein self-assembly.

$\boldsymbol{E^S{}_{80}\text{-}S^Q{}_{10}\text{-}K_{12}}$ triblocks

Discussed in Chapter 3. Using this protein we studied the effect of replacing the amorphous block for another amorphous block. In the original protein (C-$S^Q{}_{10}$-K_{12}) the amorphous C block consists of the collagen-inspired sequence $(\text{GXaaYaa})_{132}$, which we here replaced for the elastin-inspired $E^S{}_{80}$ block.

$\boldsymbol{E^S{}_{80}\text{-}S^Q{}_{10}\text{-}B}$ triblocks, with B is one of the candidate DNA binding blocks

Discussed in Chapter 4. The purpose of these proteins is to develop an artificial viral capsid protein that can encapsulate a specific DNA template via the recognition of a so-called packaging signal in the DNA. Three candidate DNA binding blocks were proposed: a zinc finger (ZnF), a basic leucine zipper (bZIP) and an AT-hook (AT). The recognition sequences of these binding blocks serve as the packaging signals.

Proteins not discussed in this thesis:

$\boldsymbol{E^S{}_m\text{-}S^Q{}_{10}\text{-}K_{12}}$, with m = 100 or 120

Discussed in the BSc report of Maarten Klaverdijk, who was supervised by me. This BSc project was designed by myself and based on the observation in Chapter 3 of the present thesis: the $E^S{}_{80}$-$S^Q{}_{10}$-K_{12} proteins (with 80 repeats of the elastin-inspired sequence) have a higher tendency to self-assemble, thereby forming many protein-only fibrils in addition to DNA-containing VLPs upon incubation with DNA. The aim of this BSc project was to study whether we can suppress the self-assembly behaviour of the $E^S{}_m$-$S^Q{}_{10}$-K_{12} proteins by increasing the length of the amorphous $E^S{}_m$ block. The choice to increase the length of the $E^S{}_m$ block was motivated by the results of Chapter 2 of this thesis, which showed that longer amorphous $E^S{}_m$ blocks suppress self-assembly of $E^S{}_m$-$S^Q{}_n$ diblocks. Unfortunately, we did not have enough time to fully characterise the proteins.

$\boldsymbol{E^S{}_{80}\text{-}S^H{}_n\text{-}K_{12}}$, with n = 10, 16, 20 or 24

Discussed in the BSc report of Martin Bongers, who was supervised by me. Motivated to improve nucleic acid delivery to cells by VLPs formed by the artificial viral capsid proteins, we in this BSc project

aimed to design proteins that are expected to induce endosomal escape and unpacking of the nucleic acid template upon acidification in the endosomes. The design was based on previous work in our group[27,28,55] showing that block copolymers with a silk-inspired $S^H{}_n$ block solubilise at low pH and assemble into fibrils at neutral pH. We replaced the pH-insensitive $S^Q{}_{10}$ block in our proteins for the pH-responsive $S^H{}_n$ block. We hypothesise that upon decreasing the pH, the histidines within these $S^H{}_n$ silk-like blocks can bind the protons that are imported into the endosome, leading to the proton sponge effect and endosomal escape. In addition, the low pH presumably also triggers destabilisation of the silk-like core of the VLPs, resulting in unpacking of the nucleic acids. For $E^S{}_{80}$-$S^H{}_{16}$-K_{12}, we indeed obtained some indications that self-assembled protein fibrils fall apart upon changing the pH from 7.4 to 2.0. Characterisation of the triblocks with other values for n, and evaluating the effect upon cellular uptake has not been done yet.

References

[1] E. Mastrobattista, M. A. E. M. van der Aa, W. E. Hennink, and D. J. A. Crommelin, "Artificial viruses: a nanotechnological approach to gene delivery.," *Nature reviews. Drug discovery*, vol. 5, pp. 115–121, 2006.

[2] H. Yin, R. L. Kanasty, A. A. Eltoukhy, A. J. Vegas, J. R. Dorkin, and D. G. Anderson, "Non-viral vectors for gene-based therapy," *Nature Reviews Genetics*, vol. 15, no. 8, pp. 541–555, 2014.

[3] U. Unzueta, M. V. Céspedes, E. Vázquez, N. Ferrer-Miralles, R. Mangues, and A. Villaverde, "Towards protein-based viral mimetics for cancer therapies," *Trends in Biotechnology*, vol. 33, no. 5, pp. 253–258, 2015.

[4] A. Jain, S. K. Singh, S. K. Arya, S. C. Kundu, and S. Kapoor, "Protein Nanoparticles: Promising Platforms for Drug Delivery Applications," *ACS Biomaterials Science & Engineering*, vol. 4, pp. 3939–3961, 2018.

[5] a. Nagarsekar and H. Ghandehari, "Genetically engineered polymers for drug delivery," *Journal of drug targeting*, vol. 7, no. 1, pp. 11–32, 1999.

[6] O. S. Rabotyagova, P. Cebe, and D. L. Kaplan, "Protein-based block copolymers," *Biomacromolecules*, vol. 12, pp. 269–289, feb 2011.

