Protein-based Polymers that Bind to DNA

Design of Virus-like Particles and Supramolecular Nanostructures

Armando Hernández García

Thesis committee

Promotor
Prof. Dr M.A. Cohen Stuart
Emeritus Professor of Physical Chemistry and Colloid Science
Wageningen University

Co-promotors
Dr R.J. de Vries
Associate professor, Laboratory of Physical Chemistry and Colloid Science
Wageningen University

Prof. Dr P. van der Schoot
Assistant professor, Department of Applied Physics, Theory of Polymers and Soft matter
Eindhoven University of Technology

Other members
Prof. Dr A. Chilkoti, Duke University, North Carolina, United States of America
Prof. Dr J.L.M. Cornelissen, Twente University, Enschede, the Netherlands
Dr E. Mastrobattista, Utrecht University, the Netherlands
Prof. Dr J. van der Oost, Wageningen University

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

Protein-based Polymers that Bind to DNA

Design of Virus-like Particles and Supramolecular Nanostructures

Armando Hernández García

Thesis
submitted in fulfillment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr M.J. Kropff,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 17th of January 2014
at 11 a.m. in the Aula.

Armando Hernández García
Protein-based Polymers that Bind to DNA
Design of Virus-like Particles and Supramolecular Nanostructures
254 pages

PhD thesis, Wageningen University, Wageningen, NL (2014)
With references, with summaries in English and Dutch

ISBN 978-94-6173-823-3

To my parents and family.

Contents

Part IV. Main concepts, summary and general discussion

Chapter 1

Introduction

1.1 Protein-based Polymers

1.1.1. Material Science Meets Molecular Biology. The field of protein-based polymers was born at the cross-roads of Material Sciences and Molecular Biology.[1, 2] In particular, polymer scientists were looking for technologies to create novel, improved polymeric materials and found such technologies in the growing area of Molecular Biology and DNA recombinant technology[2-8] (Figure 1.1). A major impetus for further development has been the development of fast and efficient (commercially available) DNA synthesis techniques.

The term "protein-based polymers" first came up in connection with the pioneering work of David A. Tirrell in University of Massachusetts (later in CALTECH)[9] and Joseph Capello in the company Protein Polymer Technologies Inc. (San Diego, California)[3,10] around 1990. Although previous work from Doel, MT, et. al. in 1980 reported the production of an artificial repetitive sequence of poly(L-aspartyl-L-phenylalanine) amino acids by inserting a synthetic gene sequence in *E. coli,*[11] that work did not point out the possibility to produce protein-based polymers in this way.

1.1.2. Two Paradigms in One Molecule. Protein-based polymers are a new class of macromolecular materials obtained by genetically encoded biosynthesis of repetitive or tandem peptide sequences.[2-4,6-10,12,13] Recombinant production through ribosomal synthesis allows for production from renewable resources, provides precise and absolute control over stereochemistry, sequence, and chain length. These are features that are required more and more often for creating new functional polymers, and that cannot (all) be provided through the chemical synthesis of polymers. The precise control over primary sequence of these polymers, ultimately gives control over the physical, chemical and biological properties of materials made out of them, such as their mechanics, supramolecular organization, biocompatibility and biodegradability. Because of their recombinant

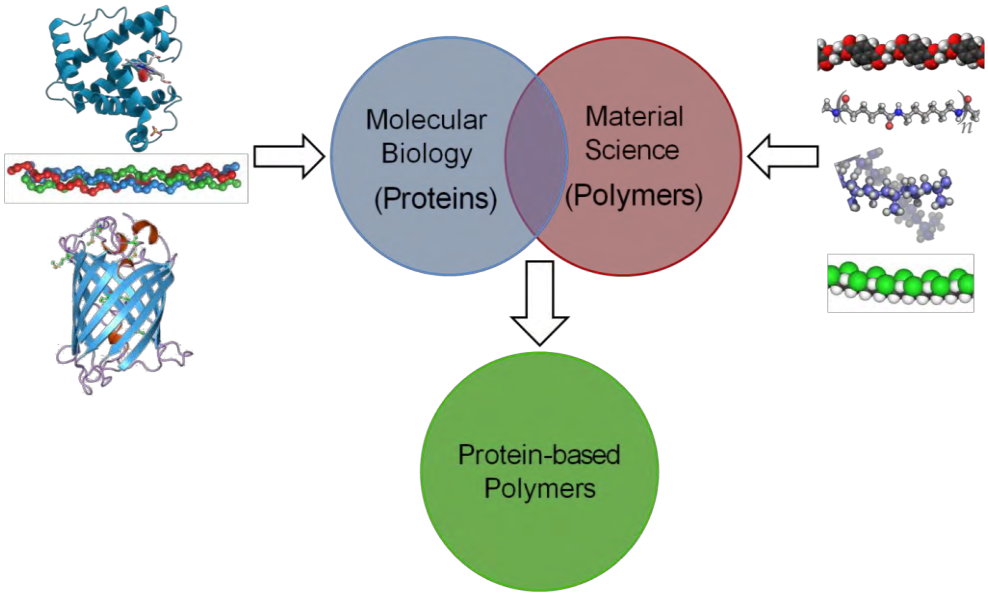


Figure 1.1. The field of protein-based polymers was born at the cross-roads of Molecular Biology and Materials Science. It combines approaches to protein- and polymer science, and properties of both proteins and polymers to create novel functional materials.

origin, protein-based polymers are composed of the 20 natural amino acids units. However, a range of technologies are now being developed that also allow for the incorporation of unnatural amino acids.[14-19] One of the most prominent advantages of using amino acids to build polymers is that one can borrow basic functional sequences from natural proteins, such as blocks or domains with defined secondary structures (α-helix, β-sheet, coiled-coils, etc) that can exist independently. Also, protein-based polymers can be combined with fully folded proteins or protein-domains (with more complicated secondary and tertiary structure) to create polymer-protein hybrids.[20-23] The flexibility to incorporate and repeat functional peptide sequences in protein-based polymers allows for the creation of new functionalities and new materials by combining multiple simple

FEATURE	Proteins	Polymers	Protein-based Polymers
Control of primary sequence?	👍	👎	👍
Control of monodispersity?	👍	👎	👍
Control of stereochemistry?	👍	👎	👍
Ease of production?	👎	👍	👎
High amounts of production?	👎	👍	👎
Supramolecular organization?	👍	👍	👍
Folded domains?	👍	👎	👍
Simplicity?	👎	👍	👍
Stability (heat, biological, etc)?	👎	👍	👍
Biodegradability?	👍	👎	👍
Biocompatibility?	👍	👎	👍
Environmentally-friendly?	👍	👎	👍
Availability of substrates?	👍	👍	👍
Low cost?	👎	👍	👎

👍 Yes or high 👎 No or low

Figure 1.2. Comparative table between properties of proteins, polymers and protein-based polymers.

block with properties that have already been established independently.[3,24] Potential problems in making protein-based polymers arise from the fact that large and repetitive sequences can be genetically unstable, or are otherwise not tolerated by some production host cells, and may lead to DNA deletion, or proteolytic degradation of products. One possible alternative synthesis method that deserves mention here is the chemical synthesis of long polypeptides. This is nowadays also used to create various types of amino acid copolymers ("polypeptide block copolymers" or "block copolypeptides"), but the chemical synthesis only allows for

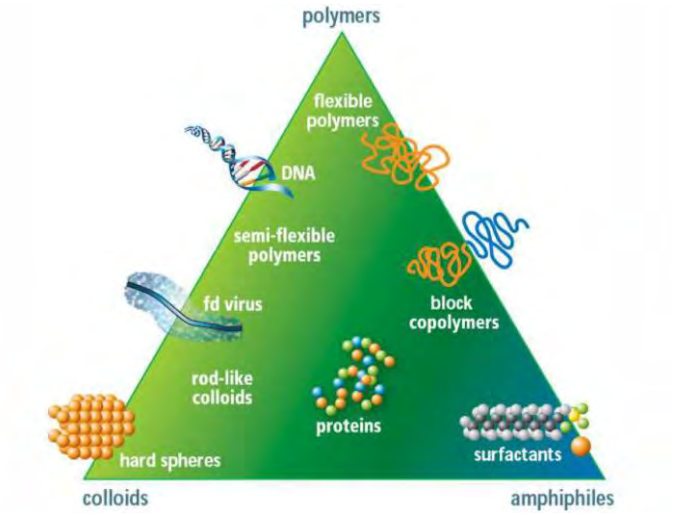


Figure 1.3. The "Soft Matter Triangle". Figure taken from Eur. Phys. J. E. 26, 1-2, (2008) [25].

blocks that are homopolymers (composed of only a single amino acid),[26-28] and therefore is much less versatile than recombinant production. For completeness, Figure 1.2 compares a number of features of (natural, folded) proteins with those of typical polymers produced through synthetic chemistry, and with those of protein-based polymers.

In a way, protein-based polymers fuse two paradigms: the world of natural proteins and the world of synthetic polymers (including block copolymers).[8,29] Protein-based polymers not only have the potential to make materials that combine the very different properties of proteins and polymers, but may also be expected to lead to materials with new properties that go beyond what can be done with proteins and polymers separately.[3, 9, 10, 30] For another perspective on the position of protein-based polymers in current science and technology, we can take a look at the so-called "soft matter triangle" (Figure 1.3).[25] While this diagram places proteins in the middle, the additional freedom for de-novo designed protein-based polymers means that in fact, protein-based polymers can access a significant fraction of the area of the entire Soft Matter triangle, which illustrates the versatility of these materials.

While at the dawn of "protein-based polymer history", researchers typically used long repeats of single units, over time, people started looking at di-block and multiblock designs. Nowadays, the use of elaborated sequences and the number of blocks used in a chain of a protein-based polymer has increased to the point that their functionality can be very specific and their performance rivals that of sophisticated natural materials.[31-38] At the same time, this increased complexity takes us away from the predictability of polymers, and brings us closer to the fundamental problem that underlies the whole field of protein design: how to reliably predict protein structure and functionality from primary amino acid sequence.[39]

1.1.3 Production of Protein Polymers. Genetic engineering followed by fermentation is the usual procedure to produce a recombinant protein-based polymer. While genetic engineering provides the tools to generate the DNA sequences for the engineered proteins, and to insert them in the final host, the fermentation step harnesses the biological systems in order to biosynthesize the protein using simple and cheap substrates.[1, 3, 4, 9, 12, 29, 32, 38, 40-47]

Since the primary sequence of amino acids is dictated by a nucleotide sequence, the first steps are to design the amino acid sequence of the protein-based polymer, to translate that into a nucleotide sequence optimized for the particular host, and then to actually produce these polynucleotide (DNA) molecules. Several techniques have been developed tp synthesize artificial genes for protein polymers. These include "concatemerization",[3, 4, 9, 30, 40] "recursive directional ligation" (RDL),[3, 4, 9, 12, 30, 32, 40, 42, 48-50] and the recently developed "overlap extension rolling circle amplification" (OERCA).[51, 52] The first one is the self-ligation of a repetitive basic unit through cohesive ends that creates in a single step multimers of the desired sequence, but without precise control in the degree of polymerization. This was the initial approach to produce the genetic sequences of the protein based polymers in 1990.[3, 9] The second one, RDL, was developed in 2002 and consists of a stepwise addition of the basic sequence unit until the final desired length is obtained.[12] If the desired length is very large, this approach therefore requires many rounds of

addition. In 2011 the OERCA[51] method was proposed, which allows for the parallel synthesis of many sequences with a large polymerization degree. In this PhD thesis we have used RDL, since the number of stepwise additions of sequences was not very large.

The most often used host organisms to produce protein polymers are the bacterium *Escherichia coli* and the yeast *Pichia pastoris.* While *E. coli* typically produces between 10-100 mg of product per L of culture, *P. pastoris* can produce between 0.1-1 g per L of cell-free medium. Production in *E. coli* is usually done on a medium based on carbohydrates such as glucose (and is induced, e.g. with IPTG) while for *P. pastoris* the most common approach is to use methanol in the dual role as both a substrate for- and inducer of the biosynthesis.[53]

1.2 Artificial DNA-Binding Proteins

1.2.1. Principles for DNA Binding from Natural Proteins. Natural DNA-binding proteins have evolved for over millions of years to perform a broad range of "DNA transactions": DNA packaging, replication, recombination, restriction and transcription.[54-57] Proteins and protein complexes involved in these transactions can be huge, but often the actual DNA-binding domains themselves are not so large.[58-61] These have been extensively studied to unravel their structures and mechanisms of interaction with DNA. Their interaction with DNA involves molecular complementarity (geometrical and chemical) that can lead to sequence-specific, sequence-aspecific or structure-specific DNA binding.[54-57, 60, 62-67] Very often, basic amino acids are found in DNA-binding domains that interact with DNA through electrostatic interactions with the phosphate backbone.[56, 64, 66, 68, 69] Specificity (for sequence or structure) is more often dictated by additional hydrogen bonding (sometimes water-mediated) and/or hydrophobic interaction with the minor and/or major grooves of the DNA structure.[54-57, 62-66, 70, 71] Many DNA-binding domains of natural proteins fall within one of a small number of characteristic structural motifs.[56, 57, 60, 62-64] Some of these are illustrated in Figure 1.4.

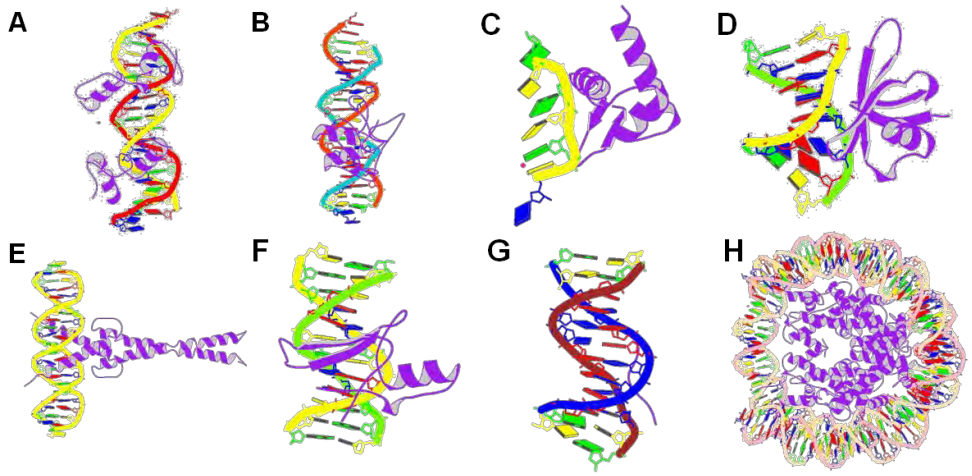


Figure 1.4. Examples of natural proteins with structural motifs for DNA-binding A) Zinc fingers (PDB code: 3UK3). B) Zinc finger from transcription factor GATA-3 (PDB code: 4HC9). C) Z-alpha domain binds to Z-DNA (PDB code: 4KMF). D) Sso7d 7kD DNA-binding domain "Histone-like" protein from *archaea* microorganisms (PDB code: 1BNZ). E) Basic/helix-loop-helix/leucine zipper DNA-binding domain of the transcription factor Max (PDB code: 1AN2). F) Methyl-CpG-binding domain (PDB code: 4LG7). G) A-T hook peptide (PDB code: 3UXW). H) Biological assembly of Histone "H4" (PDB code: 2PY0).

1.2.2 Principles of DNA-Binding and Formation of Viral Structures from Viral Coat Proteins. This thesis deals in part with mimicking viral coat proteins, hence we here wish to pay special attention to this special class of proteins, which often bind nucleic acids.[72] They mediate the condensation of nucleic acids (either DNA or RNA) into highly regular and stable supramolecular structures, generally icosahedral or cylindrical, that are composed of multiple copies of one or few different types of coat proteins.[73-75] The analysis of in-vitro viral particle formation, for example for the Tobacco Mosaic Virus (TMV),[73, 76, 77] has led to insights in the requirements that coat proteins need to fulfill in order to form such exquisite nanostructures.

Within the TMV coat protein; one can distinguish at least three regions that each encodes for a major physical-chemical functionality (see Figure 1.5): a basic region that binds to DNA, a typically hydrophobic region that leads to protein-protein

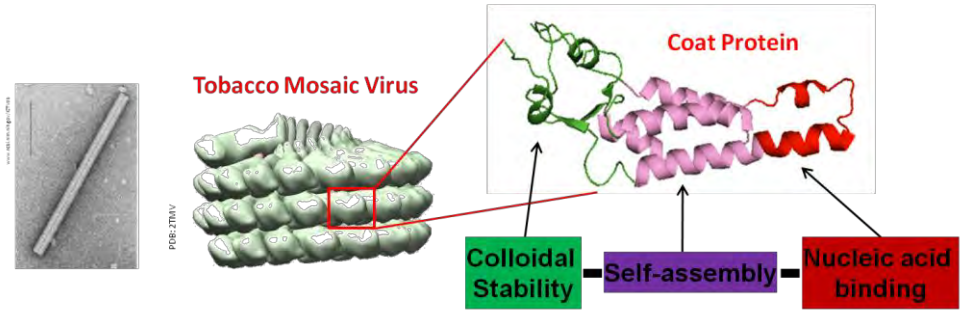


Figure 1.5. Identification of regions that encode important physico-chemical functionalities in the TMV coat protein.

interactions and self-assembly, and finally a hydrophilic outside region that provides shielding and regulates interactions of the particle with the outside world. For TMV and other viruses, the co-assembly and packing of the viral RNA with the coat protein is highly cooperative and is triggered by an allosteric change in the protein conformation upon binding.[69, 77]

Inspired by these insights we hypothesize that if we put together in a protein-based polymer molecule the following three physicochemical functionalities, the protein-based polymer will tend to form virus-like supramolecular assemblies:

(a) A basic block that binds DNA through electrostatic complexation,

(b) A block that gives rise to lateral cooperative self-assembly between proteins bound to DNA (ideally via an allosteric conformational change upon binding the DNA)

(c) A shielding block that gives colloidal stability to the self-assembled particles, such that they do not aggregate.

Below we consider each of these physicochemical functionalities in somewhat more detail.

Electrostatic complexation occurs between two molecules that are opposite charged and is often mostly entropically driven via the release of the many counter ions that were bound prior to complexation. For a DNA-protein complex, counterions are released from the positively charged amino acids such as lysine, histidine and arginine on the protein and from the negatively charged phosphate groups of the nucleic acids (DNA or RNA).

Self-assembly is the process where multiple weak bonds between building blocks (molecules, particles) leads to the (reversible) formation of larger structures. The weak bonds can be any weak noncovalent interaction such as hydrogen bonds, π- π stacking, metal coordination, electrostatic forces, Van der Waals forces or hydrophobic effects.[78-84] In many (but not all) cases, self-assembly of supramolecular structures is cooperative,[81, 84, 85] meaning that multiple (weak) bonds are formed simultaneously, which has various general consequences[86, 87] such as sharp transitions as a function of control parameters (e.g. the critical micelle concentration), and coexistence between distinctly different species (e.g. surfactant micelles and surfactant monomers). In this way, cooperativity evades energy, and time barriers, leading into preorganization of complex structures. Cooperativity is an emergent property that can be encoded in supramolecular building blocks through programming their weak noncovalent interactions.[87, 88] Indeed nature exploits cooperativity in many different ways: in protein folding, protein allostery, at the cellular level, etc.[85, 86] Often, cooperativity is linked to structural (or allosteric) changes in the building blocks upon establishing an interaction. To encode controlled cooperative self-assembly in protein-based polymer building blocks is challenging.[89] One way to encode lateral cooperative interactions in a protein-based polymer is through the use of β-sheet stacking.[48, 90] There are many reports of amino-acid sequences such as silk-like sequences, but also others that are able to fold into a beta-sheet secondary structure and stack into one-dimension elongated structures.[91-96]

Improving the *Colloidal Stability* of particle dispersions, i.e. preventing their reversible or irreversible aggregation, typically involves modulating the particle

interaction through surface modifications. A very effective general strategy, which is also employed in this thesis, is shielding off any potential attractions with a hydrophilic polymer-brush at the particle surface.

1.3 Protein-based Polymers that Bind to DNA: Applications

This thesis deals with protein-based polymers that bind to DNA, and by doing so, change various DNA properties. Although not the focus of this thesis, we believe that such polymers in the future may have a big impact on a number of application areas where control of DNA (physical) properties is crucial. Also that protein-based polymers will have an important role in the development of general capabilities for molecular manipulation and molecular engineering.[97] Some of these are reviewed below, and shown in Figure 1.6.

1.3.1. Artificial Viruses and DNA Delivery Systems. The design of artificial viruses is highly relevant as an approach to non-viral DNA delivery, and for increasing our capabilities to develop synthetic biosystems. The task of designing artificial viruses that mimic their natural counterparts, though, has been found to be extremely difficult.[98-103] Only in the last decades, with more suitable tools becoming available for engineering of functional polymeric materials, the design of such systems has become a possibility.

Given the many attractive features of protein-based polymers outlined above, protein-based polymer technology may be expected to be highly suitable for designing "artificial viruses", and thus advancing tools for synthetic biology.[39, 104] Indeed, the protein-based polymers can be designed to do all the things previously considered as being "virus mimetic"[104-112] such as co-assembly with nucleic acids into particles within a certain size range, that in addition can be targeted to bind to and enter in cells, etc.