[7] A. Hernandez-Garcia, D. J. Kraft, A. F. J. Janssen, P. H. H. Bomans, N. A. J. M. Sommerdijk, D. M. E. Thies-Weesie, M. E. Favretto, R. Brock, F. a. de Wolf, M. W. T. Werten, P. van der Schoot, M. C. Stuart, and R. de Vries, "Design and self-assembly of simple coat proteins for artificial viruses," *Nature Nanotechnology*, vol. 9, pp. 698–702, aug 2014.

[8] I. Weitzhandler, M. Dzuricky, I. Hoffmann, F. Garcia Quiroz, M. Gradzielski, and A. Chilkoti, "Micellar Self-Assembly of Recombinant Resilin-/Elastin-Like Block Copolypeptides," *Biomacromolecules*, vol. 18, no. 8, pp. 2419–2426, 2017.

[9] A. Hernandez-Garcia, M. A. Cohen Stuart, and R. de Vries, "Templated co-assembly into nanorods of polyanions and artificial virus capsid proteins," *Soft Matter*, vol. 14, no. 1, pp. 132–139, 2018.

[10] J. R. McDaniel, D. J. Callahan, and A. Chilkoti, "Drug delivery to solid tumors by elastin-like polypeptides.," *Advanced drug delivery reviews*, vol. 62, pp. 1456–1467, dec 2010.

[11] H. Fujisawa and P. Hearing, "Structure, function and specificity of the DNA packaging signals in double-stranded DNA viruses," *Seminars in Virology*, vol. 5, no. 1, pp. 5–13, 1994.

[12] P. G. Stockley, R. Twarock, S. E. Bakker, A. M. Barker, A. Borodavka, E. Dykeman, R. J. Ford, A. R. Pearson, S. E. Phillips, N. A. Ranson, and R. Tuma, "Packaging signals in single-stranded RNA viruses: Nature's alternative to a purely electrostatic assembly mechanism," *Journal of Biological Physics*, vol. 39, no. 2, pp. 277–287, 2013.

[13] R. Twarock, R. J. Bingham, E. C. Dykeman, and P. G. Stockley, "A modelling paradigm for RNA virus assembly," *Current Opinion in Virology*, no. Figure 1, pp. 1–8, 2018.

[14] N. Terasaka, Y. Azuma, and D. Hilvert, "Laboratory evolution of virus-like nucleocapsids from nonviral protein cages," *Proceedings of the National Academy of Sciences*, vol. 115, no. 21, pp. 5432–5437, 2018.

[15] P. R. Selvin, "The renaissance of fluorescence resonance energy transfer," *Nature Structural Biology*, vol. 7, no. 9, pp. 730–734, 2000.

[16] K. C. Neuman and A. Nagy, "Single-molecule force spectroscopy: Optical tweezers, magnetic tweezers and atomic force microscopy," *Nature Methods*, vol. 5, no. 6, pp. 491–505, 2008.

[17] G. Sitters, D. Kamsma, G. Thalhammer, M. Ritsch-Marte, E. J. Peterman, and G. J. Wuite, "Acoustic force spectroscopy," *Nature Methods*, vol. 12, no. 1, pp. 47–50, 2014.

[18] S. Jekhmane, R. de Haas, O. Paulino da Silva Filho, A. H. van Asbeck, M. E. Favretto, A. Hernandez Garcia, R. Brock, and R. de Vries, "Virus-Like Particles of mRNA with Artificial Minimal Coat Proteins: Particle Formation, Stability, and Transfection Efficiency," *Nucleic Acid Therapeutics*, vol. 27, no. 3, pp. 159–167, 2017.

[19] M. Durymanov and J. Reineke, "Non-viral delivery of nucleic acids: Insight into mechanisms of overcoming intracellular barriers," *Frontiers in Pharmacology*, vol. 9, pp. 1–15, 2018.

[20] A. K. Varkouhi, M. Scholte, G. Storm, and H. J. Haisma, "Endosomal escape pathways for delivery of biologicals," *Journal of Controlled Release*, vol. 151, no. 3, pp. 220–228, 2011.

[21] L. Johannes and M. Lucchino, "Current Challenges in Delivery and Cytosolic Translocation of Therapeutic RNAs," *Nucleic Acid Therapeutics*, vol. 28, no. 3, pp. 178–193, 2018.

[22] D. Pallarola, A. Bochen, H. Boehm, F. Rechenmacher, T. R. Sobahi, J. P. Spatz, and H. Kessler, "Interface immobilization chemistry of cRGD-based peptides regulates integrin mediated cell adhesion," *Advanced Functional Materials*, vol. 24, no. 7, pp. 943–956, 2014.

[23] W. Li, F. Nicol, and F. C. Szoka, "GALA: A designed synthetic pH-responsive amphipathic peptide with applications in drug and gene delivery," *Advanced Drug Delivery Reviews*, vol. 56, no. 7, pp. 967–985, 2004.

[24] U. Lächelt and E. Wagner, "Nucleic Acid Therapeutics Using Polyplexes : A Journey of 50 Years (and Beyond)," *Chemical Reviews*, vol. 115, no. 19, pp. 11043–11078, 2015.