The protein-based polymer approach bridges the two current approaches that have been largely separate from each other: one focused on "mimicking" the functional biological activity, and the other one focused on mimicking the

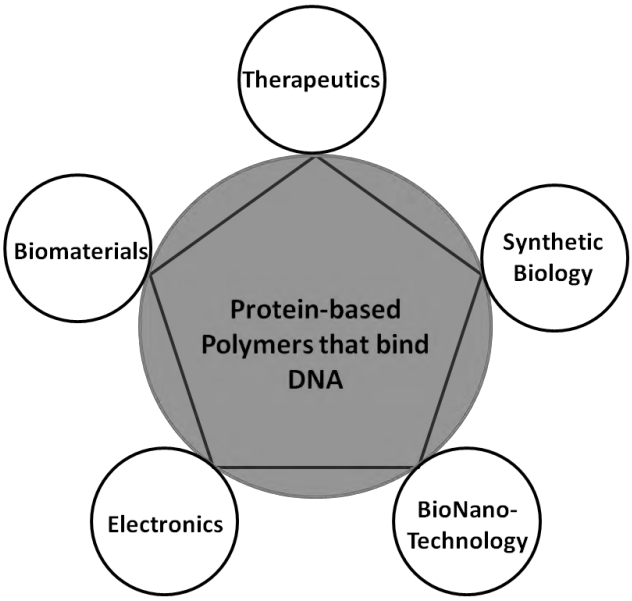


Figure 1.6. Practical applications where Protein-based polymers that bind DNA can have impact.

structural features and self-assembly. In addition, protein-based polymers can be carefully designed to encode cooperativity, which is a key parameter for viral self-assembly that has hardly received any attention in previous designs for artificial viruses. The latter negligence may in fact be related to the lack of molecular tools to precisely encode and tune cooperativity in designs for self-assembling polymers. Being able to precisely encode and tune cooperativity is not only important for mimicking viruses. Indeed, given the crucial role of cooperativity in biology, it may be expected to be also crucial for ultimately being able to design synthetic live-like systems.[39, 85, 86]

1.3.2. Optical Mapping and Sequencing of Single DNA Molecules. A very active field of DNA technology is that of single DNA molecule sequencing. The approaches that are being developed are all highly sophisticated and complex.[113] For all of them the control of DNA physical properties is crucial, and we expect that dedicated DNA-binders may contribute here. One example is the optical mapping of single DNA molecules that allows for the localization of marker

sequences in very long (genome-sized) DNA molecules, relies on being able to stretch DNA. Specific DNA binders that increase the DNA stiffness can be of significant value for that application, as we illustrate in chapter 5.

1.3.3 Protein-based Polymers for DNA-based Nanomaterials. DNA is a versatile building block for building materials at the nanoscale level. It has a well-defined recognition language based in the Watson-Crick complementary pairing that can be harnessed to program the folding of nanostructures with particular functionalities. The programmed hybridization of a template or scaffold sequence (usually ssDNA from M13 virus) and short "helper" or "staple" strands, is being used to create an ever growing collection of 2D and 3D nanostructures that can be used in a wide range of applications.[114, 115] Applications include mechanical devices that walk, cages with lids, nanotubes, sensors, nanopores, metallic nanowires, and "delivery robots".[116-121] DNA-binding protein-based polymers could be used in this field to coat and protect DNA nanostructures, to modulate their physical properties, to locate functional moieties at specific places, etc.[115] Another possibility is to use the monodisperse and precisely-defined DNA structures as a scaffold for the multivalent display of other groups.[122] This application could also benefit from coating the DNA with protein-based polymers.[32, 123-125] A final application could be to modulate DNA structures that are to be used as templates in materials science (e.g. to construct metal, or other types of inorganic nanoparticles or nanowires for sensing, electronic materials, nanomedicine, etc.) As in every new field, the limits of these technologies have not been established yet. We believe a combination with protein-based polymer technology could lead to synergy and offer yet more opportunities.

1.4 Outline of this thesis.

This thesis is divided in three main parts. Each part focuses in a different aspect of the protein-based polymers that bind to DNA. In Part I the interest is to design and produce several protein block copolymers able to complex DNA molecules in order to develop artificial virus-like particles. The design, production and detailed

characterizations are described. In Part II some applications of the previously designed protein block copolymers are depicted. Also new protein polymer designs with improved features for complexing DNA are described and characterized together with some possible applications. Part III deals with the electrostatic complexation of other anionic polyelectrolytes different to DNA, such as negative supramolecular polymers that carry metal ions or synthetic polymers. Also, the self-assembly of designed triblock copolymers without a DNA template is described. Last Part IV makes a summary of the main finding and conclusions of all the previous chapters and a general discussion about the design of protein based polymers for designing artificial viral-like systems the implications. It is also discussed the perspectives of these designed systems.

1.5 Aim of this thesis

The aim of this thesis is to design systematically polypeptide building blocks, that when combined into multiblock polypeptides can co-assemble with DNA into virus-like particles. The polypeptides are intended to mimic key properties of natural viruses. The virus-like particles that we aimed to obtain should be highly ordered, and feature precise control of their nanostructure and nanoarchitecture, as well as displaying as much functional virus properties as possible.

1.6 References

1. Maskarinec, S. A.; Tirrell, D. A., Protein engineering approaches to biomaterials design. *Curr. Opin. Biotechnol.* 2005, *16*, 422-426.

2. Krejchi, M. T.; Atkins, E. D. T.; Waddon, A. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A., Chemical sequence control of beta-sheet assembly in macromolecular crystals of periodic polypeptides. *Science* 1994, *265*, 1427-1432.

3. Cappello, J.; Crissman, J.; Dorman, M.; Mikolajczak, M.; Textor, G.; Marquet, M.; Ferrari, F., Genetic-engineering of structural protein polymers. *Biotechnol. Prog.* 1990, *6*, 198-202.

4. McGrath, K. P.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A., Genetically directed syntheses of new polymeric materials - expression of artificial genes encoding proteins with repeating (AlaGly)3ProGluGly elements. *J. Am. Chem. Soc.* 1992, *114*, 727-733.

5. Kaplan, D. L., Protein-based polymers. *Abstr. Pap. Am. Chem. Soc.* 1995, *209*, 31-PMSE.

6. Zhang, X. R.; Urry, D. W.; Daniell, H., Expression of an environmentally friendly synthetic protein-based polymer gene in transgenic tobacco plants. *Plant Cell Reports* 1996, *16*, 174-179.

7. Urry, D. W., Physical chemistry of biological free energy transduction as demonstrated by elastic protein-based polymers. *J. Phys. Chem. B* 1997, *101*, 11007-11028.

8. Barron, A. E.; Zuckermann, R. N., Bioinspired polymeric materials: in-between proteins and plastics. *Curr. Opin. Chem. Biol.* 1999, *3*, 681-687.

9. McGrath, K. P.; Tirrell, D. A.; Kawai, M.; Mason, T. L.; Fournier, M. J., Chemical and biosynthetic approaches to the production of novel polypeptide materials. *Biotechnol. Prog.* 1990, *6*, 188-192.

10. Cappello, J., The biological production of protein polymers and their use. *Trends Biotechnol.* 1990, *8*, 309-311.

11. Doel, M. T.; Eaton, M.; Cook, E. A.; Lewis, H.; Patel, T.; Carey, N. H., The expression in Escherichia coli of synthetic repeating polymeric genes-coding for poly(L-Aspartyl-L-Phenylalanine). *Nucleic Acids Research* 1980, *8*, 4575-4592.

12. Meyer, D. E.; Chilkoti, A., Genetically encoded synthesis of protein-based polymers with precisely specified molecular weight and sequence by recursive directional ligation: Examples from the elastin-like polypeptide system. *Biomacromolecules* 2002, *3*, 357-367.

13. Rabotyagova, O. S.; Cebe, P.; Kaplan, D. L., Protein-Based Block Copolymers. *Biomacromolecules* 2011, *12*, 269-289.

14. Link, A. J.; Mock, M. L.; Tirrell, D. A., Non-canonical amino acids in protein engineering. *Curr. Opin. Biotechnol.* 2003, *14*, 603-609.

15. Yoshikawa, E.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A., Genetically-engineered fluoropolymers - synthesis of repetitive polypeptides containing p-fluorophenylalalanine residues. *Macromolecules* 1994, *27*, 5471-5475.

16. Kiick, K. L.; van Hest, J. C. M.; Tirrell, D. A., Expanding the scope of protein biosynthesis by altering the methionyl-tRNA synthetase activity of a bacterial expression host. *Angew. Chem.-Int. Edit.* 2000, *39*, 2148.

17. Kwon, I.; Kirshenbaum, K.; Tirrell, D. A., Breaking the degeneracy of the genetic code. *J. Am. Chem. Soc.* 2003, *125*, 7512-7513.

18. Wang, P.; Tang, Y.; Tirrell, D. A., Incorporation of trifluoroisoleucine into proteins in vivo. *J. Am. Chem. Soc.* 2003, *125*, 6900-6906.

19. Connor, R. E.; Tirrell, D. A., Non-canonical amino acids in protein polymer design. *Polym. Rev.* 2007, *47*, 9-28.

20. Woolfson, D. N., The design of coiled-coil structures and assemblies. *Adv.Protein Chem.* 2005, *70*, 79-+.

21. Krejchi, M. T.; Cooper, S. J.; Deguchi, Y.; Atkins, E. D. T.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A., Crystal structures of chain-folded antiparallel beta-sheet assemblies from sequence-designed periodic polypeptides. *Macromolecules* 1997, *30*, 5012-5024.

22. Woolfson, D. N.; Ryadnov, M. G., Peptide-based fibrous biomaterials: some things old, new and borrowed. *Curr. Opin. Chem. Biol.* 2006, *10*, 559-567.

23. Smeenk, J. M.; Otten, M. B. J.; Thies, J.; Tirrell, D. A.; Stunnenberg, H. G.; van Hest, J. C. M., Controlled assembly of macromolecular beta-sheet fibrils. *Angew. Chem.-Int. Edit.* 2005, *44*, 1968-1971.

24. Ulijn, R. V.; Smith, A. M., Designing peptide based nanomaterials. *Chem. Soc. Rev.* 2008, *37*, 664-675.

25. Gompper, G.; Dhont, J. K. G.; Richter, D., A unified view of soft matter systems? *Eur. Phys. J. E* 2008, *26*, 1-2.

26. Deming, T. J., Facile synthesis of block copolypeptides of defined architecture. *Nature* 1997, *390*, 386-389.

27. Deming, T. J., Polypeptide materials: New synthetic methods and applications. *Advanced Materials* 1997, *9*, 299-&.

28. Deming, T. J., Methodologies for preparation of synthetic block copolypeptides: materials with future promise in drug delivery. *Adv. Drug Deliv. Rev.* 2002, *54*, 1145-1155.

29. Cunliffe, D.; Pennadam, S.; Alexander, C., Synthetic and biological polymers-merging the interface. *Eur. Polym. J.* 2004, *40*, 5-25.

30. Fournier, M. J.; Creel, H. S.; Krejchi, M. T.; Mason, T. L.; Tirrell, D. A.; McGrath, K. P.; Atkins, E. D. T., Genetic synthesis of periodic protein materials. *J. Bioact. Compat. Polym.* 1991, *6*, 326-338.

31. Garanger, E.; Lecommandoux, S., Towards Bioactive Nanovehicles Based on Protein Polymers. *Angew. Chem.-Int. Edit.* 2012, *51*, 3060-3062.

32. Hernandez-Garcia, A.; Werten, M. W. T.; Stuart, M. C.; de Wolf, F. A.; de Vries, R., Coating of Single DNA Molecules by Genetically Engineered Protein Diblock Copolymers. *Small* 2012, *8*, 3491-3501.

33. van Hest, J. C. M.; Tirrell, D. A., Protein-based materials, toward a new level of structural control. *Chem. Commun.* 2001, 1897-1904.

34. Zhang, S. G., Fabrication of novel biomaterials through molecular self-assembly. *Nat. Biotechnol.* 2003, *21*, 1171-1178.

35. Langer, R.; Tirrell, D. A., Designing materials for biology and medicine. *Nature* 2004, *428*, 487-492.

36. Diehl, M. R.; Zhang, K. C.; Lee, H. J.; Tirrell, D. A., Engineering cooperativity in biomotor-protein assemblies. *Science* 2006, *311*, 1468-1471.

37. DiMarco, R. L.; Heilshorn, S. C., Multifunctional Materials through Modular Protein Engineering. *Advanced Materials* 2012, *24*, 3923-3940.

38. Frandsen, J. L.; Ghandehari, H., Recombinant protein-based polymers for advanced drug delivery. *Chem. Soc. Rev.* 2012, *41*, 2696-2706.

39. Bromley, E. H. C.; Channon, K.; Moutevelis, E.; Woolfson, D. N., Peptide and protein building blocks for synthetic biology: From programming biomolecules to self-organized biomolecular systems. *ACS Chemical Biology* 2008, *3*, 38-50.

40. Creel, H. S.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A., Genetically directed syntheses of new polymeric materials - efficient expression of a monodisperse copolypeptide containing 14 tandemly repeated -(AlaGly)4ProGluGly-elements. *Macromolecules* 1991, *24*, 1213-1214.

41. Nagarsekar, A.; Ghandehari, H., Genetically engineered polymers for drug delivery. *J. Drug Target.* 1999, *7*, 11-32.

42. Werten, M. W. T.; Wisselink, W. H.; Jansen-van den Bosch, T. J.; de Bruin, E. C.; de Wolf, F. A., Secreted production of a custom-designed, highly hydrophilic gelatin in Pichia pastoris. *Protein Engineering* 2001, *14*, 447-454.

43. Haider, M.; Megeed, Z.; Ghandehari, H., Genetically engineered polymers: status and prospects for controlled release. *J. Control. Release* 2004, *95*, 1-26.

44. Mi, L. X., Molecular cloning of protein-based polymers. *Biomacromolecules* 2006, *7*, 2099-2107.

45. Rodriguez-Cabello, J. C.; Reguera, J.; Girotti, A.; Arias, F. J.; Alonso, M., Genetic engineering of protein-based polymers: The example of elastinlike polymers. In *Ordered Polymeric Nanostructures at Surfaces*, Vancso, G. J.; Reiter, G., Eds. Springer-Verlag Berlin: Berlin, 2006; Vol. 200, pp 119-167.

46. Dutta, N. K.; Truong, M. Y.; Mayavan, S.; Choudhury, N. R.; Elvin, C. M.; Kim, M.; Knott, R.; Nairn, K. M.; Hill, A. J., A Genetically Engineered Protein Responsive to Multiple Stimuli. *Angew. Chem.-Int. Edit.* 2011, *50*, 4428-4431.

47. Kiick, K. L., Biosynthetic methods for the production of advanced protein-based materials. *Polym. Rev.* 2007, *47*, 1-7.

48. Martens, A. A.; Portale, G.; Werten, M. W. T.; de Vries, R. J.; Eggink, G.; Stuart, M. A. C.; de Wolf, F. A., Triblock Protein Copolymers Forming Supramolecular Nanotapes and pH-Responsive Gels. *Macromolecules* 2009, *42*, 1002-1009.

49. Werten, M. W. T.; Teles, H.; Moers, A. P. H. A.; Wolbert, E. J. H.; Sprakel, J.; Eggink, G.; de Wolf, F. A., Precision Gels from Collagen-Inspired Triblock Copolymers. *Biomacromolecules* 2009, *10*, 1106-1113.

50. McDaniel, J. R.; MacKay, J. A.; Quiroz, F. G.; Chilkoti, A., Recursive Directional Ligation by Plasmid Reconstruction Allows Rapid and Seamless Cloning of Oligomeric Genes. *Biomacromolecules* 2010, *11*, 944-952.

51. Amiram, M.; Quiroz, F. G.; Callahan, D. J.; Chilkoti, A., A highly parallel method for synthesizing DNA repeats enables the discovery of 'smart' protein polymers. *Nat. Mater.* 2011, *10*, 141-148.

52. Ding, S.; Wang, X.; Barron, A. E., Protein polymers Gene libraries open up. *Nat. Mater.* 2011, *10*, 83-84.

53. Zhang, W. H.; Bevins, M. A.; Plantz, B. A.; Smith, L. A.; Meagher, M. M., Modeling Pichia pastoris growth on methanol and optimizing the production of a recombinant protein, the heavy-chain fragment C of botulinum neurotoxin, serotype A. *Biotechnology and Bioengineering* 2000, *70*, 1-8.

54. Nelson, H. C., Structure and function of DNA-binding proteins. *Current opinion in genetics & development* 1995, *5*, 180-9.

55. Jantz, D.; Amann, B. T.; Gatto, G. J.; Berg, J. M., The design of functional DNA-binding proteins based on zinc finger domains. *Chemical Reviews* 2004, *104*, 789-799.

56. Shanahan, H. P.; Garcia, M. A.; Jones, S.; Thornton, J. M., Identifying DNA-binding proteins using structural motifs and the electrostatic potential. *Nucleic Acids Res.* 2004, *32*, 4732-4741.

57. Dickey, T. H.; Altschuler, S. E.; Wuttke, D. S., Single-Stranded DNA-Binding Proteins: Multiple Domains for Multiple Functions. *Structure* 2013, *21*, 1074-1084.

58. Angrand, P. O., DNA-binding domains of eukaryotic transcription factors. *M S-Med. Sci.* 1993, *9*, 725-736.

59. Crane-Robinson, C.; Dragan, A. I.; Privalov, P. L., The extended arms of DNA-binding domains: a tale of tails. *Trends Biochem.Sci.* 2006, *31*, 547-552.

60. Singh, D. P.; Kubo, E.; Takamura, Y.; Shinohara, T.; Kumar, A.; Chylack, L. T.; Fatma, N., DNA binding domains and nuclear localization signal of LEDGF: Contribution of two helix-turn-helix (HTH)like domains and a stretch of 58 amino acids of the N-terminal to the trans-activation potential of LEDGF. *J. Mol. Biol.* 2006, *355*, 379-394.

61. Stros, M.; Launholt, D.; Grasser, K. D., The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. *Cell. Mol. Life Sci.* 2007, *64*, 2590-2606.

62. Coleman, J. E.; Oakley, J. L., Physical chemical studies of the structure and function of DNA binding (helix-destabilizing) proteins. *CRC critical reviews in biochemistry* 1980, *7*, 247-89.

63. Murphy, F. V.; Churchill, M. E. A., Nonsequence-specific DNA recognition: a structural perspective. *Structure with Folding & Design* 2000, *8*, R83-R89.

64. Norberg, J., Association of protein-DNA recognition complexes: electrostatic and nonelectrostatic effects. *Arch. Biochem. Biophys.* 2003, *410*, 48-68.

65. Podgornaya, O. I.; Voronin, A. P.; Enukashvily, N. I.; Matveev, I. V.; Lobov, I. B., Structure-specific DNA-binding proteins as the foundation for three-dimensional chromatin organization. In *International Review of Cytology - a Survey of Cell Biology, Vol 224*, Jeon, K. W., Ed. 2003; Vol. 224, pp 227-296.

66. Dhanasekaran, M.; Negi, S.; Sugiura, Y., Designer zinc finger proteins: Tools for creating artificial DNA-binding functional proteins. *Accounts Chem. Res.* 2006, *39*, 45-52.

67. Calhoun, L. N.; Kwon, Y. M., Structure, function and regulation of the DNA-binding protein Dps and its role in acid and oxidative stress resistance in Escherichia coli: a review. *Journal of Applied Microbiology* 2011, *110*, 375-386.

68. Rachez, C.; Sautiere, P.; Formstecher, P.; Lefebvre, P., Identification of amino acids critical for the DNA binding and dimerization properties of the human retinoic acid receptor alpha - Importance of lysine 360, lysine 365, and valine 361. *J. Biol. Chem.* 1996, *271*, 17996-18006.

69. van der Schoot, P.; Bruinsma, R., Electrostatics and the assembly of an RNA virus. *Phys. Rev. E* 2005, *71*, 12.

70. Gao, Y. G.; Su, S. Y.; Robinson, H.; Padmanabhan, S.; Lim, L.; McCrary, B. S.; Edmondson, S. P.; Shriver, J. W.; Wang, A. H. J., The crystal structure of the hyperthermophile chromosomal protein Sso7d bound to DNA. *Nat. Struct. Biol.* 1998, *5*, 782-786.

71. Marmorstein, R.; Fitzgerald, M. X., Modulation of DNA-binding domains for sequence-specific DNA recognition. *Gene* 2003, *304*, 1-12.

72. Liu, Z.; Qiao, J.; Niu, Z.; Wang, Q., Natural supramolecular building blocks: from virus coat proteins to viral nanoparticles. *Chemical Society Reviews* 2012, *41*, 6178-6194.

73. Klug, A., The tobacco mosaic virus particle: structure and assembly. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 1999, *354*, 531-535.

74. Zandi, R.; van der Schoot, P.; Reguera, D.; Kegel, W.; Reiss, H., Classical nucleation theory of virus capsids. *Biophysical Journal* 2006, *90*, 1939-1948.

75. Zlotnick, A., To build a virus capsid - an equilibrium-model of the self-assembly of polyhedral protein complexes. *Journal of Molecular Biology* 1994, *241*, 59-67.

76. Harrison, B. D.; Wilson, T. M. A., Milestones in the research on tobacco mosaic virus. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 1999, *354*, 521-529.

77. Kraft, D. J.; Kegel, W. K.; van der Schoot, P., A Kinetic Zipper Model and the Assembly of Tobacco Mosaic Virus. *Biophysical Journal* 2012, *102*, 2845-2855.

78. Lehn, J. M., Supramolecular chemistry - scope and perspectives molecules, supermolecules, and molecular devices. *Angew. Chem.-Int. Edit. Engl.* 1988, *27*, 89-112.

79. Whitesides, G. M.; Mathias, J. P.; Seto, C. T., Molecular self-assembly and nanochemistry - a chemical strategy for the synthesis of nanostructures. *Science* 1991, *254*, 1312-1319.

80. Hartgerink, J. D.; Beniash, E.; Stupp, S. I., Self-assembly and mineralization of peptide-amphiphile nanofibers. *Science* 2001, *294*, 1684-1688.

81. Lehn, J. M., Toward self-organization and complex matter. *Science* 2002, *295*, 2400-2403.

82. Reinhoudt, D. N.; Crego-Calama, M., Synthesis beyond the molecule. *Science* 2002, *295*, 2403-2407.

83. Whitesides, G. M.; Boncheva, M., Beyond molecules: Self-assembly of mesoscopic and macroscopic components. *Proceedings of the National Academy of Sciences of the United States of America* 2002, *99*, 4769-4774.

84. Whitesides, G. M.; Grzybowski, B., Self-assembly at all scales. *Science* 2002, *295*, 2418-2421.

85. Williamson, J. R., Cooperativity in macromolecular assembly. *Nature Chemical Biology* 2008, *4*, 458-465.

86. Whitty, A., Cooperativity and biological complexity. *Nature Chemical Biology* 2008, *4*, 435-439.

87. Hunter, C. A.; Anderson, H. L., What is Cooperativity? *Angewandte Chemie-International Edition* 2009, *48*, 7488-7499.

88. Rybtchinski, B., Adaptive Supramolecular Nanomaterials Based on Strong Noncovalent Interactions. *Acs Nano* 2011, *5*, 6791-6818.

89. Palma, C.-A.; Cecchini, M.; Samori, P., Predicting self-assembly: from empirism to determinism. *Chemical Society Reviews* 2012, *41*, 3713-3730.

90. Schor, M.; Bolhuis, P. G., The self-assembly mechanism of fibril-forming silk-based block copolymers. *Phys. Chem. Chem. Phys.* 2011, *13*, 10457-10467.

91. Aggeli, A.; Bell, M.; Boden, N.; Keen, J. N.; Knowles, P. F.; McLeish, T. C. B.; Pitkeathly, M.; Radford, S. E., Responsive gels formed by the spontaneous self-assembly of peptides into polymeric beta-sheet tapes. *Nature* 1997, *386*, 259-262.

92. Bellesia, G.; Shea, J.-E., Self-assembly of beta-sheet forming peptides into chiral fibrillar aggregates. *Journal of Chemical Physics* 2007, *126*.

93. Bowerman, C. J.; Liyanage, W.; Federation, A. J.; Nilsson, B. L., Tuning beta-Sheet Peptide Self-Assembly and Hydrogelation Behavior by Modification of Sequence Hydrophobicity and Aromaticity. *Biomacromolecules* 2011, *12*, 2735-2745.

94. Dutt, A.; Spencer, E. C.; Howard, J. A. K.; Pramanik, A., Studies of Amyloid-Like Fibrillogenesis through beta-Sheet-Mediated Self-Assembly of Short Synthetic Peptides. *Chemistry & Biodiversity* 2010, *7*, 363-375.

95. Lamm, M. S.; Rajagopal, K.; Schneider, J. P.; Pochan, D. J., Nanostructure of beta-sheet fibrils constructed by unfolded beta-hairpin peptide self-assembly. *Abstr. Pap. Am. Chem. Soc.* 2004, *228*, U377-U377.

96. Wu, C.; Biancalana, M.; Koide, S.; Shea, J.-E., Binding Modes of Thioflavin-T to the Single-Layer beta-Sheet of the Peptide Self-Assembly Mimics. *Journal of Molecular Biology* 2009, *394*, 627-633.

97. Drexler, K. E., Molecular engineering - An approach to the development of general capabilities for molecular manipulation. *Proceedings of the National Academy of Sciences of the United States of America-Physical Sciences* 1981, *78*, 5275-5278.

98. Lamarre, B.; Ryadnov, M. G., Self-Assembling Viral Mimetics: One Long Journey with Short Steps. *Macromolecular Bioscience* 2011, *11*, 503-513.

99. Grigoryan, G.; Kim, Y. H.; Acharya, R.; Axelrod, K.; Jain, R. M.; Willis, L.; Drndic, M.; Kikkawa, J. M.; DeGrado, W. F., Computational Design of Virus-Like Protein Assemblies on Carbon Nanotube Surfaces. *Science* 2011, *332*, 1071-1076.

100. Miyata, K.; Nishiyama, N.; Kataoka, K., Rational design of smart supramolecular assemblies for gene delivery: chemical challenges in the creation of artificial viruses. *Chemical Society Reviews* 2012, *41*, 2562-2574.

101. Lee, J.-h.; Choi, Y. J.; Lim, Y.-b., Self-assembled filamentous nanostructures for drug/gene delivery applications. *Expert Opinion on Drug Delivery* 2010, *7*, 341-351.

102. Lim, Y.-b.; Lee, E.; Yoon, Y.-R.; Lee, M. S.; Lee, M., Filamentous artificial virus from a self-assembled discrete nanoribbon. *Angewandte Chemie-International Edition* 2008, *47*, 4525-4528.

103. Tu, R. S.; Tirrell, M., Bottom-up design of biomimetic assemblies. *Advanced Drug Delivery Reviews* 2004, *56*, 1537-1563.

104. Mastrobattista, E.; van der Aa, M.; Hennink, W. E.; Crommelin, D. J. A., Artificial viruses: a nanotechnological approach to gene delivery. *Nature Reviews Drug Discovery* 2006, *5*, 115-121.

105. Putnam, D.; Zelikin, A. N.; Izumrudov, V. A.; Langer, R., Polyhistidine-PEG : DNA nanocomposites for gene delivery. *Biomaterials* 2003, *24*, 4425-4433.

106. Wong, S. Y.; Pelet, J. M.; Putnam, D., Polymer systems for gene delivery-past, present, and future. *Progress in Polymer Science* 2007, *32*, 799-837.

107. Martin, M. E.; Rice, K. G., Peptide-guided gene delivery. *Aaps Journal* 2007, *9*, E18-E29.

108. Zeng, J. M.; Wang, S., Enhanced gene delivery to PC12 cells by a cationic polypeptide. *Biomaterials* 2005, *26*, 679-686.

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

110. Valery, C.; Paternostre, M.; Robert, B.; Gulik-Krzywicki, T.; Narayanan, T.; Dedieu, J. C.; Keller, G.; Torres, M. L.; Cherif-Cheikh, R.; Calvo, P.; Artzner, F., Biomimetic organization: Octapeptide self-assembly into nanotubes of viral capsid-like dimension. *Proc. Natl. Acad. Sci. U. S. A.* 2003, *100*, 10258-10262.

111. Aoyama, Y.; Kanamori, T.; Nakai, T.; Sasaki, T.; Horiuchi, S.; Sando, S.; Niidome, T., Artificial viruses and their application to gene delivery. size-controlled gene coating with glycocluster nanoparticles. *Journal of the American Chemical Society* 2003, *125*, 3455-3457.

112. Kakizawa, Y.; Kataoka, K., Block copolymer micelles for delivery of gene and related compounds. *Advanced Drug Delivery Reviews* 2002, *54*, 203-222.

113. Lam, E. T.; Hastie, A.; Lin, C.; Ehrlich, D.; Das, S. K.; Austin, M. D.; Deshpande, P.; Cao, H.; Nagarajan, N.; Xiao, M.; Kwok, P. Y., Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nat. Biotechnol.* 2012, *30*, 771-776.

114. Douglas, S. M.; Dietz, H.; Liedl, T.; Hogberg, B.; Graf, F.; Shih, W. M., Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* 2009, *459*, 414-418.

115. LaBean, T. H.; Li, H. Y., Constructing novel materials with DNA. *Nano Today* 2007, *2*, 26-35.

116. Lee, H.; Lytton-Jean, A. K. R.; Chen, Y.; Love, K. T.; Park, A. I.; Karagiannis, E. D.; Sehgal, A.; Querbes, W.; Zurenko, C. S.; Jayaraman, M.; Peng, C. G.; Charisse, K.; Borodovsky, A.; Manoharan, M.; Donahoe, J. S.; Truelove, J.; Nahrendorf, M.; Langer, R.; Anderson, D. G., Molecularly self-assembled nucleic acid nanoparticles for targeted in vivo siRNA delivery. *Nature Nanotechnology* 2012, *7*, 389-393.

117. Rinker, S.; Ke, Y.; Liu, Y.; Chhabra, R.; Yan, H., Self-assembled DNA nanostructures for distance-dependent multivalent ligand-protein binding. *Nature Nanotechnology* 2008, *3*, 418-422.

118. Burns, J. R.; Stulz, E.; Howorka, S., Self-Assembled DNA Nanopores That Span Lipid Bilayers. *Nano Lett.* 2013, *13*, 2351-2356.

119. Rahman, M.; Norton, M. L., Two-Dimensional Materials as Substrates for the Development of Origami-Based Bionanosensors. *Ieee Transactions on Nanotechnology* 2010, *9*, 539-542.

120. Schlapak, R.; Danzberger, J.; Armitage, D.; Morgan, D.; Ebner, A.; Hinterdorfer, P.; Pollheimer, P.; Gruber, H. J.; Schaeffler, F.; Howorka, S., Nanoscale DNA Tetrahedra Improve Biomolecular Recognition on Patterned Surfaces. *Small* 2012, *8*, 89-97.

121. Wang, R. S.; Nuckolls, C.; Wind, S. J., Assembly of Heterogeneous Functional Nanomaterials on DNA Origami Scaffolds. *Angew. Chem.-Int. Edit.* 2012, *51*, 11325-11327.

122. Barnard, A.; Smith, D. K., Self-Assembled Multivalency: Dynamic Ligand Arrays for High-Affinity Binding. *Angewandte Chemie-International Edition* 2012, *51*, 6572-6581.

123. Ruff, Y.; Moyer, T.; Newcomb, C. J.; Demeler, B.; Stupp, S. I., Precision Templating with DNA of a Virus-like Particle with Peptide Nanostructures. *Journal of the American Chemical Society* 2013, *135*, 6211-6219.

124. Janssen, P. G. A.; Jabbari-Farouji, S.; Surin, M.; Vila, X.; Gielen, J. C.; de Greef, T. F. A.; Vos, M. R. J.; Bomans, P. H. H.; Sommerdijk, N. A. J. M.; Christianen, P. C. M.; Leclere, P.; Lazzaroni, R.; van der Schoot, P.; Meijer, E. W.; Schenning, A. P. H. J., Insights into Templated Supramolecular Polymerization: Binding of Naphthalene Derivatives to ssDNA Templates of Different Lengths. *Journal of the American Chemical Society* 2009, *131*, 1222-1231.

125. Palmer, L. C.; Stupp, S. I., Molecular Self-Assembly into One-Dimensional Nanostructures. *Accounts of Chemical Research* 2008, *41*, 1674-1684.

Chapter 2

Coating of single DNA molecules by diblock copolymers[1]

[1] This chapter is based on Armando Hernandez-Garcia, Marc W. T. Werten, Martien Cohen Stuart, Frits A. de Wolf, and Renko de Vries. Coating of single DNA molecules by genetically engineered protein diblock copolymers, Small, 2012.

2.1 Introduction

Besides being a carrier of genetic information, DNA has also become a much-explored multi-purpose material for biological and non-biological applications. Coating DNA with binder moieties is an effective way to modify physical parameters and modulate its interactions with the environment.[1-5] For example, non-viral gene delivery has been achieved by complexation of DNA with cationic lipids and/or polymers to promote passage across the cell membrane.[6] A controlled DNA coating would be advantageous to regulate doses of transfected DNA and also to direct vector shape. Furthermore, a protein coating on single stranded DNA has been used to facilitate passage through nanopores.[5] Binder moieties could finally be used for modulating DNA physical properties such as bending rigidity, thickness, and interactions with other entities.[2, 7] These properties are crucial for example for nanoscale self-assembly ("DNA origami"),[8] templating of nanowires,[3] transport in nano-channels, and the stretching, separation and sequencing of DNA.[1, 4, 5]

A key issue is control over the architecture of the DNA coating. So far, only a few studies have focused on (1) understanding the requirements for coating of DNA without any intra- or inter-molecular bridging, and on (2) a rational designed of DNA-binding polymers based on these requirements. Typical for current polymeric DNA binders is the formation of random aggregates of uncontrolled size and irregular shape, with many nucleic acid molecules per aggregate.[9-12] This limits the applicability of the current polymers in precisely defined coatings. An understanding of the characteristics of polymer design, that result in a monodisperse polymer-DNA complex and prevent intra- or inter-molecular bridging between DNA segments, could be beneficial for the field of DNA-based nanobiotechnology.

The simplest design for a macromolecule capable of coating DNA would be that of a diblock copolymer containing a cationic domain, connected to another domain that confers colloidal stability.[13, 14] A number of such molecules have been

studied in which the binding domain is a simple, flexible polycation. For example, Kostiainen et al. have conjugated various globular proteins with polycationic blocks and showed that the conjugates are capable of coating double strand DNA (dsDNA).[15] A number of studies have also been performed on synthetic diblock copolymers consisting of a polycationic domain coupled to an uncharged, flexible and inert hydrophilic domain of variable size.[9, 11, 12, 15-23]

DeRouchey et al. have reported controlled coating of single DNA molecules using 20 kDa poly(oxyethylene) (PEG) linked to poly(iminoethylene) (PEI) carrying 50 positive charges.[17] Hartmann et al., used PEG-polyamidoamine (2.7 kDa-20 positive charges) to form a mixed population of single-plasmid dsDNA toroids with different degrees of folding and supercoiling.[18] Osada et al. reported a PEG-poly-L-Lysine (12 kDa-17 positive charges) polymer that induced multiple folding in single plasmid DNA molecules (pDNA).[19] Analysis of these studies suggests that the length of the two blocks -hydrophilic and cationic- and also their combination in an asymmetric diblock copolymer controls the final architecture of the DNA molecule-polymer complex. For the studies with cationic-neutral diblocks performed so far, binding domains were around 20 to hundreds of positive charges, and hydrophilic blocks were between 2.7 and 20 kDa. Most studies report aggregation and/or folding of single DNA molecules by polycationic "bridging". Bridging can occur between several DNA molecules (intermolecular aggregation) and/or between several neighboring parts of the same DNA molecule (intramolecular aggregation). Such bridging creates a diverse population of shapes and sizes.[9, 19, 20]

Bridging does not distinguish between parts of the same or different molecules. Therefore, if intramolecular bridging occurs, it is very likely that intermolecular aggregation will also occur in time, or at high DNA concentrations.[23,24] This can be prevented if the binding domain is short and a protective cover such as a large hydrophilic block is present around the DNA. DeRouchey et al. reported neither intermolecular aggregation nor intramolecular folding, despite the fact that they used very long binding blocks (50 positive charges, roughly the size of the neutral

block). This may be related to the fact that they have used rather short DNA fragments (75-1018 base pairs) and low concentrations (<0.33 mg/L).[17, 24, 25]

For artificial DNA binders, a relatively long colloidal stability block and a relatively short DNA-binding block could be successful for long-term prevention of random aggregates of uncontrolled size and shape.[9] Probably, a small binding block coupled to a very large steric block (*i.e.* a highly asymmetric design) will promote coating of single DNA molecules and prevent intermolecular aggregation and/or intramolecular folding. This will significantly reduce the presence of structures with different sizes and shapes and thus make the preparation more homogeneous, with evident advantages for technological applications.[26] Structural asymmetry/symmetry is a feature that is also exploited by nature to build functional nanoparticles.[27, 28]

An attractive technology to produce custom-designed polymers is the use of genetic engineering.[16, 23, 29-31] This offers the possibility to produce protein polymers with exact control of the length and sequence, monomer by monomer, of the binding and protective domains. Moreover, it creates the additional possibility of inserting biofunctional domains such as for binding to living cells. In addition, production of such polymers as (monodisperse, exactly defined) heterologous proteins in suitable recombinant microbial production hosts would create possibilities for production at relatively low cost and large scale without any loss of product quality. This contrasts with traditional chemosynthetic production, which cannot combine monodispersity and a defined monomer sequence with low cost and large scale of production. Previously, we have demonstrated the biosynthetic production of various large protein block copolymers using recombinant *Pichia pastoris* yeast strains.[32-34] The polypeptides were secreted into the medium at g/L levels and could be recovered using a simple downstream processing suitable for scale-up scaffold.

Here, we use recombinant *P. pastoris* for the biosynthetic production of two neutral-cationic protein diblock copolymers that coat long single molecules of plasmid DNA, both linear and supercoiled without any signs of intermolecular

aggregation even at high DNA concentrations. The present protein polymers could be used for future step-wise addition of blocks for further detailed control of structural and functional properties. The development of our capabilities to control and modulate the properties of single DNA molecules by coating them with designer protein polymers could put forward a number of emergent nanotechnologies such as single DNA sequencing and single nucleic acid molecule delivery.

2.2 Materials and Methods

2.2.1. Materials. Supercoiled plasmid DNA pMTL23-C4K12 (3.7 kb), used for binding studies, was prepared using a Megaprep kit (Qiagen, Germany) from a recombinant *E. coli* culture grown overnight. The purified DNA was dissolved in sterile MQ-water at a final concentration of 232 mg/L (Nanodrop ND-1000 Spectrophotometer; Thermo Scientific, Waltham, MA). Several portions were diluted to a final concentration of 116 mg/L in 10 mM acetate buffer, pH 5, and microfiltrated (0.2μm) for light scattering analysis. Restriction enzymes used for molecular cloning of the cationic protein polymers were purchased from New England Biolabs (Ipswich, MA) or Fermentas (Lithuania). Acetate buffer was prepared from acetic acid and sodium acetate of analytical quality purchased from Sigma (St. Louis, MO).

2.2.2. Construction of Expression Vectors and Strains. The previously described vector pMTL23-P4 contains a gene encoding an artificial highly hydrophilic random coiled protein "P4".[33] In the present article this collagen-like sequence is denoted as the colloidal stability block "C_4". The vector was digested with *Van*91I/*Eco*RI (3' to the C4 gene). The binding blocks were prepared by annealing of complementary codon-optimized oligonucleotides (Eurogentec, Seraing, Belgium; see Appendix Table 2.2 for DNA sequences). The resulting double-stranded adaptors encode the amino acid sequences as shown in Figure 2.1D, and provide *Van*91I/*Eco*RI sticky ends. The adaptors were separately ligated into the previously digested vector, resulting in plasmids pMTL23-C4-K12 and pMTL23-

C4-HK6. The fragments encoding the C_4-K_{12} and C_4-$(HK)_6$ diblock copolymers were released through digestion of pMTL23-C4-K12 and pMTL23-C4-HK6 with *Xho*I/*Eco*RI and ligated into the correspondingly digested *P. pastoris* expression vector pPIC9 (Invitrogen, Breda, The Netherlands). The resulting plasmids pPIC9-C4-K12 and pPIC9-C4-HK6 were linearized with *SalI* and electroporated into *P. pastoris* strain GS115 (Invitrogen). The plasmid integrates into the genome through homologous recombination at the *his4* locus providing normal growth on methanol. The presence of the genes was verified by PCR.

2.2.3. Protein Polymer Biosynthesis. Fed-batch fermentations using minimal basal salts medium were performed in 2.5-L Bioflo 3000 fermentors (New Brunswick Scientific, Edison, NJ), similarly to Zhang et al.[35] The methanol fed-batch phase for protein production lasted two to three days. A homemade semiconductor gas sensor-controller was used to monitor the methanol level in the off-gas and to maintain a constant level of 0.2% (w/v) methanol in the broth. The pH was maintained at 3.0 throughout the fermentation by addition of ammonium hydroxide. At the end of the fermentation, the cells were separated from the broth by centrifugation for 15 min at 10000 × g (room temperature or 4 °C) in an SLA-3000 rotor (Thermo Scientific, Waltham, MA), and the supernatant was microfiltered (Pall Corporation, Port Washington, NY) and immediately stored at 4 °C for subsequent purification.

2.2.4. Protein Polymer Purification. All centrifugation was done for 30 min at 20000 × g at 4 °C, interchangeably in a Sorvall SLA-1500 or SLA-3000 rotor (Thermo Scientific, Waltham, MA). First, medium salts were removed from the cell-free broth by adjustment of the pH to 8.0 with NaOH, followed by centrifugation. Subsequently, the diblock copolymers were selectively precipitated from the solution, by adding ammonium sulphate to a saturation of 45%, incubating overnight at 4 °C, and subsequent centrifugation. The pellet was resuspended in an equal volume (relative to the cell-free broth) of Milli-Q water and precipitation was repeated once at 4°C, using incubation time of 1-2 hrs. The pellet was resuspended in 0.2 volumes (relative to the cell-free broth) of Milli-Q

water and sodium chloride and acetone were added to a final concentration of 50 mM and 40% (v/v), respectively. After centrifugation the acetone concentration of the supernatant was raised to 80% (v/v), and the solution was centrifuged in order to precipitate the pure diblock copolymer. The pellet was dried overnight, resuspended in Milli-Q water, extensively desalted by dialysis against Milli-Q water and lyophilized.