[25] O. Boussif, F. Lezoualc'h, M. A. Zanta, M. D. Mergny, D. Scherman, B. Demeneix, and J.-P. Behr, "A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: Polyethylenimine," *Proc. Nat. Acad. Sci. USA*, vol. 92, no. 16, pp. 7297–7301, 1995.

[26] J.-P. Behr, "The Proton Sponge: a Trick to Enter Cells the Viruses Did Not Exploit," *Chimia*, vol. 51, pp. 34–36, 1997.

[27] L. H. Beun, I. M. Storm, M. W. T. Werten, F. A. De Wolf, M. A. Cohen Stuart, and R. De Vries, "From micelles to fibers: Balancing self-assembling and random coiling domains in pH-responsive silk-collagen-like protein-based polymers," *Biomacromolecules*, vol. 15, no. 9, pp. 3349–3357, 2014.

[28] M. D. Golinska, M. K. Włodarczyk-Biegun, M. W. Werten, M. A. Stuart, F. A. De Wolf, and R. De Vries, "Dilute self-healing hydrogels of silk-collagen-like block copolypeptides at neutral pH," *Biomacromolecules*, vol. 15, no. 3, pp. 699–706, 2014.

[29] D. Chow, M. L. Nunalee, D. W. Lim, A. J. Simnick, and A. Chilkoti, "Peptide-based biopolymers in biomedicine and biotechnology," *Materials Science and Engineering R: Reports*, vol. 62, no. 4, pp. 125–155, 2008.

[30] A. Hernandez-Garcia, M. W. T. Werten, M. C. Stuart, F. A. de Wolf, and R. de Vries, "Coating of single DNA molecules by genetically engineered protein diblock copolymers.," *Small (Weinheim an der Bergstrasse, Germany)*, vol. 8, pp. 3491–501, nov 2012.

[31] F. Araste, K. Abnous, M. Hashemi, S. M. Taghdisi, M. Ramezani, and M. Alibolandi, "Peptide-based targeted therapeutics: Focus on cancer treatment," *Journal of Controlled Release*, vol. 292, pp. 141–162, 2018.

[32] A. Berlec and B. Štrukelj, "Current state and recent advances in biopharmaceutical production in Escherichia coli, yeasts and mammalian cells," *Journal of Industrial Microbiology and Biotechnology*, vol. 40, no. 3-4, pp. 257–274, 2013.

[33] V. P. Chauhan, Z. Popović, O. Chen, J. Cui, D. Fukumura, M. G. Bawendi, and R. K. Jain, "Fluorescent nanorods and nanospheres for real-time in vivo probing of nanoparticle shape-dependent tumor penetration," *Angewandte Chemie - International Edition*, vol. 50, no. 48, pp. 11417–11420, 2011.

[34] R. Fernandes, N. R. Smyth, O. L. Muskens, S. Nitti, A. Heuer-Jungemann, M. R. Ardern-Jones, and A. G. Kanaras, "Interactions of skin with gold nanoparticles of different surface charge, shape, and functionality," *Small*, vol. 11, no. 6, pp. 713–721, 2015.

[35] G. L. Butterfield, M. J. Lajoie, H. H. Gustafson, D. L. Sellers, U. Nattermann, D. Ellis, J. B. Bale, S. Ke, G. H. Lenz, A. Yehdego, R. Ravichandran, S. H. Pun, N. P. King, and D. Baker, "Evolution of a designed protein assembly encapsulating its own RNA genome," *Nature*, vol. 552, no. 7685, pp. 415–420, 2017.

[36] K. Powell, "How biologists are creating life-like cells from scratch.," *Nature*, vol. 563, no. 7730, pp. 172–175, 2018.

[37] K. A. Hajj and K. A. Whitehead, "Tools for translation: Non-viral materials for therapeutic mRNA delivery," *Nature Reviews Materials*, vol. 2, pp. 1–17, 2017.

[38] N. Pardi, M. J. Hogan, F. W. Porter, and D. Weissman, "mRNA vaccines a new era in vaccinology," *Nature Reviews Drug Discovery*, vol. 17, no. 4, pp. 261–279, 2018.

[39] S. L. Ginn, A. K. Amaya, I. E. Alexander, M. Edelstein, and M. R. Abedi, "Gene Therapy Clinical Trials Worldwide To 2017 - an Update," *The Journal of Gene Medicine*, no. January, p. e3015, 2018.

[40] L. M. Bryant, D. M. Christopher, A. R. Giles, C. Hinderer, J. L. Rodriguez, J. B. Smith, E. a. Traxler, J. Tycko, A. P. Wojno, and J. M. Wilson, "Lessons learned from the clinical development and market authorization of Glybera.," *Human gene therapy. Clinical development*, vol. 24, no. 2, pp. 55–64, 2013.

[41] A. Léger, C. Guiner, M. L. Nickerson, K. Im, N. Ferry, P. Moullier, R. O. Snyder, and M. Penaud-Budloo, "Adeno-associated viral vector-mediated transgene expression is independent of dna methylation in primate liver and skeletal muscle," *PLoS ONE*, vol. 6, no. 6, p. e20881, 2011.