2.2.5. SDS-PAGE and Mass Spectrometry. SDS-PAGE was carried out using the NuPAGE Novex system (Invitrogen, Carlsbad, CA) with 10% Bis-Tris gels, MES SDS as running buffer and SeeBlue Plus2 prestained molecular mass markers. Gels were stained with Coomassie SimplyBlue SafeStain (Invitrogen). Matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) was carried out in an Ultraflex mass spectrometer (Bruker, Billerica, MA). Proteins samples were prepared by the dried droplet method. The matrix was made up of 5 mg/mL 2.5-dihydroxyacetophenone, 1.5 mg/mL diammonium hydrogen citrate, 25% (v/v) ethanol, and 1% (v/v) trifluoroacetic acid on a 600 μm AnchorChip target (Bruker). An external mass calibration was done based on Protein Calibration Standard II (Bruker).

2.2.6. Agarose Gel Mobility Retardation Assay. Aliquots 46 ng of pDNA (3.7 kb) dissolved in 10 mM acetate buffer pH 5 were mixed with different volumes of a 1g/L C_4-B_{12} solution dissolved in the same buffer to a final volume of 15 µL. After 30-60 min at room temperature, the mixtures were electrophoresed in a 1% agarose gel for 60 min at 90V using 1x TAE buffer (pH 8). Bands were visualized using ethidium bromide.

2.2.7. Dynamic Light Scattering. Light scattering measurements were performed with a Zetasizer NanoZS apparatus (Malvern Instruments, UK) equipped with a 4mW He-Ne ion laser at a wavelength of 633nm. DNA-protein complexes were formed at room temperature by mixing 5.6 µL of DNA (116 mg/L) dissolved in 10 mM acetate buffer pH 5 together with a certain volume of desalted and purified C_4-B_{12} (1-10 g/L) dissolved in the same buffer, as necessary to obtain the desired N/P ratio. Buffer was added to the DNA and diblock copolymer premixed solutions to a

final volume of 15 μL and mixed. DNA, protein and buffer solutions were previously filtrated using 0.2μm filters for all the further experiments. The solutions were left to equilibrate for 1-2 hrs at room temperature before DLS measurements. The intensity of light scattered by the DNA-protein complexes was determined from an average of five autocorrelation measurements carried out in the course of 60 seconds at 25ºC using a scattering angle of 12.8 and/or 173 degrees. Effective hydrodynamic radius was automatically obtained from the apparent diffusion coefficient calculated by the apparatus.

2.2.8. Molecular Weight Estimation. Determination of the molecular weight using static light scattering requires the absolute intensity of light scattered at an angle Θ, called the Rayleigh ratio R_θ:

$$R_\theta = \frac{I_{sample} - I_{solvent}}{I_{toluene}} \frac{n^2_{solvent}}{n^2_{toluene}} R_{toluene} \qquad (2.1)$$

Where I_{sample}, $I_{solvent}$, and $I_{toluene}$ are, respective, the relative scattered intensities of the sample, the solvent and the toluene standard, which has an absolute scattering intensity of $R_{toluene}$ = 0.0013522 m^{-1}. Furthermore, $n_{solvent}$ and $n_{toluene}$ are the solvent and toluene refractive indices (1.333 & 1.49366, respectively). The Rayleigh constant is:

$$K_R = \frac{4\pi^2 n^2_{solvent}(dn/dC)^2}{\lambda^4 N_A} \qquad (2.2)$$

where N_A is Avogadro's number, dn/dC is the refractive index increment. For low scattering angles and low concentrations, the absolute intensity is independent of scattering angle and linear in both the weight concentration C and weight-averaged molar mass M_w:

$$\frac{K_R C}{R_\theta} = \frac{1}{M_w} \qquad (2.3)$$

2.2.9. Atomic Force Microscopy. DNA-diblock copolymer complex solutions (prepared the same way as for DLS experiments) were diluted 10 or 20 times with filtrated Milli-Q water. Immediately 3-5 μL of the resulting solution was added onto a clean hydrophilic 1x1 cm silicon or freshly cleaved mica wafer and left for 1-2 min. Then it was washed with 500 μL of filtrated Milli-Q water to remove salts and non-absorbed particles, followed by soaking up of excess water using a tissue and slow drying under a nitrogen stream. Samples were analyzed using a Digital Instruments NanoScope V equipped with a silicon nitride probe (Veeco, NY, USA) with a spring constant of 0.32 N/m in *ScanAsyst*™ imaging mode. Images were recorded between 0.488-0.965 Hz and 384-1024 samples/line. Image processing was done with NanoScope Analysis 1.20 software. Contour length and long axis length measurements were performed with *ImageJ* software.

2.3 Results and Discussion

2.3.1. Design of Protein Diblock Copolymers. In order to coat single DNA molecules and avoiding inter and/or intramolecular bridging of DNA we propose a diblock copolymer design that consist of a colloidal block and a cationic binding DNA block (Figure 2.1A and 2.1B). The colloidal stability block has to be bulky and large enough to confer strong stabilization without impeding binding and also unable to promote DNA condensation. The cationic DNA binding block needs to have a short length to avoid intra and intermolecular bridging of the DNA molecules but long enough to bind DNA strongly. For the colloidal stabilizing block we use an highly soluble and biocompatible collagen-like sequence, based on our prior work (previously referred to as P4 block; denoted here as C_4-block), that consists of approx. 400 amino acids (~36 kDa) and exhibits random coil structure and not any self assembly properties in solution,[32-34] Figure 2.1B and 2.1C). The choice of the length of C_4, is based on our previous work on self-stacking silk-like polymers stabilized by C_4 blocks.[32] For the binding block, we chose a short stretch of 12 cationic amino acids (denoted here as binding block B_{12}) (Figure 2.1D).

The length of the binding block, twelve cationic monomers still gives good electrostatic DNA binding.[10] Two different binding blocks were produced: B^{K12} (12 lysines) and $B^{(HK)6}$ (six histidine-lysine dyads), which could show a different (pH dependent) binding profile. The long C_4 (~400 amino acids) and the short B_{12} blocks (12 positively charged amino acids) should make sure that the diblock gives a thermodynamically stable coating of single DNA molecules, without any bridging, in opposite to previous work that have found DNA condensation to some extent (intramolecular bridging).[18, 29, 36] For this design, we expect thermodynamically stable single-DNA complexes, and not just complexes that are kinetically stable,[13] that is, at low concentrations or for a limited time.[37]

The binding blocks were placed at the C-terminus of the protein polymer to ensure that any truncated proteins hypothetically resulting from premature termination of transcription or translation do not bind the DNA. A Gly residue was added to the C-terminus of the binding block to prevent exoproteolysis of the basic amino acids by the *P. pastoris* homolog of Kex1 protease.[38]

2.3.2. Biosynthesis of Protein Polymers. The two protein diblock copolymers were successfully produced using recombinant strains of yeast *P. pastoris* carrying the artificial genes coding for the desired amino acid sequences. After differential salt precipitation (see materials and methods) the intact C_4-B^{K12} and C_4-$B^{(HK)6}$ were recovered with a yield of up to ~1g/L (weight of lyophilized salt-free protein to volume of cell-free medium). The precipitation steps were chosen in such a way that autologous *P. pastoris* proteins and exopolysaccharides present in the extracellular medium were not precipitated, but only the heterologous protein polymer products were precipitated.

2.3.3. Molecular Characterization. Purity and molecular weight of C_4-B^{K12} and C_4- $B^{(HK)6}$ after purification were characterized using SDS-PAGE and MALDI-TOF (Figure 2.2). SDS-PAGE showed well-defined single bands for both purified diblocks that corresponded to the dominant bands in the lane of the harvested medium. The molecular weight of the purified polypeptides as apparent from SDS-PAGE (~98 kDa) did not correspond to the expected value (38 kDa).

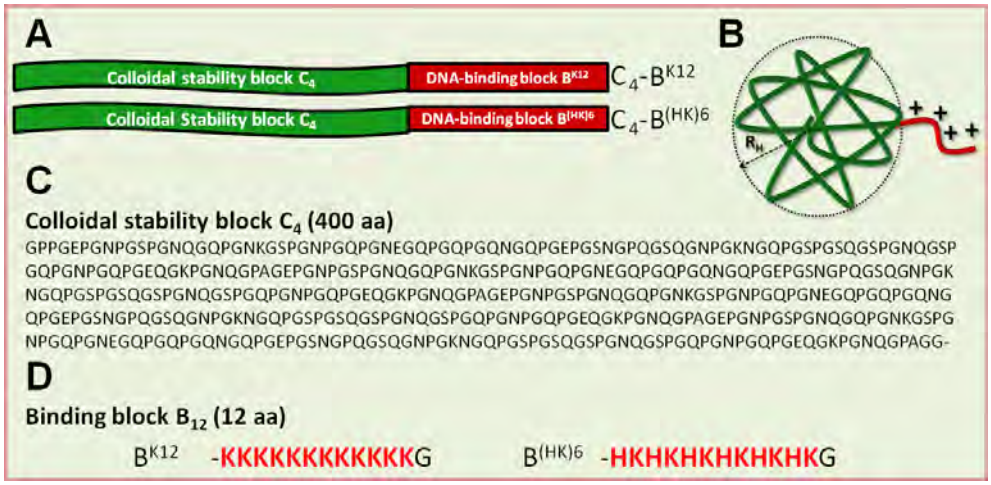


Figure 2.1. Protein diblock copolymer design and composition. (A) Gross structure (arrangement of blocks) in the presently-used DNA-binding polymers. Green section is the hydrophilic polyamino acid coil for colloidal stability. Red section is the cationic block for DNA binding. (B) Representation of the protein diblock copolymer as a random coil, R_H: Hydrodynamic radius. (C) Amino acid sequence of the colloidal stability block and (D) Amino acid sequence of the DNA-binding blocks.

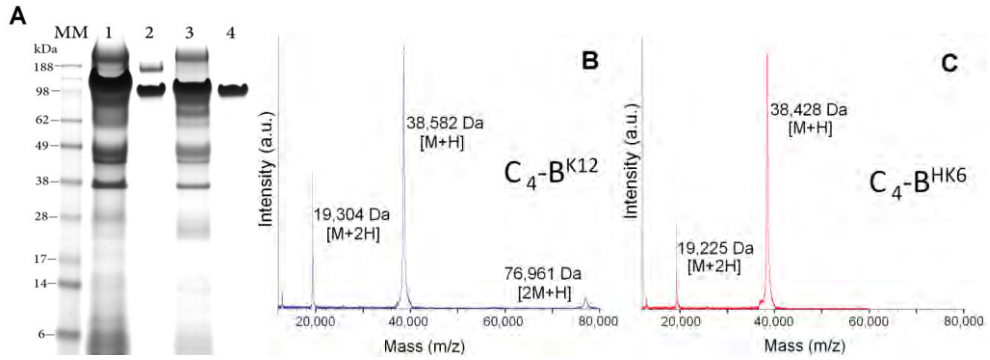


Figure 2.2 molecular characterizations of protein diblock copolymers. (A) SDS-PAGE: MM: molecular weight marker, lane 1: C4-BK12 harvested medium at the end of the fermentation (10 µL loaded), 2: Purified C4-BK12 (10 µg loaded), 3: C4-B(HK)6 harvested medium at the end of the fermentation, 4: Purified C4-B(HK)6 sample (10 µg loaded). (B) MALDI-TOF of purified C4-BK12.

A probable cause for the reduced mobility is the small size of the amino acids in the C_4 block and its very low content of hydrophobic amino acids, which leads to poor binding of SDS.[33] In addition, the highly cationic character of the B block

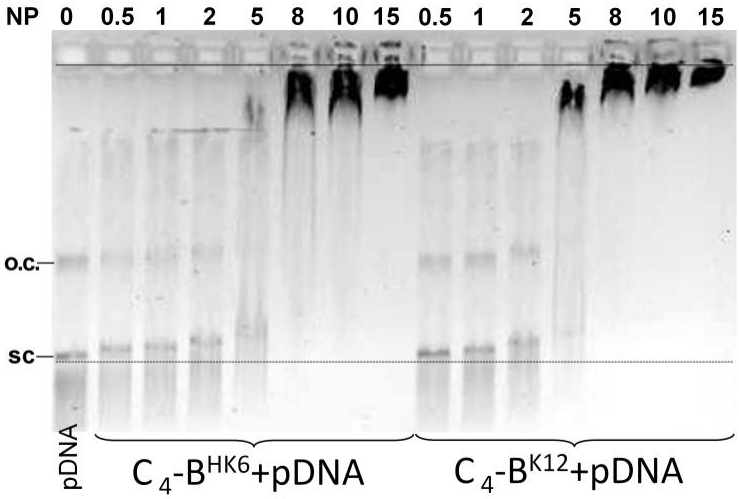


Figure 2.3. Electrophoretic mobility of pDNA-Protein diblock copolymer complexes on a 1% agarose gel. The N/P ratio is indicated at the top. The solid line at the top indicates the starting position of samples before electrophoretic migration and the dashed line at the bottom indicates the final migration position of plamid DNA (pDNA) alone (N/P = 0). "sp" indicates band position of supercoiled pDNA and "o.c." open circular pDNA. The mobility of protein-DNA complexes was drastically reduced at higher N/P.

might have been of influence. The minor upper band visible in purified C_4-B^{K12} is probably a remaining protein impurity.

The molecular weights of the diblock protein polymers produced by *P. pastoris* were analyzed by MALDI-TOF. For C_4-B^{K12} peaks at 19,304, 38,582 and 76,961 Da were found, corresponding to M+2H, M+H and 2M+H, respectively. For C_4-$B^{(HK)6}$ 19,225 and 38,428 Da corresponding to M+2H and M+H, respectively. These mass values are in good agreement with the expected theoretical mass values ~38,407 and 38,461 Da for C_4-B^{K12} and C_4-$B^{(HK)6}$, respectively.

2.3.4. Electrophoretic mobility shift assay of DNA-protein complexes. Agarose gel mobility retardation was used as an initial assay for the interaction between the two protein diblock copolymers and plasmid DNA. Binding of the protein polymer to the DNA changes the charge, size and shape (and thus, the electrophoretic mobility) of the naked DNA into that of the DNA-polymer complex. Figure 2.3

shows that the electrophoretic mobility of DNA is reduced when protein block copolymers are added. At higher N/P (ratio of the amino groups of lysine side chains in the binding block to the phosphates from DNA backbone) the effect increases and saturates at about N/P =5. Although the net charge of the resulting complexes was almost neutral at high N/P it never became positive. Each colloidal stabilization block carries a net charge of about -5 under the conditions of the agarose gel electrophoresis (pH 8) and thus, complete neutralization of DNA charges with the C_4-B_{12} polymers will leave a net negatively charged complex. On the other hand, it is possible that the coating cannot become dense enough to neutralize all the charge even at high N/P. This behavior is similar to a reported recombinant large protein carrying a short binding block, where total retardation of the protein-DNA complexes was not observed either.[23] Other recombinant diblock copolymers using a short poly-lysine block and a large elastine-like polypeptide were able to condense DNA even at N/P < 1, because of the ability of elastine-like sequences to collapse.[29] In contrast, neutral-cationic diblock polymers with a smaller colloidal stabilization block and larger binding block were observed to completely neutralize the DNA charge, or even reverse the mobility probably by overcharging.[17, 18, 22] As a control, we have verified that there is no mobility shift when DNA is mixed with C_4 block alone (not shown).

2.3.5. Light Scattering Studies. Light Scattering was used to further study the interaction between DNA and the protein block copolymers, since it is a convenient technique that does not require any labeling in order to obtain the hydrodynamic size of the complexes in solution and an estimate of their molecular weights. Figure 2.4A shows the absolute intensity of scattered light or Rayleigh ratio R_{θ}, (see materials and methods) of the complexes as a function of the N/P ratio. For both copolymers, the scattered intensity increased rapidly with increasing N/P ratio in the range N/P < 1, and then saturated at N/P >> 1. The scattered intensity of the fully covered DNA was about a factor 100 larger than that of the naked DNA. For a rough estimate of the degree of coverage, it may be assumed that the scattering at N/P>>1 is dominated by the complexes and that the contribution

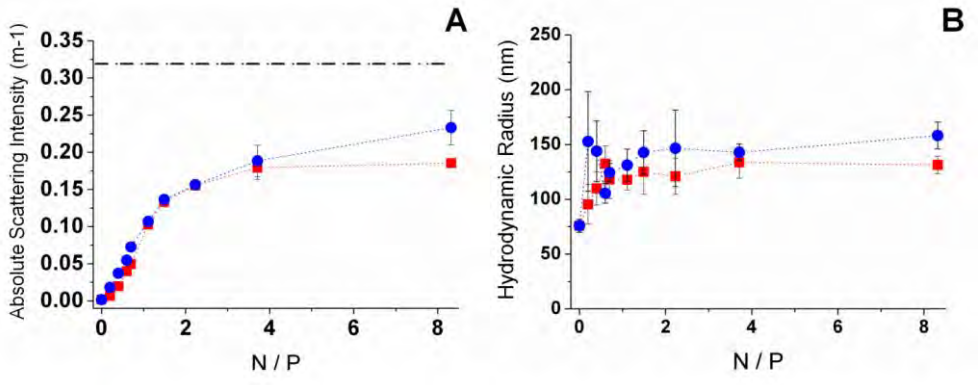


Figure 2.4. Light scattering intensity and hydrodynamic radius for pDNA-protein diblock copolymer complexes formed at different N/P at θ = 12.8°. A: Absolute light scattering intensity; broken line means the predicted scattering of fully protein coated pDNA. B: Hydrodynamic radius of the complexes. Closed red square: C_4-B^{K12} complexes, blue circle: C_4-$B^{(HK)6}$ complexes.

of excess free polymer to the scattering can be neglected. Assuming full charge neutralization, the predicted molar mass M of a single coated DNA molecule is

$$M = (1+\alpha)M_{DNA}$$
$$\alpha = \frac{1}{n_p}\frac{M_p}{M_0} \qquad (2.4)$$

where M_{DNA} and M_p are respectively, the DNA and polymer molar mass, n_p is the number of base pairs covered by a single bound C_4-B_{12} polymer, and M_0 ~ 600 Da is the average molar mass of a single DNA base pair (bp) without sodium counterions. Assuming n_p = 6 bp (one C_4-B_{12} polymer per 12 DNA phosphate charges or 6 base pairs), gives α = 10.7. In the limit of zero scattering angle and zero concentration, the predicted absolute scattering intensity R_θ for the complexes is

$$R_\theta = K_R C M \qquad (2.5)$$

where $C = (1+\alpha)C_{DNA}$ is the weight concentration of complexes, C_{DNA} is the weight concentration of DNA and M is as defined above. Literature values for the

refractive index increments for dsDNA and the collagen proteins are within the same range of dn/dc = 0.16 to 0.18 ml/g (see materials and methods). Using a single average value of dn/dc = 0.17 for C_4-B_{12} gives a value of the Rayleigh constant of K_R = $2.105 \cdot 10^{-5}$ mol m^2 kg^{-2} (see materials and methods). Roughly speaking, both C and M increase by a factor of about 10 when the DNA is coated by the protein polymer, such that the light scattering increases by a factor of about 100, as observed. The estimated absolute scattering intensity of $R_\theta = 0.319$ m^{-1} for charge neutral complexes is indicated as a dashed line in figure 2.4A, and lies somewhat above the saturating values for both types of complexes. In line with the results of the electrophoretic mobility shift assay (Figure 2.3), this indicates that at saturation, for both complexes, an important part, but not all of the DNA phosphate charges have been neutralized by complexation with binding domains.

Figure 2.4B shows the effective hydrodynamic radii (R_h) determined using DLS, as a function of N/P. The R_h values correspond to a single diffusing species at each N/P. For C_4-$B^{(HK)6}$, the size rapidly increases from 75 nm in the absence of polymer, to 150 nm at N/P = 1.5. For C_4-B^{K12} + DNA, the final size is somewhat smaller, Rh = 135 nm at N/P = 1.5. These values are close to the reported for similar bulky cationic polypeptides.[23] For neutral-cationic diblock copolymers with longer binding blocks, smaller solution sizes (<100 nm) have been reported, most likely because of DNA condensation.[22] The later was also found for elastine-like diblock copolymers since their well known ability to collapse.[29] Finally, for the individual diblocks, in the absence of DNA, the scattered intensity was very low. The hydrodynamic radius was 5.4 nm for C_4-B^{K12} and 6.1 nm for C_4-$B^{(HK)6}$. These values are consistent with the reported radius of gyration for C_4 of ~7 nm.[34]

It is important to note that the reported effective hydrodynamic radius R_h of the complexes is calculated from a measured diffusion D constant assuming a spherical particle size,

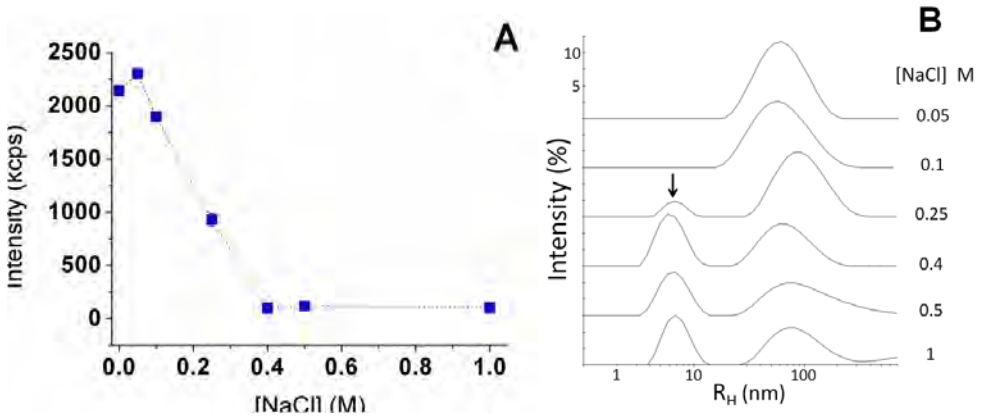


Figure 2.5. Evolution of DNA-protein C_4-B^{K12}-pDNA nanoplexes disassembly at different salt concentrations by dynamic light scattering (θ = 173°, N/P = 1.1). A: Intensity of scattered light of nanoplexes at different salt concentrations. The large drop of intensity at [NaCl] = 0.4 M indicates virtual total disassembly of nanoplexes. B: Light scattered intensity by size distribution of nanoplexes at different salt concentrations. Small ~6 nm (indicated by an arrow) and large ~75nm peaks correspond to free diblock copolymers and naked pDNA in solution, respectively.

$$D = \frac{k_B T}{6\pi\eta R_h} \qquad (2.6)$$

where k_BT is the thermal energy, and η is the solvent viscosity. A 3.7kb supercoiled plasmid, both with and without adhering diblock polymers, presumably is closer to a rod-like than to a spherical particle. Assuming the diffusion of both naked and coated supercoiled plasmid can be approximated by that of a cylindrical particle with length L and diameter d, the predicted diffusion constant is:

$$D = \frac{Ak_B T}{3\pi\eta L}$$
$$A = \ln(L/d) + 0.312 + 0.565/(L/d) - 0.1/(L/d)^2 \qquad (2.7)$$

This approximation was used by deRouchey et al. for short (rod-like) linear DNA (<1kb) coated with cationic-neutral diblock polymers to extract an estimated thickness d of the coated DNA, using the known length L.[17] For a very rough

estimate of the effective thickness of the (coated) plasmid "cylinder", we use L = 629 nm, which is half the contour length of 3.7kb DNA (using a contour length per basepair of 0.34 nm). Then, for uncoated DNA we find d = 12nm, which is not unreasonable for the diameter of a plectonemic supercoil. For the fully coated supercoils we find d = 87 nm (assuming R_h = 135nm for C_4-B^{K12}) and d = 112nm (assuming R_h = 150 nm for C_4-$B^{(HK)6}$). We wish to emphasize that these are only rough estimates of the hydrodynamic thickness of the coated supercoils. In particular, the effect of the flexibility of the coated supercoils should be included for a more precise determination.

2.3.6. Charge Screening by salt addition. To confirm that complexes are formed due to electrostatic interactions, N/P =1.1 complexes were exposed to increasing concentration of salt (NaCl) and their disassembly was followed using DLS (Figure 2.5A and 2.5B). These experiments were carried out at a larger scattering angle of 173°, corresponding to a larger value of the wavevector q (see materials and methods) to allow for a more sensitive detection of the diffusion of the much smaller free polymers. Figure 2.5A shows that the scattering intensity decreases rapidly for salt concentrations between 0.1M and 0.25 M. Figure 2.5B shows the size distributions as determined using DLS. The contribution due to diffusion of free diblocks (R_H = 6 nm) starts to become visible at 0.25 M. The peak at around 100 nm most probably is due to the naked plasmid DNA, implying that at this salt concentration, most diblocks are no longer bound to the DNA. Above 0.4M the scattering intensity is constant and no polymers are bound anymore. These experiments demonstrate that the complexation is clearly electrostatic in nature. This also means that further tuning of the binding strength may be achieved via changes of pH that affect the charge of the binding blocks. The design of the binding block could, in that case, be tailored to specific pH-dependent behavior.

2.3.7. Atomic Force Microscopy. AFM studies in air were carried out to determine the morphology of the complexes and to confirm that complexes consist of individual DNA molecules coated by diblock copolymer. We have found that the collagen-like C_4 blocks rapidly adsorb on the mica and silica surfaces that we have

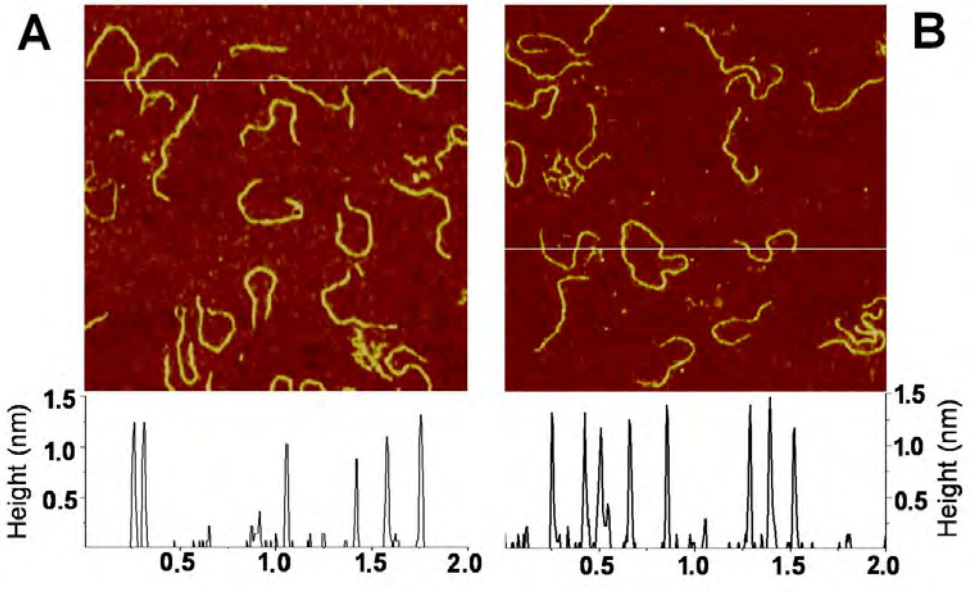


Figure 2.6. Atomic Force Microscopy images in air and height profiles of typical rod-like shape nanoplexes formed between protein diblock copolymers and pDNA 3.7 kb on silica. A: C_4-B^{K12}-pDNA (N/P = 2.2). B: C_4-$B^{(HK)6}$-pDNA (N/P = 1.5). Images sizes are 2x2 µm.

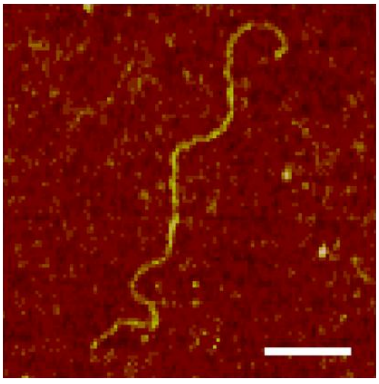


Figure 2.7. Atomic Force Microscopy image in air of linear dsDNA coated with C_4-B^{K12} protein diblock copolymers at N/P = 1 deposited on mica. Contour length L = 1258 nm, height is 0.45 nm. Scale bar represents 200 nm.

used as a substrate for AFM. Whereas naked DNA does not adhere on bare mica or silica, all of the complexes did, without requiring divalent cations or poly-Lysine.

This is advantageous because it excludes any possible influence of these additives on the morphology of the adhering DNA complexes.[39]

Figure 2.6 shows typical images of C_4-B^{K12} and C_4-$B^{(HK)6}$-DNA complexes. In both cases, the complexes are well-dispersed. They appear as semiflexible single rods of regular shape. The average contour length of the nanoplexes is 649.4 ± 15 nm for C_4-B^{K12} and 624.7 ± 29.3 for C_4-$B^{(HK)6}$ (N = 20). The height for both complexes is uniform with an average around 1 nm (1.07 ± 0.1 and 0.99 ± 0.2 (N = 20) for C_4-B^{K12} and C_4-$B^{(HK)6}$, respectively). This indicates that the C_4 block lies flat on the surface. The contour lengths for the complexes are very close to half the contour length of the plasmid DNA, as expected for plectonemic supercoils. The length of the pDNA is 3.7kb, which corresponds to a contour length of L_{DNA} = 1258 nm, assuming a length per bp of 0.34 nm.[40] This indicates that there is no intermolecular DNA bridging.

For the plectonemic supercoils we cannot determine from AFM whether the two strands are bridged or not. To check for bridging, we have therefore also imaged complexes with linearized pDNA. A representative image of a C_4-B^{K12}-linearized pDNA complex at N/P = 1 is shown in Figure 2.7. In this case the contour length of the complex indeed matches the expected contour length of the naked DNA, indicating that individual DNA molecules get coated, and that there is no intramolecular folding due to bridging of the binding blocks.

A series of images of individual complexes at increasing values of N/P (for both polymers) is shown in figure 2.8 and their average heights in Table 2.1. For both polymers, the complexes become thicker, stiffer and more homogeneously coated at higher N/P. At low N/P, the complexes appear to be thin and less rigid. The C_4-$B^{(HK)6}$-DNA complexes appear to be thicker at lower N/P as compared to C_4-K_{12} complexes: N/P = 0.4 complexes of C_4-$B^{(HK)6}$ are similar to N/P = 1 complexes of C_4-B^{K12}. Also, both types of complexes continue to increase in thickness for N/P > 1. The latter finding is in agreement with the light scattering and electrophoresis results. Whereas at low N/P, occasionally we can still distinguish the two duplexes of the superhelix, at higher N/P only a single thick fiber is observed. In the

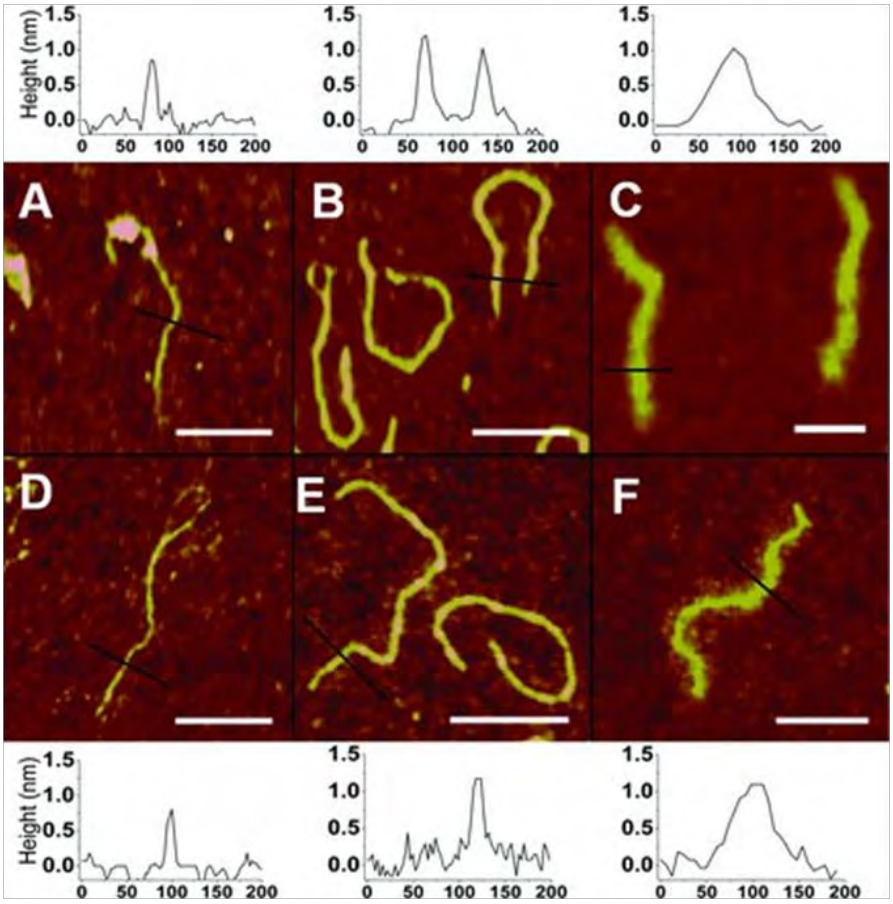


Figure 2.8. AFM images in air of complexes between single pDNA molecules and protein diblock copolymers at different N/P ratios on silica, and height profiles. Top row images: C_4-B^{K12}–pDNA nanoplexes; A) N/P = 0.95, B) N/P = 2.2,C) N/P = 8.3;the corresponding height profiles are at the top. Bottom row images: C_4-$B^{(HK)6}$–pDNA nanoplexes; D) N/P = 0.4, E) N/P = 1.5, F) N/P = 8.5;the corresponding height profiles are at the bottom. Scale bars and *x*-axes of height profiles correspond to 200 nm. Black bars are segments of height profiles.

polymer/DNA complex, the two duplexes of the superhelix may be either bridged by the DNA binding blocks or not, but the resolution of the AFM images is not sufficient to distinguish between these cases.

In conclusion, the AFM measurements are completely consistent with the light scattering measurements and indicate that the complexes consist of individual

Table 2.1. Average height profiles of protein-coated pDNA.

C_4-B^{K12}		C_4-$B^{(HK)6}$	
N/P	**Height [nm]**	**N/P**	**Height [nm]**
0.95	0.82 ± 0.18	0.4	0.87 ± 0.19
2.2	1.03 ± 0.14	1.5	1.20 ± 0.15
8.3	1.06 ± 0.10	8.5	1.10 ± 0.12

DNA molecules, coated with diblock copolymers. Furthermore, the observed heights are very uniform, which would not have been the case if there had been a spread in the number of dsDNA chains per complex.

2.3.8. Numerical Estimates. DNA has two charges every 0.34 nm (every bp), such that the bare linear charge density υ of dsDNA (elementary charges *e* per unit length) is $\upsilon \approx 5.9$ nm^{-1}. The distance between two successive amino acids in a fully stretched polypeptide (protein) chain is about 0.4 nm.[41] This agrees well with the above-mentioned distance between two successive phosphate groups in the DNA chain. If we assume that a single B_{12} block neutralizes 12 DNA phosphate charges, the B_{12} block covers an estimated length of dsDNA of $h_{\min} = 2.0$ nm. DNA-bound diblock copolymers start to come into contact with each other when they are bound at a spacing *h* along the DNA contour smaller than the coil diameter of the C_4 block. This occurs at a critical spacing h^* and critical degree of neutralization f^*

$$h^* \approx 2R \approx 12\text{nm}$$
$$f^* \approx h_{\min} / h^* \approx 0.17 \qquad (2.8)$$

where $R \approx 6$ nm is the coil radius of the C_4 block, as determined using DLS. For N/P > f^*, electrostatic binding of the diblock copolymers is opposed by steric repulsion of neighboring C_4 blocks. As is clear from the experimental data, binding continues to increase well beyond N/P = f^*, and this will lead to stretching of the C_4 blocks (Figure 2.9). When the free energy penalty due to side-chain stretching exceeds the binding free energy of the B_{12} block, no more additional polymers are

bound, even if not all of the DNA charges have been neutralized. Both the electrophoresis and the static light scattering suggest that in the case of our diblock copolymers this should happen at N/P > 8.

Side-chain stretching has been discussed extensively in the literature in relation to so-called bottle-brush polymers, consisting of a main chain covered with regularly spaced side chains (see the recent review of Hsu et. al.).[42] Assuming good solvent conditions, the scaling estimate for the degree of stretching of the grafted side chains (each consisting of N segments of length l) is:

$$\frac{R_s}{R} \propto \Gamma^{1/4}$$
$$\Gamma = N\frac{l}{h} \qquad (2.9)$$

where R_s is the mean-square end-to-end distance of the stretched chain and R is the mean-square end-to-end distance of the unstretched chain. Note that in the absence of a numerical prefactor for this scaling estimate, it can only be used to get order-of-magnitude estimates. Assuming l = 0.4 nm and N=400, we find $\Gamma^{1/4} \approx 3.0$, for the fully charge-neutral complex with h=h_{min} =2nm. This implies that there is only moderate stretching.

Based on these estimates, one ideally would like to calculate the expected thickness of a coated plectonemic supercoil (to be compared with the rough estimate of $D \approx 100$ nm obtained by DLS). This is not straightforward, however, since the two coated duplexes of the plectonemic supercoil may be expected to interact quite strongly (with some interpenetration of the polymer brushes), or even show some bridging at high N/P. Clearly, more accurate determinations of the DNA bottle-brush diameters and grafting densities are needed to resolve this issue.[17, 43, 44]

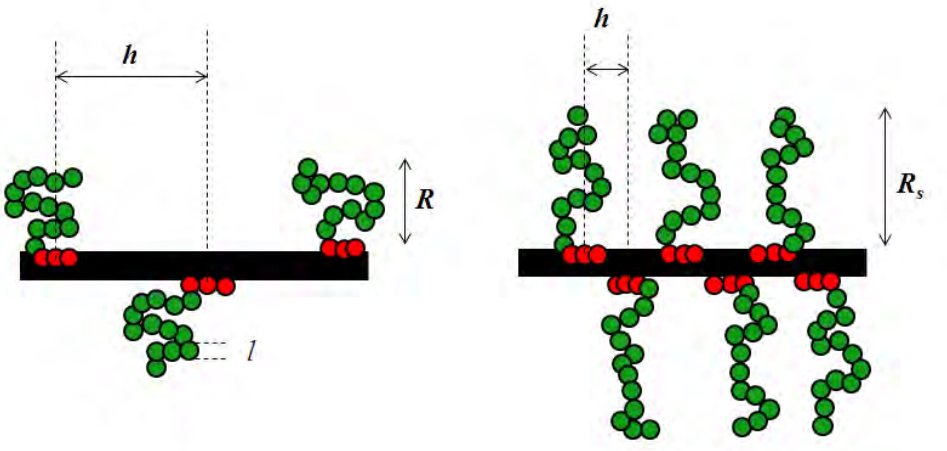


Figure 2.9. Schematic representation of supercoiled DNA interacting with cationic-neutral diblock copolymers. Thick black line is dsDNA, green circles are monomers of the colloidal block, red circles are monomers of the binding block. Numbers do not match the actual number of monomers used in this study. Left image: dilute brush $h < R$, where h is the grafting density and R is the (unstretched) rms end-to-end distance of the neutral block. Right image: brush with stretched side chains, $h < R_s$ where R_s is the rms end-to-end distance of the stretched side chains.

A further issue that is highly debated in the literature on bottle-brush polymers is the stiffening of the main chain, induced by the side chains.[7, 45] Current simulations suggest that at least for relatively short side-chains ($N < 100$) the contribution $l_{p,s}$ of the side chains to the total persistence length l_p of dense bottle-brush polymer is typically on the order of its radius R_s (see Figure 2.9).[7] This would mean that for rather short colloidal stability blocks, adhering diblocks can never straighten long DNA to the extent that it becomes rod-like. On the other hand, according to numerical self-consistent field calculations, very strong stiffening may occur for densely grafted, extremely long side chains ($N >> 1000$):[42]

$$l_{p,s} \approx \mu \cdot \frac{l^3}{h^2} N^2 \qquad (2.10)$$

with a very small numerical prefactor, $\mu \approx 0.02$. For our case (again considering the saturation case, full charge neutralization, h = 2 nm), the latter equation would

imply $l_{p,s} \approx 100\,\text{nm}$ for a single coated dsDNA and $l_p \approx 2(l_{p,0} + l_{p,s}) \approx 300nm$ for a coated plectoneme, assuming the two duplexes are both coated and are not bridged, and using $l_{p,0} \approx 50\,\text{nm}$ for the intrinsic persistence length of uncoated dsDNA. This order-of-magnitude seems to be consistent with the nearly rod-like shape of the fully coated 3.7kb supercoiled plasmid molecules in AFM. If we consider the entire supercoil as a semiflexible object, its estimated persistence length after coating would increase to about half of the total supercoil contour length of about 650nm, implying that it should indeed appear as close to rod-like.

2.4 Concluding Remarks

The short binding block and long colloidal stability block of the protein polymers that we have designed and produced play an important role in the reliable coating of single dsDNA molecules that we observe. Effectively, they prevent bridging and, hence, intramolecular folding and intermolecular aggregation. Indeed, even during a six-day incubation, complexes of C_4-B^{K12}-dsDNA did not show any aggregation (See Appendix, Figure 2.10). This is in agreement with coarse-grained simulations that have shown that a large colloidal stability block interfered with the packing of single polyanion molecules.[44] The increased stabilization could be because the C_4 block is highly hydrophilic and relatively inert, which is a desired property for biomedical applications.[13] Also, the five negative charges present along the C_4 block could contribute to the stabilization. On the other hand, the steric repulsion also affects the coating process of DNA: once a certain density of diblock copolymers is bound to DNA, steric repulsion starts to oppose the binding of further diblock copolymers. These physical properties of C_4 block do not have any temperature dependence in opposite to reported protein diblocks using as thermo-responsive elastine-like sequences which condense DNA instead of just coating single DNA molecules without any folding.[29]

The short length of the binding block contributes to the absence of bridging in two ways. First, it leads to a high density of steric blocks along the DNA duplex.

Second, the binding strength is not as high as for much longer binding blocks. Hence binding competes less efficiently with the steric repulsion due to the colloidal stability domain. As a consequence charge neutralization is incomplete and the complex has a net negative charge.