[42] C. Chakraborty, A. R. Sharma, G. Sharma, C. G. P. Doss, and S. S. Lee, "Therapeutic miRNA and siRNA: Moving from Bench to Clinic as Next Generation Medicine," *Molecular Therapy - Nucleic Acids*, vol. 8, pp. 132–143, 2017.

[43] J. DeVincenzo, J. E. Cehelsky, R. Alvarez, S. Elbashir, J. Harborth, I. Toudjarska, L. Nechev, V. Murugaiah, A. V. Vliet, A. K. Vaishnaw, and R. Meyers, "Evaluation of the safety, tolerability and pharmacokinetics of ALN-RSV01, a novel RNAi antiviral therapeutic directed against respiratory syncytial virus (RSV)," *Antiviral Research*, vol. 77, no. 3, pp. 225–231, 2008.

[44] T. Li, Y. Xue, G. Wang, T. Gu, Y. Li, Y. Y. Zhu, and L. Chen, "Multi-target siRNA: Therapeutic strategy for hepatocellular carcinoma," *Journal of Cancer*, vol. 7, no. 10, pp. 1317–1327, 2016.

[45] S. Aguiar, B. van der Gaag, and F. A. B. Cortese, "RNAi mechanisms in Huntington's disease therapy: SiRNA versus shRNA," *Translational Neurodegeneration*, vol. 6, p. 30, 2017.

[46] Y. Peng, K. J. Clark, J. M. Campbell, M. R. Panetta, Y. Guo, and S. C. Ekker, "Making designer mutants in model organisms," *Development*, vol. 141, pp. 4042–4054, oct 2014.

[47] W. Y. Wu, J. H. G. Lebbink, R. Kanaar, N. Geijsen, and J. van der Oost, "Genome editing by natural and engineered CRISPR-associated nucleases," *Nature Chemical Biology*, vol. 14, no. 7, pp. 642–651, 2018.

[48] J. B. Miller, S. Zhang, P. Kos, H. Xiong, K. Zhou, S. S. Perelman, H. Zhu, and D. J. Siegwart, "Non-Viral CRISPR/Cas Gene Editing In Vitro and In Vivo Enabled by Synthetic Nanoparticle Co-Delivery of Cas9 mRNA and sgRNA," *Angewandte Chemie - International Edition*, vol. 56, no. 4, pp. 1059–1063, 2017.

[49] M. A. Marć, E. Domínguez-Álvarez, and C. Gamazo, "Nucleic acid vaccination strategies against infectious diseases," *Expert Opinion on Drug Delivery*, vol. 12, no. 12, pp. 1851–1865, 2015.

[50] A. M. Wen and N. F. Steinmetz, "Design of virus-based nanomaterials for medicine, biotechnology, and energy," *Chemical Society Reviews*, vol. 45, no. 15, pp. 4074–4126, 2016.

[51] M. A. Bruckman, A. E. Czapar, A. VanMeter, L. N. Randolph, and N. F. Steinmetz, "Tobacco mosaic virus-based protein nanoparticles and nanorods for chemotherapy delivery targeting breast cancer," *Journal of Controlled Release*, vol. 231, pp. 103–113, 2016.

[52] A. E. Czapar, Y.-R. Zheng, I. A. Riddell, S. Shukla, S. G. Awuah, S. J. Lippard, and N. F. Steinmetz, "Tobacco Mosaic Virus Delivery of Phenanthriplatin for Cancer therapy," *ACS Nano*, vol. 10, no. 4, pp. 4119–4126, 2016.

[53] S. B. P. E. Timmermans and J. C. M. van Hest, "Self-assembled nanoreactors based on peptides and proteins," *Current Opinion in Colloid & Interface Science*, vol. 35, pp. 26–35, 2018.

[54] M. Rother, M. G. Nussbaumer, K. Renggli, and N. Bruns, "Protein cages and synthetic polymers: a fruitful symbiosis for drug delivery applications, bionanotechnology and materials science," *Chem. Soc. Rev.*, vol. 45, no. 22, pp. 6213–6249, 2016.

[55] L. H. Beun, X. J. Beaudoux, J. M. Kleijn, F. a. De Wolf, and M. a. Cohen Stuart, "Self-assembly of silk-collagen-like triblock copolymers resembles a supramolecular living polymerization," *ACS Nano*, vol. 6, no. 1, pp. 133–140, 2012.

Summary

Although most people associate viruses with diseases, virus particles are also highly valuable as nano-tools in medicine and biotechnology, either as a nano-containers or nano-delivery tools themselves, or as a source of inspiration for the development of new materials with virus-like functions. An example of such a new material is the artificial viral capsid protein that was recently developed in our group. This *de novo* designed protein mimics three key properties of the capsid proteins of the tobacco mosaic virus: nucleic acid binding, self-assembly into rod-shaped particles and colloidal stability. Each of these properties is encoded using a separate block or domain within our proteins, so that we obtain a triblock protein abbreviated as C-$S^Q{}_{10}$-K_{12}. The oligolysine K_{12} realises binding to a nucleic acid template via non-specific electrostatic interactions. The proteins self-assemble into rod-shaped virus-like particles (VLPs) as a consequence of the stacking of multiple $S^Q{}_{10}$ blocks. The $S^Q{}_{10}$ block consist of 10 repeats of the sequence (GAGAGAGQ), a motif inspired by amino acid sequences found in natural silks. Finally, the C block, that consists of a collagen-inspired $(\mathrm{GXaaYaa})_{132}$ sequence, adopts a random coil conformation and forms a flexible corona around the rod-shaped VLPs that provides colloidal stability and protects the encapsulated nucleic acid template.