In other studies, formation of monomolecular DNA complexes with diblock copolymers carrying a larger binding block and a shorter protective block has been achieved by combining either a low DNA concentration and/or short incubation times.[17, 18] In this way, monomolecular DNA coating is expected to occur through kinetic intermediates that ultimately will aggregate and/or fold over time in order to decrease the total free energy. Indeed, it was observed that after 24h of incubation at low temperature PEG-poly-Lys(17-19) diblock copolymers folded the plasmid DNA (observed as a polydisperse population of multi-folded monomolecular DNA thick rods) upon complexation.[19, 20] Even monomolecular assembly using homo cationic poly-Lysine (19 – 117 amino acids) or poly-R-Ornitine (227 amino acids) is attainable, but only at very low DNA concentration (~3mg/L).[24, 25]

The complexes formed between pDNA and our highly asymmetric diblock copolymers showed rod morphology that resembles thin filamentous viruses along large range of N/Ps. This remarkable property could be the used as initial step into developing viral-mimetic polypeptides.[46, 47] This contrasts with DNA complexes displaying shapes such as thick compact rods with multiple DNA molecules, toroids or spheres attained with diblock polymers that carry a larger binding block and longer shielding group.[9, 19, 22] A filamentous morphology has been proven to give better stability against clearance and increased circulation time for *in vivo* delivery.[48]

Coating DNA strands could have several applications in DNA nanotechnology, such as DNA templating,[3] translocation of stretched single DNA molecules through nanopores,[5] and in non-viral gene therapy by giving protection against DNAse degradation.[49] If further stabilization of the coating would be necessary,

cross-linking of either the shielding block or the binding block could be considered.[50, 51]

In conclusion, two artificial protein block copolymers, C_4-B^{K12} and C_4-$B^{(HK)6}$, carrying a hydrophilic bulky colloidal stability domain and a short DNA binding domain were produced as secreted monodisperse proteins in the yeast *P. pastoris* at high yields. Owing to their bulky and rather stabilizing block and short binding block, the polymers form a polymer brush around each single DNA molecule and do not form bridges between different patches of the same DNA molecule or between different DNA molecules. This results in semi-flexible protein-DNA nanorods that are highly soluble and stable in time, each rod containing one single DNA molecule. These DNA complexes could be very suitable in nanotechnology and in novel functional biomaterials. The diblock copolymers presented here may serve as scaffolds for further incorporation of functional blocks, for instance, to develop viral-mimetic polypeptides.

2.5 References

1. Hall, A. R.; van Dorp, S.; Lemay, S. G.; Dekker, C., Electrophoretic Force on a Protein-Coated DNA Molecule in a Solid-State Nanopore. *Nano Lett.* 2009, *9*, 4441-4445.

2. Hegner, M.; Smith, S. B.; Bustamante, C., Polymerization and mechanical properties of single RecA-DNA filaments. *Proc. Natl. Acad. Sci. U. S. A.* 1999, *96*, 10109-10114.

3. Keren, K.; Berman, R. S.; Buchstab, E.; Sivan, U.; Braun, E., DNA-Templated Carbon Nanotube Field-Effect Transistor. *Science* 2003, *302*, 1380-1382.

4. Kowalczyk, S. W.; Hall, A. R.; Dekker, C., Detection of Local Protein Structures along DNA Using Solid-State Nanopores. *Nano Lett.* 2009, *10*, 324-328.

5. Smeets, R. M. M.; Kowalczyk, S. W.; Hall, A. R.; Dekker, N. H.; Dekker, C., Translocation of RecA-Coated Double-Stranded DNA through Solid-State Nanopores. *Nano Lett.* 2009, *9*, 3089-3095.

6. Wong, S. Y.; Pelet, J. M.; Putnam, D., Polymer systems for gene delivery-past, present, and future. *Prog. Polym. Sci.* 2007, *32*, 799-837.

7. Feuz, L.; Leermakers, F. A. M.; Textor, M.; Borisov, O., Bending rigidity and induced persistence length of molecular bottle brushes: A self-consistent-field theory. *Macromolecules* 2005, *38*, 8891-8901.

8. Douglas, S. M.; Dietz, H.; Liedl, T.; Hogberg, B.; Graf, F.; Shih, W. M., Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* 2009, *459*, 414-418.

9. Lam, J. K. W.; Ma, Y.; Armes, S. P.; Lewis, A. L.; Baldwin, T.; Stolnik, S., Phosphorylcholine-polycation diblock copolymers as synthetic vectors for gene delivery. *J. Controlled Release* 2004, *100*, 293-312.

10. Martin, M. E.; Rice, K. G., Peptide-guided gene delivery. *AAPS J.* 2007, *9*, E18-E29.

11. Numata, K.; Subramanian, B.; Currie, H. A.; Kaplan, D. L., Bioengineered silk protein-based gene delivery systems. *Biomaterials* 2009, *30*, 5775-5784.

12. Walker, G. F.; Fella, C.; Pelisek, J.; Fahrmeir, J.; Boeckle, S.; Ogris, M.; Wagner, E., Toward synthetic viruses: Endosomal pH-triggered deshielding of targeted polyplexes greatly enhances gene transfer in vitro and in vivo. *Mol. Ther.* 2005, *11*, 418-425.

13. Kabanov, A. V.; Kabanov, V. A., Interpolyelectrolyte and block ionomer complexes for gene delivery: Physicochemical aspects. *Adv. Drug Delivery Rev.* 1998, *30*, 49-60.

14. Kakizawa, Y.; Kataoka, K., Block copolymer micelles for delivery of gene and related compounds. *Adv. Drug Delivery Rev.* 2002, *54*, 203-222.

15. Kostiainen, M. A.; Szilvay, G. R.; Lehtinen, J.; Smith, D. K.; Linder, M. B.; Urtti, A.; Ikkala, O., Precisely defined protein-polymer conjugates: construction of synthetic DNA binding domains on proteins by using multivalent dendrons. *ACS Nano* 2007, *1*, 103-113.

16. Canine, B. F.; Wang, Y.; Hatefi, A., Evaluation of the effect of vector architecture on DNA condensation and gene transfer efficiency. *J. Controlled Release* 2008, *129*, 117-123.

17. DeRouchey, J.; Walker, G. F.; Wagner, E.; Radler, J. O., Decorated Rods: A "Bottom-Up" Self-Assembly of Monomolecular DNA Complexes. *J. Phys. Chem. B* 2006, *110*, 4548-4554.

18. Hartmann, L.; Haefele, S.; Peschka-Suess, R.; Antonietti, M.; Boerner, H. G., Tailor-made poly(amidoamine)s for controlled complexation and condensation of DNA. *Chem.--Eur. J.* 2008, *14*, 2025-2033.

19. Osada, K.; Oshima, H.; Kobayashi, D.; Doi, M.; Enoki, M.; Yamasaki, Y.; Kataoka, K., Quantized Folding of Plasmid DNA Condensed with Block Catiomer into Characteristic Rod Structures Promoting Transgene Efficacy. *J. Am. Chem. Soc.* 2010, *132*, 12343-12348.

20. Osada, K.; Yamasaki, Y.; Katayose, S.; Kataoka, K., A synthetic block copolymer regulates S1 nuclease fragmentation of supercoiled plasmid DNA. *Angew. Chem., Int. Ed.* 2005, *44*, 3544-3548.

21. Putnam, D.; Zelikin, A. N.; Izumrudov, V. A.; Langer, R., Polyhistidine-PEG : DNA nanocomposites for gene delivery. *Biomaterials* 2003, *24*, 4425-4433.

22. Tang, R.; Palumbo, R. N.; Nagarajan, L.; Krogstad, E.; Wang, C., Well-defined block copolymers for gene delivery to dendritic cells: Probing the effect of polycation chain-length. *J. Controlled Release* 2010, *142*, 229-237.

23. Zeng, J. M.; Wang, S., Enhanced gene delivery to PC12 cells by a cationic polypeptide. *Biomaterials* 2005, *26*, 679-686.

24. Mann, A.; Khan, M. A.; Shukla, V.; Ganguli, M., Atomic force microscopy reveals the assembly of potential DNA "nanocarriers" by poly-L-ornithine. *Biophys. Chem.* 2007, *129*, 126-136.

25. Mann, A.; Richa, R.; Ganguli, M., DNA condensation by poly-L-lysine at the single molecule level: Role of DNA concentration and polymer length. *J. Controlled Release* 2008, *125*, 252-262.

26. Hartmann, L.; Boerner, H. G., Precision Polymers: Monodisperse, Monomer-Sequence-Defined Segments to Target Future Demands of Polymers in Medicine. *Adv. Mater.* 2009, *21*, 3425-3431.

27. Jutz, G.; Boeker, A., Bionanoparticles as functional macromolecular building blocks - A new class of nanomaterials. *Polymer* 2011, *52*, 211-232.

28. Lamarre, B.; Ryadnov, M. G., Self-Assembling Viral Mimetics: One Long Journey with Short Steps. *Macromol. Biosci.* 2011, *11*, 503-513.

29. Chen, T. H. H.; Bae, Y.; Furgeson, D. Y., Intelligent biosynthetic nanobiomaterials (IBNs) for hyperthermic gene delivery. *Pharm. Res.* 2008, *25*, 683-691.

30. Kopecek, J., Smart and genetically engineered biomaterials and drug delivery systems. *Eur. J. Pharm. Sci.* 2003, *20*, 1-16.

31. Rabotyagova, O. S.; Cebe, P.; Kaplan, D. L., Protein-Based Block Copolymers. *Biomacromolecules* 2011, *12*, 269-289.

32. Schor, M.; Martens, A. A.; Dewolf, F. A.; Stuart, M. A. C.; Bolhuis, P. G., Prediction of solvent dependent beta-roll formation of a self-assembling silk-like protein domain. *Soft Matter* 2009, *5*, 2658-2665.

33. Werten, M. W. T.; Wisselink, W. H.; Jansen-van den Bosch, T. J.; de Bruin, E. C.; de Wolf, F. A., Secreted production of a custom-designed, highly hydrophilic gelatin in Pichia pastoris. *Protein Eng.* 2001, *14*, 447-454.

34. Werten, M. W. T.; Teles, H.; Moers, A. P. H. A.; Wolbert, E. J. H.; Sprakel, J.; Eggink, G.; de Wolf, F. A., Precision Gels from Collagen-Inspired Triblock Copolymers. *Biomacromolecules* 2009, *10*, 1106-1113.

35. Zhang, W. H.; Bevins, M. A.; Plantz, B. A.; Smith, L. A.; Meagher, M. M., Modeling Pichia pastoris growth on methanol and optimizing the production of a recombinant protein, the heavy-chain fragment C of botulinum neurotoxin, serotype A. *Biotechnol. Bioeng.* 2000, *70*, 1-8.

36. Osada, K.; Oshima, H.; Kobayashi, D.; Doi, M.; Enoki, M.; Yamasaki, Y.; Kataoka, K., Quantized Folding of Plasmid DNA Condensed with Block Catiomer into Characteristic Rod Structures Promoting Transgene Efficacy. *J. Am. Chem. Soc.* 2010, *132*, 12343-12348.

37. DeRouchey, J.; Walker, G. F.; Wagner, E.; Radler, J. O., Decorated Rods: A Bottom-Up Self-Assembly of Monomolecular DNA Complexes. *J. Phys. Chem. B* 2006, *110*, 4548-4554.

38. Dmochowska, A.; Dignard, D.; Henning, D.; Thomas, D. Y.; Bussey, H., Yeast kex1 gene encodes a putative protease with a carboxypeptidase-b-like function involved in killer toxin and alpha-factor precursor processing. *Cell* 1987, *50*, 573-584.

39. Bussiek, M.; Mucke, N.; Langowski, J., Polylysine-coated mica can be used to observe systematic changes in the supercoiled DNA conformation by scanning force microscopy in solution. *Nucleic Acids Res.* 2003, *31*.

40. Mandelkern, M.; Elias, J. G.; Eden, D.; Crothers, D. M., The dimensions of DNA in solution. *J. Mol. Biol.* 1981, *152*, 153-161.

41. Ainavarapu, S. R. K.; Brujic, J.; Huang, H. H.; Wiita, A. P.; Lu, H.; Li, L.; Walther, K. A.; Carrion-Vazquez, M.; Li, H.; Fernandez, J. M., Contour length and refolding rate of a small protein controlled by engineered disulfide bonds. *Biophys. J.* 2007, *92*, 225-33.

42. Hsu, H.-P.; Paul, W.; Binder, K., One- and two-component bottle-brush polymers: Simulations compared to theoretical predictions. *Macromol. Theory Simul.* 2007, *16*, 660-689.

43. DeRouchey, J.; Schmidt, C.; Walker, G. F.; Koch, C.; Plank, C.; Wagner, E.; Radler, J. O., Monomolecular assembly of siRNA and poly(ethylene glycol)-peptide copolymers. *Biomacromolecules* 2008, *9*, 724-732.

44. Ziebarth, J.; Wang, Y., Coarse-Grained Molecular Dynamics Simulations of DNA Condensation by Block Copolymer and Formation of Core-Corona Structures. *J. Phys. Chem. B* 2010, *114*, 6225-6232.

45. Stigter, D.; Dill, K. A., Binding of ionic ligands to polyelectrolytes. *Biophys. J.* 1996, *71*, 2064-2074.

46. Lee, J.-h.; Choi, Y. J.; Lim, Y.-b., Self-assembled filamentous nanostructures for drug/gene delivery applications. *Expert Opin. Drug Delivery* 2010, *7*, 341-351.

47. Lim, Y. B.; Lee, E.; Yoon, Y. R.; Lee, M. S.; Lee, M., Filamentous artificial virus from a self-assembled discrete nanoribbon. *Angew. Chem., Int. Ed.* 2008, *47*, 4525-4528.

48. Geng, Y.; Dalhaimer, P.; Cai, S.; Tsai, R.; Tewari, M.; Minko, T.; Discher, D. E., Shape effects of filaments versus spherical particles in flow and drug delivery. *Nat. Nanotechnol.* 2007, *2*, 249-255.

49. Itaka, K.; Yamauchi, K.; Harada, A.; Nakamura, K.; Kawaguchi, H.; Kataoka, K., Polyion complex micelles from plasmid DNA and poly(ethylene glycol)-poly(L-lysine) block copolymer as serum-tolerable polyplex system: physicochemical properties of micelles relevant to gene transfection efficiency. *Biomaterials* 2003, *24*, 4495-4506.

50. Miyata, K.; Kakizawa, Y.; Nishiyama, N.; Harada, A.; Yamasaki, Y.; Koyama, H.; Kataoka, K., Block catiomer polyplexes with regulated densities of charge and disulfide cross-linking directed to enhance gene expression. *J. Am. Chem. Soc.* 2004, *126*, 2355-2361.

51. Oupicky, D.; Parker, A. L.; Seymour, L. W., Laterally stabilized complexes of DNA with linear reducible polycations: Strategy for triggered intracellular activation of DNA delivery vectors. *J. Am. Chem. Soc.* 2002, *124*, 8-9.

2.6 Appendix

Table 2.2. Complementary oligonucleotides sequences used to construct the B^{K12} and $B^{(HK)6}$ binding blocks.

B^{K12}	Amino acids: KKKKKKKKKKKG*	
FW	48bp	5'*GTGGTAAGAAAAAGAAGAAAAAGAAGAAGAAGAAAAAGAAGGGTTAAG* 3'
RW	55bp	5'*AATTCTTAACCCTTCTTTTTCTTCTTCTTCTTTTTCTTCTTTTTCTTACCACCGG* 3'
$B^{(HK)6}$	Amino acids: HKHKHKHKHKHG*	
FW	48bp	5' *GTGGTCATAAGCACAAGCATAAACATAAGCATAAGCACAAAGGTTAAG* 3'
RW	55bp	5' *AATTCTTAACCTTTGTGCTTATGCTTATGTTTATGCTTGTGCTTATGACCACCGG* 3'

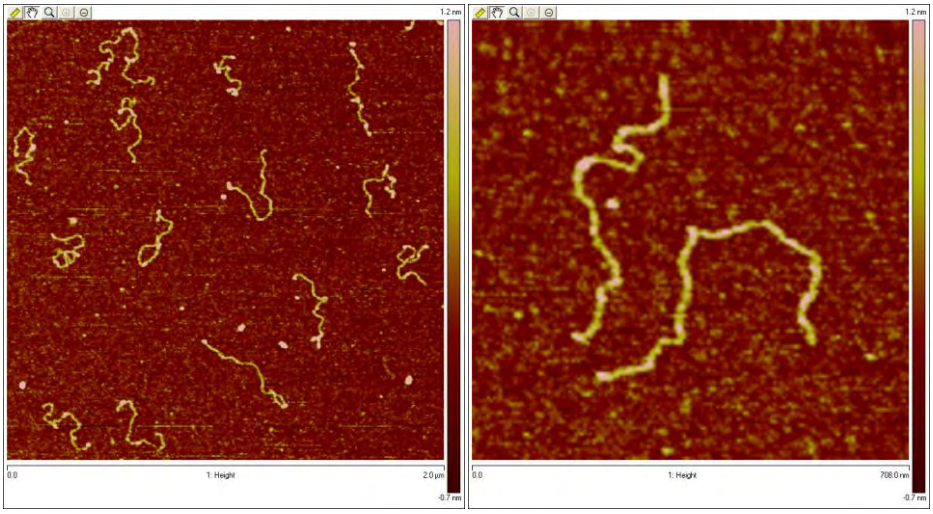


Figure 2.10. Atomic Force Microscopy images in air of typical rod-like shape nanoplexes formed between C_4-B^{K12} protein diblock copolymer and dsDNA 2.5 kb at N/P = 2 deposited on mica after 6 days incubation at room temperature (phosphate buffer 10 mM, pH 7.4). Sample was prepared similarly to that described in the methods section. The nanorods look well dispersed and intact even when incubated for such a long time without any antimicrobial agent. Left image size is 2x2 μm, right one is 708x708 nm.

Chapter 3

Supramolecular Design of a Minimal Coat-Protein for an Artificial Virus[1]

1 Chapter based on letter: Armando Hernandez-Garcia, Daniela J. Kraft, Anne F. J. Janssen, Paul H. H. Bomans, Nico A. J. M. Sommerdijk, Marco Favretto, Roland Brock, Frits A. de Wolf, Marc W. Werten, Paul van der Schoot, Martien Cohen Stuart, Renko de Vries. Supramolecular Design of a Minimal Coat-Protein for an Artificial Virus. Submitted, 2013.

3.1 Introduction

Virus particles are highly effective vehicles to deliver genetic material into susceptible host cells.[1] The simplest of viruses consist of no more than a polynucleotide genome encapsulated in a protective shell made up of multiple copies of a single coat protein.[2, 3] Not surprisingly, there have been many attempts to mimic this concept and to construct self-assembling virus-like particles.[4,5,6,7,8,9,10] A necessary condition identified for the successful *in-vitro* formation of infective virus particles is precisely tuned co-operativity of the self-assembly process.[3,11,12] Natural viruses seem have this property, which is also highlighted by theoretical models for their self-assembly,[13] but to date it has not been explicitly incorporated in any design of artificial viruses. We here rationally design a minimal viral coat protein which does feature precise control over the co-operativity of its co-assembly with nucleic acids. The design is inspired by the structure of the coat protein of Tobacco Mosaic Virus (TMV), and based on requirements identified by theories for its assembly.[13] Our biosynthetic artificial viral coat protein consists of three simple polypeptide domains for nucleic acid binding, self-assembly and shielding. In contrast to previous designs, we find highly co-operative, spontaneous co-assembly with single DNA molecules into rod-shaped virus-like-particles (VLPs). We confirm the validity of our design principles by showing that the kinetics of self-assembly of our VLPs follows our previous model for TMV assembly. Mature VLPs protect DNA against enzymatic degradation and transfect HeLa cells with considerable efficiency, making them promising delivery vehicles for nucleic acids, for example for gene therapy. In addition they are interesting monodisperse one-dimensional templates for materials science. Being biosynthetic and protein-based, our design paves the way for developing viruses that are completely artificial and yet can replicate in a cellular host.[30]

3.2 Methods Summary

3.2.1. Production of Polypeptides. Genes coding for the proteins and recombinant strains of *Pichia pastoris* were prepared using standard methods. Proteins were

secreted into the medium during fermentation and purified from the filtered cell-free medium by fractional precipitation. Dialyzed and freeze-dried proteins were stored in sealed tubes. Purity and integrity was corroborated by SDS-PAGE and MALDI-TOF MS.

3.2.2. Sample Preparation. DNA-protein complexes were prepared by mixing aliquots of stock solutions of DNA (*NoLimits* Thermo Scientific, Waltham, MA) and protein in buffer (10mM phosphate, pH 7.4, containing 0.1 mM DTT) at the desired charge ratio N/P (molar ratio between positively charged NH_2 groups (N) from binding block "B" to negatively charged PO_3 groups of the DNA template (P)). Samples were incubated at room temperature.

3.2.3. Atomic Force Microscopy (AFM). 3-5 μL of sample ([DNA] = 1 μg/mL) were deposited onto a clean silicon surface. After 2 minutes the surface was rinsed with 1 mL of Milli-Q water to remove salts and non-absorbed particles, followed by soaking up of excess water using a tissue and slow drying under a N_2 stream.

3.2.4. Cryo-Transmission Electron Microscopy (Cryo-TEM). Samples ([DNA] = 30 μg/mL) were prepared using a vitrification robot (FEI Vitrobot Mark III) using grids R2/2 (Quantifoil Micro Tools GmbH, Jena, Germany) previously surface plasma treated (Cressington 208 carbon coater). The cryo-TEM characterizations were performed on a CryoTitan (FEI) equipped with a field emission gun operating at 300 kV and a postcolumn Gatan energy filter. Images were recorded using a 2 k×2 k Gatan CCD camera.

3.2.5. Electrophoretic Mobility Shift Assay (EMSA). DNA-protein samples ([DNA] = 15 ng/μL) incubated for 1.5h were electrophoresed in a 1% agarose gel (95 min / 60V) using 1x TAE buffer (pH 8). Bands were visualized with ethidium bromide.

3.3 Results and Discussion

Viruses are among the simplest biological systems and can be considered a testing ground for developing artificial life. Artificial viruses are potentially useful as model systems for providing insights into natural viruses. They have biomedical and biotechnological applications such as targeted delivery of nucleic acids in the context of gene therapy,[1,8] and as scaffolds in material science.[14,15,16] In Nature, survival of viruses requires that a significant fraction of replicated viral genomes is completely protected by coat proteins in infected host cells, based on just weak and reversible protein-protein and protein-nucleic acid bonds. Co-operative binding, e.g., through allostery, provides the mechanism for this.[3,11,12] A rational design of an artificial virus must therefore hinge on the control of co-operativity.

In natural viruses, such as Tobacco Mosaic Virus (TMV), co-operativity arises through allosteric conformational switching of the coat proteins upon binding nucleic acid, often initially at specific origin-of-assembly regions. Indeed, taking into account allostery and a specific origin-of-assembly region, we have recently put forward a model that accurately describes the in-vitro assembly kinetics of TMV.[13]

Based on these mechanistic insights, and inspired by the structure of the TMV coat protein[2] we here rationally design a minimalistic artificial viral coat protein in which each of three essential physicochemical functionalities of viral coat proteins are encoded into simple polypeptide blocks. For nucleic acid binding we use an oligolysine block called B^{K12} that binds sequence specifically through electrostatic interactions.[17] Precisely tuned co-operativity is provided by a silk-like sequence S_n = $(GAGAGAGQ)_n$ that in solution folds and stacks into stiff filamentous structures.[18,19] The number of silk strands n dictates the level of co-operativity of the binding of the protein and thus is our key variable. To prevent aggregation of the assembled artificial virus particles we include a 407 amino acid-long hydrophilic random-coil sequence called C_4 with a high fraction of prolines and hydrophilic (mostly uncharged) amino acids.[20] The latter block may be modified to

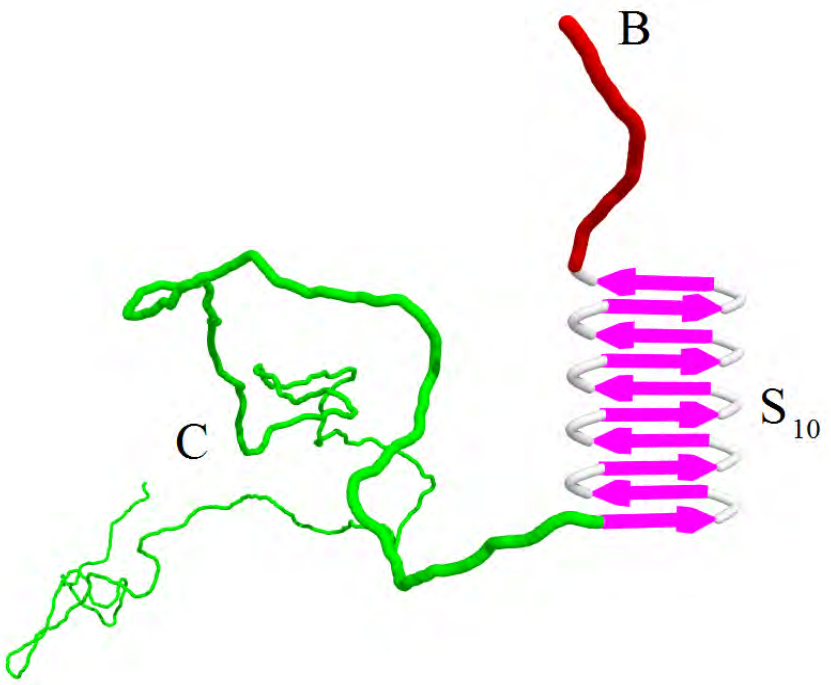


Figure 3.1. Illustration of the C_4-S_n-B^{K12} protein-based polymer design. Shown is the case of n = 10, where B^{K12} is the C-terminal dodecalysine DNA binding block (in red), S_{10} is a tenfold repetition of a beta-roll forming octapeptide S = (GAGAGAGQ) that constitutes the self-assembly block (in pink), and C_4 is a N-terminal 407 amino acid hydrophilic random coil block (in green).

encode other interactions with the environment as well. Polypeptides C_4-S_n-B^{K12} (shown schematically in Figure 3.1) with n = 0, 2, 4, 10 and 14 were produced biosynthetically by expression in the yeast *Pichia pastoris*. For details on production, purification and characterization see the Appendix section 3.6.1, Table 3.1, Table 3.2 and Figure 3.7.

We will now show that beyond a certain number of repeats n of the central block S_n our rationally designed artificial coat proteins behave in many ways as coat proteins of natural viruses: binding to nucleic acids induces folding and stacking of the blocks S_n, which leads to the co-operative formation of compact, rod-shaped virus-like particles (VLPs). Each of the VLPs encapsulates a single nucleic acid

molecule. The assembly kinetics of the VLPs follows the theoretical model for TMV assembly kinetics that inspired the original design; inside the VLPs, nucleic acids are protected against degradation by nucleases, and HeLa cells are transfected by VLPs with an efficiency that is similar to that of current non-viral gene delivery agents. Our design goes far beyond existing approaches for artificial viruses that do not feature precise control over co-operativity.[5,6,7,8,9,10,21,22] Being purely biosynthetic and protein-based, our VLPs can be a stepping stone to the development of a completely artificial virus that self-replicates in a cellular host.[30]

A small number of repeats of the central block S_n is insufficient to induce the co-operative formation of compact VLPs. Complexes of linear double stranded DNA with polypeptides C_4-S_n-B^{K12}, with n = 2, 4 have contour lengths that are exactly equal to that of the dsDNA template, and retain much of its original flexibility (Figure 3.2b-2c). These complexes have a 'bottle brush' structure that we have previously found for the polypeptide C_4-S_0-B^{K12} without the central block (Figure 3.2a): a single DNA chain surrounded by a dense array of pendant side chains (thickness in solution around 15 nm, see Appendix Figure 3.8). Other groups have also demonstrated non-co-operative coating of single nucleic acid molecules by polymers featuring DNA-binding blocks in combination with large hydrophilic shielding blocks.[6,17,21,22,23]

We do, however, achieve full co-operativity with C_4-S_n-B^{K12} polypeptides for n = 10 and 14. In this case, rod-shaped VLPs are formed with contour lengths much smaller than that of the DNA template, as shown in Figure 3.2d-2g (Appendix Figure 3.9 and 3.10). Over a wide range of template lengths, DNA in VLPs is compacted to about 1/3 of its original length (Appendix Figure 3.11 and 3.12). From the solution molar mass of the complexes, we estimate that 95% of the DNA phosphate charges are neutralized by compensating charges on binding blocks B (Appendix Figure 3.13). Consistent with the idea that charge neutralization determines the stoichiometry of the VLPs, the compaction factor found for single stranded DNA is 1/6 (Appendix Figure 3.14). Similar to natural viral capsid proteins, the VLP-forming polypeptides tend to form rod-like self-assemblies in

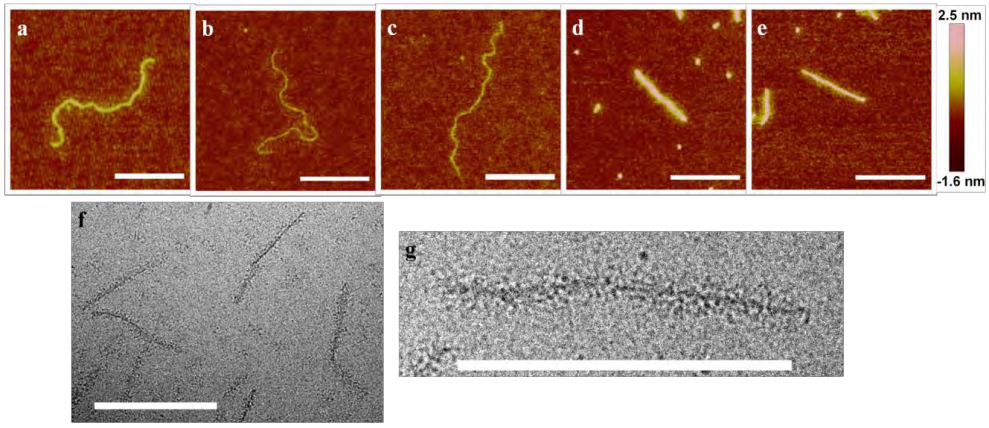


Figure 3.2. Dependence of DNA + C_4-S_n-B^{K12} polymer complex morphology on size *n* of the silk-like self-assembly block. AFM and Cryo-TEM images of complexes with linear dsDNA (2,500 bp). (a-c) *n* = 0, 2 and 4: non-cooperative coating of DNA templates, charge ratios N/P = 10 (*n* = 0) and N/P = 2 (*n* = 2 and 4), incubation time 1h. Contour lengths of complexes are close to that of the template DNA (850nm): for, respectively, *n* = 0, 2, 4, contour lengths of complexes shown are 830 nm, 869 nm, and 837 nm. (d-e) *n* = 10 and 14: formation of virus-like particles (VLPs). Charge ratio N/P = 7, incubation time 24h. Contour lengths of VLPs shown in the images are 315 nm (*n* = 10) and 304 nm (*n* = 14). f) and g) cryo-TEM images of VLPs for *n* = 10, N/P = 7, incubation time 24h. Scale bar in all images is 300 nm.

the absence of a template as well, but only at much higher concentrations (Appendix Figure 3.15).

Like natural viruses, the VLPs protect their nucleic acid cargo against enzymatic attack and are capable of delivering their cargo into cells. To assess protection, we incubated VLPs with high concentrations of DNAse I and analyzed the degradation using agarose-gel electrophoresis. Conditions were chosen such that no intact bare DNA was present after 1 minute of incubation. We find that the complexes offer a protection that increases with the length *n* of the central block S_n. In particular, VLPs formed by C_4-S_n-B^{K12} polypeptides with *n* = 10 and 14 resist attack for times longer than 60 minutes (Figure 3.3). This suggests a link between compaction and protection of the nucleic acids and the co-operativity of the VLP assembly that we investigate in more detail below.

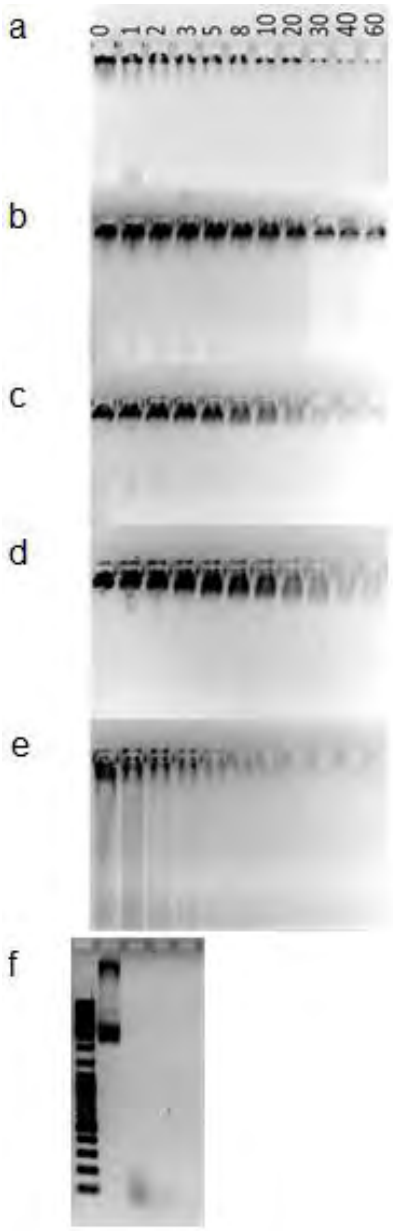


Figure 3.3. Enzymatic Degradation Assay of pDNA Complexed with Proteins. a) C_4-S_{14}-B^{K12}, b) C_4-S_{10}-B^{K12}, c) C_4-S_4-B^{K12}, d) C_4-S_2-B^{K12}, e) C_4-S_0-B^{K12} (or just C_4-B^{K12}) and f) naked DNA (first lane are the molecular markers). Protein-DNA complexes were incubated with DNAseI for different times and analyzed with agarose gel electrophoresis. The incubation time of complexes with enzyme in minutes is indicated at the top.

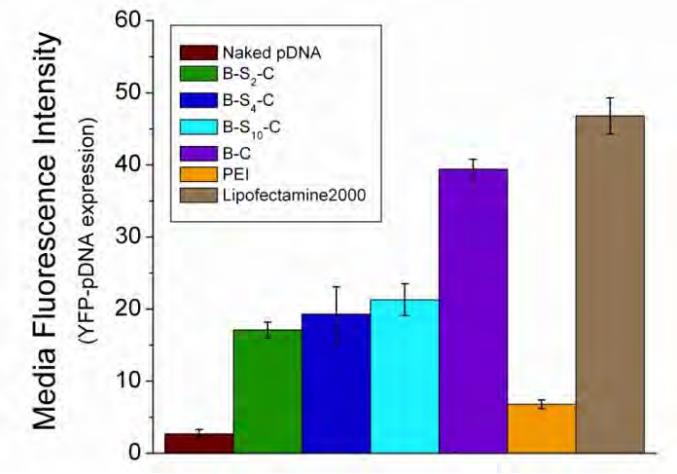


Figure 3.4. Transfection Efficiency of YGFP-pDNA complexed with different protein-based polymers and standard non-viral transfection agents.

Transfection of cells by complexes of plasmid DNA with C_4-S_n-B^{K12} polypeptides was tested for n = 0, 2, 4, and 10. All complexes transfect HeLa cells and give rise to the expression of a fluorescent reporter protein, with a similar efficiency as the established non-viral delivery standards poly-ethyleneimine and Lipofectamine 2000 (Figure 3.4). Interestingly, we find that polypeptides without the central block C_4-S_n-B^{K12} have the highest transfection efficiency and that it is somewhat lower for VLPs based on C_4-S_{10}-B^{K12}. We speculate that the reduced transfection efficiency is related to the more difficult disassembly that is concomitant with the increased protection due to co-operativity and the associated compaction of VLPs. This ties in with what is known about natural viruses that have to strike a balance between the degree of protection and ease of disassembly.[24]

A defining test for co-operativity is "all-or-nothing"-type behavior.[25,26] For *in-vitro* assembly of viruses, this translates into the co-existence of fully encapsulated and bare nucleic acids.[3, 13] Our AFM images of the early stages of the encapsulation of DNA by C-S_{10}-B show DNA that is either mostly empty or fully encapsulated

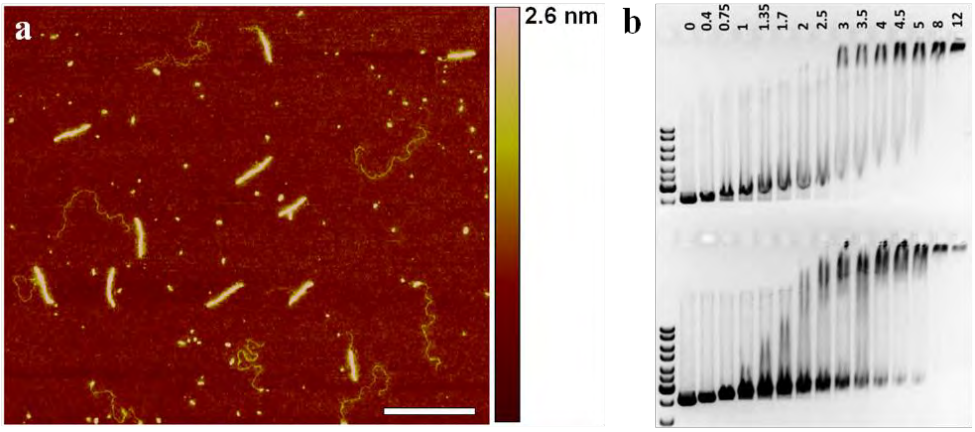


Figure 3.5. Cooperativity of VLP formation for C_4-S_{10}-B^{K12} polypeptide. a) AFM image of incompletely assembled virus-like particles (incubation time 1 hour) of C_4-S_{10}-B^{K12} with 2,500 bp linear dsDNA at charge ratio N/P = 15, [DNA] = 1 μg/mL. Smaller objects are protein-only assemblies (See Appendix Figure 3.15). Scale bar is 500nm b) Electrophoretic Mobility Shift Assay (EMSA) for C_4-S_{10}-B^{K12} interacting with 2,500 bp linear DNA (bottom), compared with that for C_4-B^{K12} diblock lacking the self-assembly domain (top). Charge ratio N/P for different lanes is indicated. EMSA for C-S_{10}-B shows coexistence of different types of complexes, but not that for the C_4-B^{K12} control polymer.

(Figure 3.5a). Cooperative binding is confirmed by electrophoretic mobility shift assays of DNA, incubated with increasing concentrations of polypeptide. For C_4-S_0-B^{K12} we observe a single type of complex with gradually shifting mobility (Figure. 3.5b, upper panel), whereas for C_4-S_{10}-B^{K12}, complexes of high mobility co-exist with complexes of low mobility (Figure 3b, lower panel). The latter indicates the co-existence of empty and fully encapsulated nucleic acids and hence supports our case for co-operative binding.

Nucleation of VLPs mostly occurs at one end of the DNA hence the ends of the DNA act as an effective origin of assembly. The TMV model that motivated our design should therefore be quantitatively applicable to the kinetics of our VLP assembly. We have quantified the time evolution of the distribution of lengths of condensed sections of VLPs from AFM images, and compared this with the predictions of the aforementioned kinetic model (Figure 3.6a-c).[13] The model assumes that a first subunit, which binds to an origin of assembly on a linear array

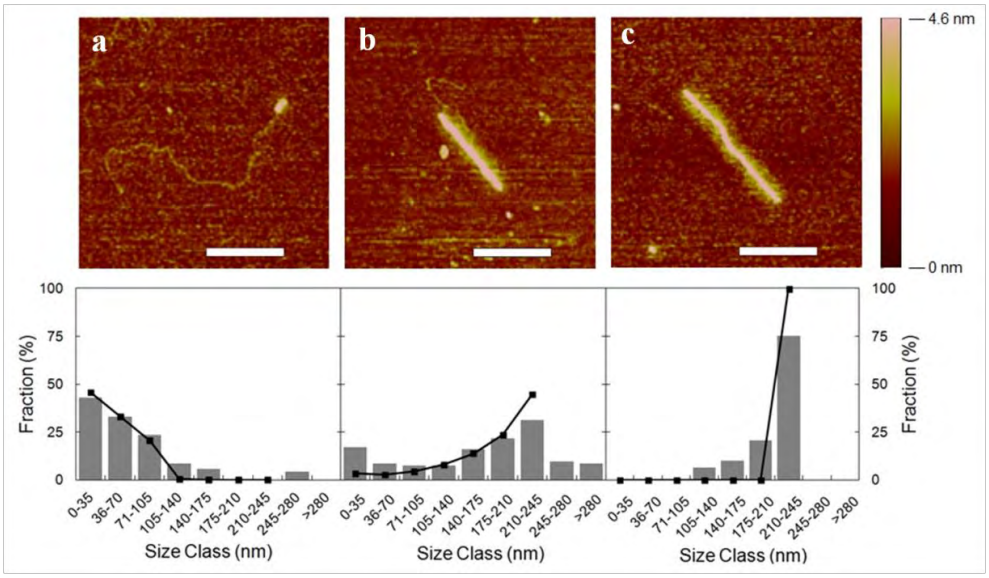


Figure 3.6. Kinetics of VLP formation. Top panel: representative AFM images of VLPs of 2,500 bp linear dsDNA with C_4-S_{10}-B^{K12} polymers at a charge ratio N/P = 3, [DNA] = 1 μg/mL. Bar = 200nm. Incubation times: a) 10 min b) 5h 50min c) 24h. Bottom row: size distributions of condensed sections as determined from AFM images (grey bars) and fit to kinetic model of TMV self-assembly[20] (black symbols, solid black lines).

of binding sites, pays a free energy penalty h due to conformational switching, while binding of subsequent subunits involves a favorable binding free energy, g. In addition, bound protein subunits gain a free energy ε per bond due to favorable protein-protein interactions. A comparison with our experimental data leads to estimated parameter values of $h-\varepsilon \approx 5.3k_BT$ for the effective penalty for nucleation, and $g+\varepsilon \approx -18k_BT$ for the effective binding free energy (Appendix Figure 3.16). These values are very similar in magnitude to those found for TMV.[13] again emphasizing the striking similarity between the assembly of our artificial VLPs and natural viruses such as TMV.

3.4 Summary and Conclusion

Based on a current understanding of the self-assembly of natural viruses we have rationally designed artificial viral coat proteins by combining multiple bio-inspired functional blocks of low sequence complexity. These protein-based block-copolymers[27] co-operatively form rod-shaped VLPs that protect the encapsulated DNA and feature significant transfection activity. The capsids that our polypeptides make are scaffold materials that could be optimized for targeted delivery of nucleic acids,[8] and may act as monodisperse templates in material science.[16, 23, 28] Our generic protein-based block-copolymer approach can be used to create many other intricate nanostructures.[29] Being completely biosynthetic, our design for an artificial viral coat protein paves the way for the development of completely artificial viruses that could self-replicate in a cellular host.[30]

3.5 References

1. Naldini, L.; Blomer, U.; Gallay, P.; Ory, D.; Mulligan, R.; Gage, F. H.; Verma, I. M.; Trono, D., In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* 1996, *272*, 263-267.

2. Namba, K.; Pattanayek, R.; Stubbs, G., Visualization of protein-nucleic acid interactions in a virus - refined structure of intact tobacco mosaic-virus at 2.9-a resolution by x-ray fiber diffraction. *Journal of Molecular Biology* 1989, *208*, 307-325.