In this thesis we studied the self-assembly and co-assembly with a nucleic acid template of a number of variants of the previously designed artificial viral capsid proteins, with alternative sequences, in order to obtain a better understanding of how the sequence-design of these proteins affects their physical behaviour. In addition, we studied the cellular uptake of the VLPs *in vitro*, such that in future studies we can also improve the protein design with respect to the ability of the VLPs to deliver nucleic acids into cells.

First, we changed the production host of the artificial viral capsid proteins from the yeast *Pichia pastoris* to the bacterium *Escherichia coli*. This was expected to facilitate the fast production of larger numbers of artificial viral capsid protein variants. As the collagen-inspired C block is known to be poorly produced in *E. coli*, this block was replaced by another polypeptide sequence: the elastin-inspired sequence $E^S{}_m = (\mathrm{GSGVP})_m$. Like the C block, the $E^S{}_m$ block adopts a random coil structure.

In Chapter 2 we addressed the self-assembly behaviour of artificial viral capsid proteins that lack the binding block. The emphasis was on elucidating the effect of changing the lengths of the flexible amorphous block and the silk-inspired crystallisable block on (template-induced) assembly. A series of nine $E^S{}_m$-$S^Q{}_n$ diblock proteins with varying lengths of the $E^S{}_m$ and $S^Q{}_n$ blocks was produced in *E. coli*, where $m = 40$, 60 or 80 and $n = 6$, 10 or 14. Using this set of proteins, we found that a minimum length of the crystallisable silk-like $S^Q{}_n$ block ($n \geq 10$) was required for self-assembly of these proteins. For diblocks with an $S^Q{}_n$ block length just above this minimum, self-assembly was in addition sensitive to the length of the amorphous $E^S{}_m$ block, where increasing lengths of the amorphous

block suppressed self-assembly of the crystallisable $S^Q{}_n$ block. We conclude that for template-induced assembly proteins should have a crystallisable $S^Q{}_n$ block length just above the minimum required length for self-assembly, as well as long, hydrophilic, amorphous blocks. This minimises the self-assembly of the proteins in absence of template-binding while still allowing for self-assembly at the high local concentrations on templates, induced by binding blocks.

Next, in Chapter 3 we kept the block lengths constant, and compared self-assembly of artificial viral capsid proteins with amorphous *C* blocks as colloidal stability block, with artificial viral capsid proteins with elastin-inspired amorphous $E^S{}_m$ blocks of the same length. By comparing the behaviour of the original *C*-$S^Q{}_{10}$-K_{12} protein with the newly produced $E^S{}_{80}$-$S^Q{}_{10}$-K_{12}, we found that, in addition to the length of the colloidal stability block, also the chemical details of this block affect the self-assembly behaviour of the artificial viral capsid proteins. Specifically, the highly hydrophilic *C* blocks seem to suppress self-assembly of the proteins in the absence of templates to a greater extent than the less hydrophilic $E^S{}_{80}$ blocks. Co-assembly with a DNA template, on the other hand, was largely unaffected by changing from the *C* to the $E^S{}_{80}$ colloidal stability blocks. The hydrophilicity of the amorphous block is therefore a key design parameter to favour templated over non-templated assembly of artificial viral capsid proteins.

In Chapter 4 we developed more variants of the artificial viral capsid proteins by changing the nature of the binding block. The aim was to develop a capsid protein that only encapsulates DNA templates that have a so-called "packaging signal", a sequence element that strongly promotes nucleation of fibrils. Except for better mimicking natural viruses, this could also facilitate the encapsulation of specific templates within complex environments such as cells, and the development of the artificial viral capsid proteins towards efficient nucleic acid delivery vehicles by using directed evolution approaches. We designed a new triblock protein ($E^S{}_{80}$-$S^Q{}_{10}$-ZnF) in which we replaced the non-specific K_{12} binding domain with the sequence-specific DNA binding domain ZnF, which is based on a zinc finger found in the *Drosophila melanogaster* GAGA factor. We produced the protein in *E. coli*, but were not able yet to purify a fully intact protein. Nevertheless, the work in this chapter is a step towards the generation of an artificial viral capsid protein for which VLP assembly and specific encapsulation of a template is induced by the recognition of a packaging signal.

Finally, in Chapter 5 we investigated the ability of VLPs, formed by co-assembly of the artificial viral capsid proteins *C*-$S^Q{}_{10}$-K_{12} with mRNA, to enter HeLa cells and deliver their nucleic acids into the cytoplasm. We also investigated whether attaching cell-binding peptides to the ends of the *C* block would promote cellular uptake and hence delivery. To track the VLPs during entry, we used fluorescently-labelled *C*-$S^Q{}_{10}$-K_{12} proteins and mRNA templates. We found that the VLPs entered HeLa cells presumably via endocytosis. Modifying the VLPs with RGD cell-binding peptides, resulted in only a small increase in cell-binding and uptake of the VLPs. We assume this limited effect is due to the shielding effect of the flexible amorphous *C* block and may be solved by placing the RGD peptide on long, stiff linkers. We also concluded that further improvements of the design of the artificial viral capsid proteins may be required to also enhance endosomal escape and unpacking of the

VLPs, in order for these particles to really become useful as delivery vehicles.

This thesis concludes, in Chapter 6, with a general discussion of the work presented in this thesis, and with some recommendations on how to further improve the design of these artificial viral capsid proteins. Due to their modular design they can be optimised relatively easily for specific applications, such that they indeed have the potential to be used in a broad range of applications.

List of publications

This thesis:

- **L. Willems**, S. Roberts, I. Weitzhandler, A. Chilkoti, E. Mastrobattista, J. van der Oost and R.J. de Vries, *Inducible fibril formation of silk-elastin diblocks*. (Submitted; Chapter 2)
- **L. Willems**, L. van Westerveld, S. Roberts, I. Weitzhandler, C. Calcines Cruz, A. Hernandez-Garcia, A. Chilkoti, E. Mastrobattista, J. van der Oost and R.J. de Vries, *The nature of the amorphous hydrophilic block affects self-assembly of an artificial viral coat polypeptide*. (Submitted; Chapter 3)

Other work:

- O. Wicht, W. Li, **L. Willems**, T.J. Meuleman, R.W. Wubbolts, F.J.M. van Kuppeveld, P.J.M. Rottier and B.J. Bosch, *Proteolytic activation of the porcine epidemic diarrhea coronavirus spike fusion protein by trypsin in cell culture*, Journal of Virology 88(14): 7952-7961 (2014).

Acknowledgements

There was one place I did not want to go to for a PhD project: that Agricultural School in the middle of nowhere. Despite this thought, in November 2014 I did start as a PhD candidate at Wageningen University in the group of Physical Chemistry and Soft Matter (PCC) and in the laboratory of Microbiology (MIB). Now 4.5 years later, I am happy for that decision as I was able to work on a challenging project to design an artificial virus in a lab with amazing colleagues. I very much enjoyed my time here in Wageningen and I would like to thank several people for their scientific, mental and social support during my PhD project.

At first, I want to thank my (co-)promotors, **Renko**, **John** and **Enrico**, for giving me the opportunity to work on this exciting project and for the support you gave me throughout these past 4.5 years. I very much appreciate your time and input during our ECHO project meetings. I especially want to thank my daily supervisor **Renko**. I really enjoyed our (bi-)weekly meetings during which we discussed my results and came up with plans for future experiments. With these meetings you helped me to keep smiling throughout my PhD, by keeping me motivated when experiments had failed again or by showing me that the data wasn't as bad as I thought they were. It was interesting to see how these meetings worked in two directions, where you could teach me a lot about physical chemistry and where I was able to show you more about microbiological research. I also want to thank your wife and kids for the wonderful dinners we have had with the Protein Materials subgroup at your house. This was a great way to get to know each other a little better. I'm afraid, though, that for the next dinner you'll have to search for a new person to lit the fireplace.

To all my (former) colleagues at PCC, thank you for your input and suggestions during my PhD, for the scientific discussions we have had and most of all for the wonderful atmosphere that will be very difficult to exceed. In particular, I want to thank the following people:

Mara, you definitely belong to one of the most important people at PCC. Thank you for all your help in administrative issues throughout my time at PCC, for your time and support with any personal issues, and for the "gezelligheid" during coffee breaks and social activities. And of course, thank you for helping me to stay fit by booking a spinning-bike for me every Tuesday. **Josie**, besides arranging our administrative matters, you would always sense it if there were any other (personal) issues. Thank you for all your support and advises. **Anita**, thanks to you I was able to run all my experiments without having to think about money. Also, lots of thanks for the conversations during coffee breaks and for your interesting approaches to win games. **Bert**, thank you for always taking care of our hungry stomachs by regularly filling up the candybox with the best sweets and chocolates. I'm afraid I will have to recover from a candy addiction after leaving PCC.

Jasper, **Joris**, **Joshua**, **Marleen**, **Mieke**, **Thomas** and **Frans**, thank you for

your scientific input during groupmeetings and your social input during our yearly labtrips and the occasional drinks. **Frans**, thank you also for your help in fine-tuning my propositions, which led to their approval in only a single try, and for the lovely bbq's after the yearly Veluweloop.

Hannie, **Diane**, **Remco** and **Anton**, thank you for all your technical support and the wonderful conversations during coffee breaks and the many social activities. Special thanks to **Hannie** for the fun and sometimes serious conversations in the AFM- and preparation-lab. You really helped me to get an idea of what I would like to do after I finish this PhD project.

Joanne, you became my office-mate when you started at PCC about 3.5 years ago. Although from time to time a bit talkative, you were a very pleasant, caring and supportive office-mate. You were always willing to listen to my frustrations about work or to talk about personal stuff. I am very happy that you are my paranimph and that you yourself are also very much on track to finish your own PhD on time. **Ruben H.**, many thanks for taking me along to the coffee breaks, lunches, Friday drinks and game nights when I had just started at PCC. You and **Ties** made me feel very welcome. After we moved to the Helix building, you joined Joanne and me in the office. Except for being a very pleasant office-mate who made sure that Joanne and me put our frustrations into perspective, I also owe you many thanks for your help with LaTeX, which resulted in the transfer of my thesis from Word to LaTeX in only a single day (for which I thought it would take me at least a whole week). **Jessica** and **Gomathi**, you later also became part of office 6.072. Thank you for the serious and sometimes not so serious conversations. Thank you for being such pleasant office-mates.

To everyone of the **Protein Materials Group**, I really enjoyed the scientific discussions during our weekly meetings and in the lab. Thank you also for the wonderful dinners at Renko's house. **Fabiola** and your husband **Victor**, thank you so much for inviting us to your wedding in Mexico and for sharing this special moment with us. Special thanks also to my Mexico travel-mates **Nicolò** and **Dana**. Despite the upcoming deadline for handing in my PhD thesis, I could really relax during that holiday. Those three weeks in Mexico were amazing.

Raisa, it has been amazing to have you as a colleague. Your jokes and your effort to organise drinks and parties are definitely an important factor in the great atmosphere at PCC. I am happy that you are my paranimph and maybe you can help me to turn the day that I defend my thesis into a party as legendary as those Christmas drinks. Just make sure we leave the tables alone. **Marco**, thank you for your hospitality, both in Wageningen and at your family house in Italy. **Niek**, I think we have organised an amazing and successful P(h)D weekend in 2018. Thanks for being such a great co-organising. **Bob**, thank you for the well-timed chats at work in the weekends while I was working a crazy amount of hours a week to finish my thesis in time. **Cooking Mamas**, thank you all for the lovely food that you brought for our lunches and for the new recipes that I discovered.

I should also not forget a few of my former colleagues. **Inge** my co-author and -producer of several PCC "stukjes", **Marcel** my partner in crime for the Friday Fries from Gerard's Corner and **Wolf** my indoor and sometimes also outdoor cycling buddy. Thank you for creating such a wonderful atmosphere at PCC.

And finally, my students **Myrthe**, **Larissa**, **Martin** and **Maarten**. You all worked on different topics within my PhD project and thereby helped us to understand more about the development of an artificial viral capsid protein. I really enjoyed supervising you. **Rob**, you were not really my student and also didn't really need my supervision, but I am happy that I could help you out sometimes in the lab and I am grateful for our scientific discussions.

To my collaborators: **Ashutosh Chilkoti**, thank you for giving me the opportunity to work in your lab at Duke University for three months. It was a great experience. **Stefan** and **Isaac**, thank you for helping me out in the lab during those three months and for the scientific discussions. I also want to thank my collaborators at the Radboud UMC in Nijmegen, **Roland** and **Omar**, for the interesting scientific discussions and the welcoming atmosphere during my visits to your lab.

To my colleagues at **BacGen**, thanks a lot for your input when I had issues with the cloning or expression of my proteins, for the scientific discussion after BacGen meetings, and for the fun during BBQ's and social activities.

To **BioNT**, thank you all for the great time at the Dreijen. I really enjoyed all the drinks and BBQ's.

Naast de vele collega's wil ik ook een aantal vrienden en mijn familie bedanken:

Judith en **Lisa**, mijn lieve shorttrack-matties. Het is heerlijk om alle gedachten rondom een PhD project even van je af te kunnen laten glijden op het ijs. Helaas staan we niet zo vaak meer samen op het ijs en zien we elkaar niet heel regelmatig, maar het is geweldig om te merken dat ik altijd bij jullie terecht kan en dat we gewoon lekker verder kletsen waar we gebleven waren. Ik kijk uit naar onze volgende appeltaart-momenten, en wie weet staan we nog wel weer een keer met z'n drieën in onze onesies op het ijs.

Daniëlle, **Susanne**, **Anouk** en **Albert**, de regelmatige etentjes, de geweldige vakantie in Kroatië, en het heerlijke weekendje in de Ardennen zijn heel waardevol geweest tijdens mijn PhD. Suus en Anouk, het was ontzettend fijn om frustraties en problemen rondom een PhD met jullie te kunnen delen. Danielle en Albert, dank voor jullie steun als wij weer eens aan het klagen waren, maar vooral ook voor de afleiding. Ik kijk er naar uit om, nadat we allen klaar zijn met onze PhD, weer samen op vakantie te gaan en daarmee alle promotie-perikelen achter ons te laten.

Sophie en **Erik**, jullie waren echt geweldige huisgenoten aan de Churchilllaan in Utrecht: samen "Wie is de mol?" kijken, spelletjes spelen en raadsels oplossen tijdens het koken. Ik ben blij dat we, ook nadat we ons huisje hebben moeten verlaten, nog contact hebben weten te houden. De raadsels, Escape Rooms en spelletjes zijn altijd perfecte middelen geweest om even over wat anders na te denken dan aan de obstakels en raadsels rondom mijn PhD project. Wellicht zijn alle raadsels, Escape Rooms en spelletjes ook wel goede oefeningen geweest voor het oplossen van obstakels en raadsels rondom mijn PhD project.

Arjan, **Brenda**, **Ruben W.** en **Rida**, en natuurlijk mijn lieve neefje en nichtje **Morris** en **Milou**. Wat een geweldige familie heb ik toch. Jullie interesse in mijn onderzoek heeft ervoor gezorgd dat ik altijd ben blijven nadenken over waarom

we bepaalde experimenten eigenlijk uitvoeren. Tegelijkertijd zorgde die nuchter blik van jullie ervoor dat ik nooit te lang over problemen rondom mijn promotie heb lopen piekeren, en de discussies over allerlei onderwerpen dat ik ook zo nu en dan eens over wat anders heb nagedacht.

Papa, ik weet nog dat je met mij mee bent geweest naar de open dag voor de bachelor Biomedische Wetenschappen in Utrecht. Je hebt helaas niet meer mee kunnen maken dat ik ook echt mijn middelbare school diploma heb gehaald en dat ik ben begonnen aan deze studie. Nu, 11 jaar later, heb ik mijn bachelor en master diploma op zak en mag ik mijn proefschrift verdedigen. Helaas moet ik jou daarbij missen.

Lieve **mama**, jij staat altijd voor ons allemaal klaar, zelfs wanneer je zelf niet helemaal lekker in je vel zit. Dankjewel voor jouw steun en medeleven als het tijdens mijn PhD niet helemaal liep zoals gepland, en voor de adviezen op alle vlakken van het leven die je mij meegeeft. Ik ben blij dat ik dit belangrijke moment van het verdedigen van mijn proefschrift met jou kan delen.

About the author

Lione Willems was born on the 22nd of April 1990 in Leiden, the Netherlands. She grew up in Geldermalsen and obtained her high school diploma (VWO) in 2008 at Lek en Linge in Culemborg. That same year she started her bachelor studies in Biomedical Sciences at Utrecht University from which she graduated (*cum laude*) in 2011. Interested in pathogens, she continued her studies at Utrecht University with a master in Infection and Immunity. She did two virology internships at Utrecht University and the University Medical Centre Utrecht, and in addition went to the United Kingdom for an internship of four months on protein quality control at the University of Manchester. She obtained her master degree in 2014. In November 2014, Lione started as a PhD candidate at Wageningen University in the group of Physical Chemistry and Soft Matter and in the laboratory of Microbiology. Under supervision of Dr. Renko J. de Vries and Prof. Dr. John van der Oost of Wageningen University and of Prof. Dr. Enrico Mastrobattista of Utrecht University, she worked for four years on the design of a protein that mimics the capsid proteins of viruses. She performed part of that work in the group of Prof. Dr. Ashutosh Chilkoti at Duke University in the United States, where she stayed for three months in 2015. The results are presented and discussed in this thesis.

As a next step in her career, Lione decided to leave the science to others and to continue as an advisor in laboratory design in the company dr. heinekamp Benelux in De Meern. She has started on the 1st of May and will consult on how to design new labs or how existing ones can best be renovated.

Completed training activities

Discipline specific

GRC "Physical Virology"	Ventura (US)	2015
Training in the lab of Prof. A. Chilkoti	Durham (US)	2015
Dutch Biophysics Days (NWO)‡	Veldhoven	2015
CHAINS (NWO)‡	Veldhoven	2015
NVGCT Spring Meeting	Lunteren	2016
Soft Matter course (WUR-PCC)	Wageningen	2016
Atomic Force Microscopy Workshop	Porto (PT)	2016
GRC & GRS "Bioinspired Materials"‡	Switzerland	2016
CryoTEM course (TU Eindhoven)	Eindhoven	2016
Dutch Materials (4TU.HTM)†	Utrecht	2016
Dutch Soft Matter Meeting	Wageningen	2016
CHAINS (NWO)‡	Veldhoven	2016
GRC & GRS "Physical Virology"†‡	Lucca (IT)	2017
CHAINS (NWO)‡	Veldhoven	2017

General

Journal Club (PCC)	Wageningen	2014-2018
PhD week (VLAG)	Baarlo	2015
Photoshop & Illustrator CS6 Workshop	Rotterdam	2015
In'to Dies Pitch (WUR)†	Wageningen	2016
Scientific Publishing (WGS)	Wageningen	2016
Teaching and Supervising Thesis students (WGS)	Wageningen	2016
Career Orientation (WGS)	Wageningen	2017

Optionals

Preparation of Research Proposal	Wageningen	2014
Colloquia & Seminars	Wageningen	2014-2018
Group meetings (PCC & MIB)	Wageningen	2014-2018

† oral presentation, ‡ poster presentation

This research was financially supported by the Netherlands Organisation for Scientific Research (NWO).

Cover design by Lione Willems and Ruben Higler

Printed by: Digiforce || ProefschriftMaken


ATG