3. Klug, A., The tobacco mosaic virus particle: structure and assembly. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 1999, *354*, 531-535.

4. Percec, V.; Heck, J.; Johansson, G.; Tomazos, D.; Ungar, G., Towards tobacco mosaic virus-like self-assembled supramolecular architectures. *Macromol. Symp.* 1994, *77*, 237-265.

5. Remy, J. S.; Kichler, A.; Mordvinov, V.; Schuber, F.; Behr, J. P., Targeted gene-transfer into hepatoma-cells with lipopolyamine-condensed dna particles presenting galactose ligands - a stage toward artificial viruses. *Proc. Natl. Acad. Sci. U. S. A.* 1995, *92*, 1744-1748.

6. Aoyama, Y.; Kanamori, T.; Nakai, T.; Sasaki, T.; Horiuchi, S.; Sando, S.; Niidome, T., Artificial viruses and their application to gene delivery. size-controlled gene coating with glycocluster nanoparticles. *J. Am. Chem. Soc.* 2003, *125*, 3455-3457.

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

8. Mastrobattista, E.; van der Aa, M.; Hennink, W. E.; Crommelin, D. J. A., Artificial viruses: a nanotechnological approach to gene delivery. *Nature Reviews Drug Discovery* 2006, *5*, 115-121.

9. Lim, Y.-b.; Lee, E.; Yoon, Y.-R.; Lee, M. S.; Lee, M., Filamentous artificial virus from a self-assembled discrete nanoribbon. *Angewandte Chemie-International Edition* 2008, *47*, 4525-4528.

10. Miyata, K.; Nishiyama, N.; Kataoka, K., Rational design of smart supramolecular assemblies for gene delivery: chemical challenges in the creation of artificial viruses. *Chemical Society Reviews* 2012, *41*, 2562-2574.

11. Caspar, D. L.; Namba, K., Switching in the self-assembly of tobacco mosaic virus. *Advances in biophysics* 1990, *26*, 157-85.

12. Butler, P. J. G., Self-assembly of tobacco mosaic virus: the role of an intermediate aggregate in generating both specificity and speed. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 1999, *354*, 537-550.

13. Kraft, D. J.; Kegel, W. K.; van der Schoot, P., A Kinetic Zipper Model and the Assembly of Tobacco Mosaic Virus. *Biophysical Journal* 2012, *102*, 2845-2855.

14. Comellas-Aragones, M.; Engelkamp, H.; Claessen, V. I.; Sommerdijk, N.; Rowan, A. E.; Christianen, P. C. M.; Maan, J. C.; Verduin, B. J. M.; Cornelissen, J.; Nolte, R. J. M., A virus-based single-enzyme nanoreactor. *Nat. Nanotechnol.* 2007, *2*, 635-639.

15. Grigoryan, G.; Kim, Y. H.; Acharya, R.; Axelrod, K.; Jain, R. M.; Willis, L.; Drndic, M.; Kikkawa, J. M.; DeGrado, W. F., Computational Design of Virus-Like Protein Assemblies on Carbon Nanotube Surfaces. *Science* 2011, *332*, 1071-1076.

16. Douglas, T.; Young, M., Host-guest encapsulation of materials by assembled virus protein cages. *Nature* 1998, *393*, 152-155.

17. Hernandez-Garcia, A.; Werten, M. W. T.; Stuart, M. C.; de Wolf, F. A.; de Vries, R., Coating of Single DNA Molecules by Genetically Engineered Protein Diblock Copolymers. *Small* 2012, *8*, 3491-3501.

18. Krejchi, M. T.; Atkins, E. D. T.; Waddon, A. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A., Chemical sequence control of beta-sheet assembly in macromolecular crystals of periodic polypeptides. *Science* 1994, *265*, 1427-1432.

19. Beun, L. H.; Beaudoux, X. J.; Kleijn, J. M.; de Wolf, F. A.; Stuart, M. A. C., Self-Assembly of Silk-Collagen-like Triblock Copolymers Resembles a Supramolecular Living Polymerization. *Acs Nano* 2012, *6*, 133-140.

20. Werten, M. W. T.; Wisselink, W. H.; van den Bosch, T. J. J.; de Bruin, E. C.; de Wolf, F. A., Secreted production of a custom-designed, highly hydrophilic gelatin in Pichia pastoris. *Protein Engineering* 2001, *14*, 447-454.

21. DeRouchey, J.; Walker, G. F.; Wagner, E.; Radler, J. O., Decorated rods: A "bottom-up" self-assembly of monomolecular DNA complexes. *Journal of Physical Chemistry B* 2006, *110*, 4548-4554.

22. Osada, K.; Oshima, H.; Kobayashi, D.; Doi, M.; Enoki, M.; Yamasaki, Y.; Kataoka, K., Quantized Folding of Plasmid DNA Condensed with Block Catiomer into Characteristic Rod Structures Promoting Transgene Efficacy. *J. Am. Chem. Soc.* 2010, *132*, 12343-12348.

23. Ruff, Y.; Moyer, T.; Newcomb, C. J.; Demeler, B.; Stupp, S. I., Precision Templating with DNA of a Virus-like Particle with Peptide Nanostructures. *J. Am. Chem. Soc.* 2013, *135*, 8.

24. Lecuyer, K. A.; Behlen, L. S.; Uhlenbeck, O. C., Mutants of the bacteriophage-ms2 coat protein that alter its cooperative binding to RNA. *Biochemistry* 1995, *34*, 10600-10606.

25. Whitty, A., Cooperativity and biological complexity. *Nature Chemical Biology* 2008, *4*, 435-439.

26. Williamson, J. R., Cooperativity in macromolecular assembly. *Nature Chemical Biology* 2008, *4*, 458-465.

27. Rabotyagova, O. S.; Cebe, P.; Kaplan, D. L., Protein-Based Block Copolymers. *Biomacromolecules* 2011, *12*, 269-289.

28. Zhang, S. G., Fabrication of novel biomaterials through molecular self-assembly. *Nat. Biotechnol.* 2003, *21*, 1171-1178.

29. King, N. P.; Sheffler, W.; Sawaya, M. R.; Vollmar, B. S.; Sumida, J. P.; Andre, I.; Gonen, T.; Yeates, T. O.; Baker, D., Computational Design of Self-Assembling Protein Nanomaterials with Atomic Level Accuracy. *Science* 2012, *336*, 1171-1174.

30. Bromley, E. H. C.; Channon, K.; Moutevelis, E.; Woolfson, D. N., Peptide and protein building blocks for synthetic biology: From programming biomolecules to self-organized biomolecular systems. *ACS Chemical Biology* 2008, *3*, 38-50.

3.6 Appendix

3.6.1. Production, purification and characterization of protein-based polymers. Restriction enzymes and T4 DNA ligase were purchased from NEB (Ipswich, MA) or Thermo Scientific (Waltham, MA). GeneJET Plasmid miniprep kit and GeneJET Gel Extraction and DNA Cleanup Micro Kit were purchased from Thermo Scientific. Cloning of control protein C_4-B^{K12} was reported previously.[17] The cloning of the proteins incorporating the self-assembly block were done accordingly to the following procedure.

Cloning of Binding Block (B). A gene segment coding for binding block B (Table S1) was prepared by annealing two complementary oligonucleotides (Eurogentec, Belgium) in buffer 50 mM Tris-HCl containing 50 mM $MgCl_2$, 100 mM NaCl and 1mM dithioerythritol at pH 7.5 by heating at 95°C for 5 min and decreasing the temperature slowly for two or three hours (same procedure for annealing of other oligonucleotides). The resulting double-stranded segment was inserted into a modified pMTL23 vector, presently denoted as pMTL23ΔBsaI, from which the normal BsaI site has been removed.[22] After digestion of pMTL23Δ*Bsa*I with *Xho*I/*Eco*RI and ligation of the insert, pMTL23-Δ-*Bsa*I-B was obtained

Cloning of Colloidal Stability Block (C_4). A gene segment coding for the "C_4" block (Table S1) was obtained by digestion of the previously described vector pMTL23-C4 (ref. 17, chapter 2) with *Dra*III/*Van*91I. A pMTL23-Δ-*Bsa*I -adaptor with restriction sites required for coupling of the C_4-encoding segment to other building blocks (Table S1) was prepared by annealing two complementary oligonucleotides (Eurogentec, Belgium) and the double-stranded segment was ligated into the *Xho*I/*Eco*RI-digested plasmid pMTL23-Δ-*Bsa*I. After digestion of the resulting pMTL23-Δ-*Bsa*I-ad with *Dra*III and dephosphorylation, the C4 segment was ligated into it, so as to obtain pMTL23-Δ-*Bsa*I-C4.

Cloning of Self-assembly Blocks (S_n). A self-assembly block S_2 (Table S1) was prepared by annealing two complementary oligonucleotides (Eurogentec, Belgium); the resulting double stranded product was inserted into the *XhoI/EcoRI*

digested plasmid pMTL23-Δ-BsaI to get pMTL23-Δ-*BsaI*-S2. In order to get pMTL23-Δ-*Bsa*I-S4, an *Xho*I/*Ban*I-digested S_2 fragment from pMTL23-Δ-*Bsa*I-S2 was ligated into previously *Xho*I/*Bsa*I-digested pMTL23-Δ-*Bsa*I-S2. Similarly, to get pMTL23-Δ-*Bsa*I-S6, an *Xho*I/*Ban*I-digested S_2 fragment from pMTL23-Δ-*Bsa*I-S2 was ligated into previously *Xho*I/*Bsa*I-digested pMTL23-Δ-*Bsa*I-S4.

Cloning of C_4-S_n. The vector pMTL23-Δ-*Bsa*I-C4 was digested with *Xho*I/*Ban*I and the released insert was ligated into pMTL23-Δ-*Bsa*I-S2/4/6, previously digested with *Xho*I/*Bsa*I. In this way, pMTL23-Δ-*Bsa*I-C4-S2, pMTL23-Δ-*Bsa*I-C4-S4 and pMTL23-Δ-*Bsa*I-C4-S6 were obtained. The plasmid pMTL23-Δ-*Bsa*I-C4-S10 was obtained by ligation of an *Xho*I/*Ban*I fragment released from vector pMTL23-Δ-*Bsa*I-C4-S6 into *Xho*I/*Bsa*I-digested vector pMTL23-Δ-*Bsa*I-S4. Plasmid pMTL23-Δ-*Bsa*I-C4-S14 was obtained in a similar way but using pMTL23-Δ-*Bsa*I-C4-S10 instead of pMTL23-Δ-*Bsa*I-C4-S6.

Cloning of C_4-S_n-B^{K12} and incorporation into expression vector pPIC9. To finally obtain pMTL23-Δ-*Bsa*I-C4-S2-K12, pMTL23-Δ-*Bsa*I-C4-S4-K12, pMTL23-Δ-*Bsa*I-C4-S10-K12 and pMTL23-Δ-*Bsa*I-C4-S14-K12, which code for the artificial viral-coat polypeptides of this study, C4-S2, C4-S4, C4-S10 and C4-S14 segments were released from pMTL23-Δ-*Bsa*I-C4-S2, pMTL23-Δ-*Bsa*I-C4-S4, pMTL23-Δ-*Bsa*I-C4-S10 and pMTL23-Δ-*Bsa*I-C4-S14 by digestion with *Xho*1/*Ban*I and inserted into pMTL23-Δ-*Bsa*I-K12 that had been previously digested with *Xho*I/*Bsa*I. The fragments encoding for all the polypeptides were released through digestion of the corresponding plasmids with *XhoI/EcoRI* and ligated into the correspondingly digested *P. pastoris* expression vector pPIC9 (Invitrogen). The pPIC9 plasmid contains an HIS4 selectable marker, an alcohol oxidase 1(AOX1) promoter/terminator cassette and a *S. cerevisiae* α-factor prepro secretory signal. Sequences (Table S2) were verified by sequencing the plasmids. The plasmids pPIC9-C4-S2-K12, pPIC9-C4-S4-K12, pPIC9-C4-S10-K12 and pPIC9-C4-S14-K12 were linearized with *SalI* and electroporated into *P. pastoris* strain GS115 (Invitrogen). The plasmid integrates into the genome through homologous

Table 3.1. Oligonucleotide sequences used in cloning of the proteins.

	Binding block (B^{K12})
Fw	5'tcgagggagctgccggtctcggtgctaagaaaaagaagaaaaagaagaagaagaaaaagaagggtgcctaag 3'
Rv	5' aattcttaggcacccttcttttttcttcttcttcttttttcttcttttttcttagcaccgagaccggcagctccc 3'
	Self-assembly block (S_2)
Fw	5'tcgagggagctgccggtctcggtgctggtgctggtgctggtcaaggagccggtgctggagccggccaaggtgcctaag 3'
Rv	5'aattcttaggcaccttggccggctccagcaccggctccttgaccagcaccagcaccagcaccgagaccggcagctccc 3'
	Adaptor (for C_1-block)
Fw	5' tcgagaagagagaagctgaagctggttgtggtctcggtgctggtgcacccggtgctggtgcctaag 3'
Rv	5'aattcttaggcaccagcaccgggtgcaccagcaccgagaccacaaccagcttcagcttctctcttc 3'

recombination at the *his4* locus providing normal growth on methanol. The presence of the genes was verified by PCR.

Fermentation and purification. Biosynthesis of protein polymers was performed as described in ref. 17 except for small changes described below. Production was carried out by methanol fed-batch fermentation performed in 2.5-L Bioflo 3000 fermentors (New Brunswick Scientific, Edison, NJ) during two or three days at pH 3.0. At the end of the fermentation, the cells were separated from the broth by centrifugation for 15 min at 10000 × g (room temperature or 4 °C) in an SLA-3000 rotor (Thermo Scientific, Waltham, MA), and the supernatant was microfiltered (0.2 μm, Pall Corporation, Port Washington, NY) and immediately stored at 4 °C for subsequent purification. First, parts of the medium salts were removed from the cell-free broth by adjustment of the pH to 8.0 with NaOH, followed by centrifugation (20-30 min at 15000 × g at 4°C). Subsequently, the polypeptides were selectively precipitated from the solution by adding ammonium sulphate to a saturation of 45% for C_4-S_2-B^{K12} and C_4-S_4-B^{K12}, and 50% for C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12}. After the addition of ammonium sulphate the cell-free broth was incubated overnight at 4 °C, and subsequently centrifuged (15000 × g, SLA-1500, 30 min / 4°C). Pellets were resuspended in 300 ml of Milli-Q water heated at 60-65°C for 15 min (with frequent stirring) and precipitated overnight with ammonium sulphate at 50% saturation at 4°C for all the proteins. After centrifugation at 15,000 × g for

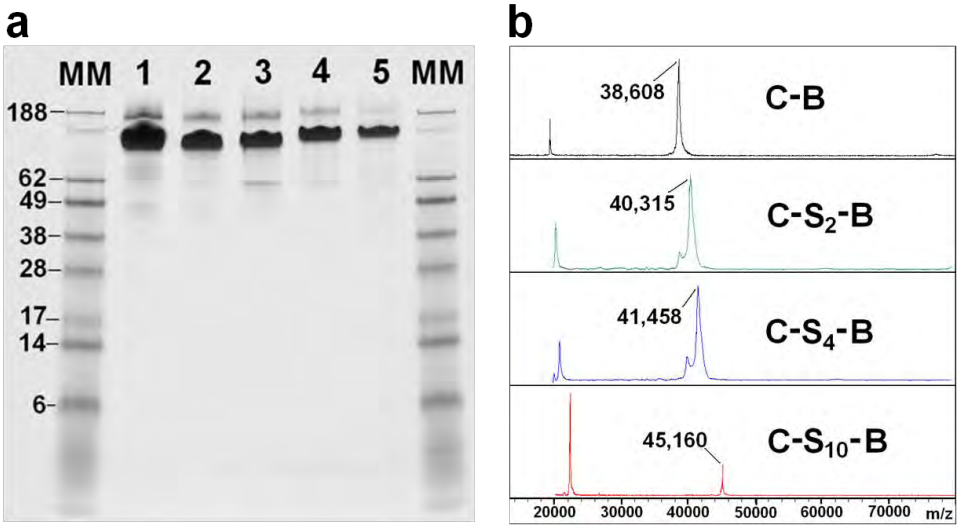


Figure 3.7. Molecular Characterization of Recombinant Proteins. A) SDS-PAGE, lane 1 C_4-B^{K12}, 2: C_4-S_2-B^{K12}, 3: C_4-S_4-B^{K12}, 4: C_4-S_{10}-B^{K12}, 5: C_4-S_{14}-B^{K12}, MM: Molecular markers. B) MALDI-TOF mass spectrum of recombinant proteins displaying their measured molar mass.

~1h/4ºC, the pellets of C_4-S_2-B^{K12} and C_4-S_4-B^{K12} were resuspended in 50 mL of Milli-Q water, extensively desalted by dialysis against Milli-Q water (until conductometry measurements were constant and similar to Milli-Q water) and finally lyophilized.

The pellets for C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12} were resuspended in 140 mL and 250 mL respectively. NaCl was added to 50 mM and acetone to 40% (v/v). Supernatants were recovered after 35-50 min centrifugation at 11,500 rpm, and 4ºC, and acetone was added to a final concentration of 80% (v/v). A precipitate was formed and was recovered by centrifugation. After drying the pellets overnight in a hood, they were dissolved in Milli-Q water with frequent mixing and continuous heating at 65 ºC for several minutes. Finally, proteins were extensively dialyzed against Milli-Q water, lyophilized and stored in sealed tubes. Protein production yields (mg of protein per litre of recovered cell free broth) were 524 mg/L for C_4-S_2-B^{K12}, 325 mg/L for C_4-S_4-B^{K12}, 120 mg/L for C_4-S_{10}-B^{K12} and 94 mg/L for C_4-S_{14}-B^{K12}.

Table 3.2. Full Amino Acid Sequences of Recombinant Proteins from Gene Sequences

C_4-B^{K12}
MetRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSDLEGDFDVAVLPFSNSTNNGLLFINTTIASIAAKEEGVSLE KREAEAGPPGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQGNPGKNGQPGSPGSQGSPGNQGSP GQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQGNPGKNGQPGSP GSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQ GNPGKNGQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQP GEPGSNGPQGSQGNPGKNGQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGG*KKKKKKKKKKKKG*
C_4-S_2-B^{K12}
MetRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSDLEGDFDVAVLPFSNSTNNGLLFINTTIASIAAKEEGVSLE KREAEAGCGLGAGAPGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQGNPGKNGQPGSPGSQGSP GNQGSPGQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQGNPGKN GQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSN GPQGSQGNPGKNGQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQP GQNGQPGEPGSNGPQGSQGNPGKNGQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGA**GAGAGAGQGAGAGAGQGA***KKKK KKKKKKKKGA*
C_4-S_4-B^{K12}
MetRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSDLEGDFDVAVLPFSNSTNNGLLFINTTIASIAAKEEGVSLE KREAEAGCGLGAGAPGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQGNPGKNGQPGSPGSQGSP GNQGSPGQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQGNPGKN GQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSN GPQGSQGNPGKNGQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQP GQNGQPGEPGSNGPQGSQGNPGKNGQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGA**GAGAGAGQGAGAGAGQGAGAGA GQGAGAGAGQGA***KKKKKKKKKKKKGA*
C_4-S_{10}-B^{K12}
MetRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSDLEGDFDVAVLPFSNSTNNGLLFINTTIASIAAKEEGVSLE KREAEAGCGLGAGAPGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQGNPGKNGQPGSPGSQGSP GNQGSPGQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQGNPGKN GQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSN GPQGSQGNPGKNGQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQP GQNGQPGEPGSNGPQGSQGNPGKNGQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGA**GAGAGAGQGAGAGAGQGAGAGA GQGAGAGAGQGAGAGAGQGAGAGAGQGAGAGAGQGAGAGAGQGAGAGAGQGAGAGAGQGA***KKKKKKKKKKKKGA*
C_4-S_{14}-B^{K12}
MetRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSDLEGDFDVAVLPFSNSTNNGLLFINTTIASIAAKEEGVSLE KREAEAGCGLGAGAPGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQGNPGKNGQPGSPGSQGSP GNQGSPGQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQGNPGKN GQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSN GPQGSQGNPGKNGQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQP GQNGQPGEPGSNGPQGSQGNPGKNGQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGA**GAGAGAGQGAGAGAGQGAGAGA GQGAGAGAGQGAGAGAGQGAGAGAGQGAGAGAGQGAGAGAGQGAGAGAGQGAGAGAGQGAGAGAGQGAGAGAGQGAGAGAGQGAGAG AGQGA***KKKKKKKKKKKKGA*

Codes: α-factor prepro secretory signal, colloidal stability "C_4" block, **self-assembly "Sn" block** and *cationic "B^{K12}" block.*

Characterization. SDS-PAGE was carried out using the NuPAGE Novex system (Invitrogen, Carlsbad, CA) with 10% Bis-Tris gels, MES SDS as running buffer and SeeBlue Plus2 pre-stained molecular mass markers. 10 μL of protein solutions (10^{-1} mg/mL) were run. Gels were stained with Coomassie SimplyBlue SafeStain (Invitrogen). Matrix-assisted laser desorption ionization time-of-flight mass

spectrometry (MALDI-TOF MS) was carried out in an Ultraflex mass spectrometer (Bruker, Billerica, MA). Proteins samples were prepared by the dried droplet method. The matrix was made up of 5 mg/mL 2.5-dihydroxyacetophenone, 1.5 mg/mL diammonium hydrogen citrate, 25% (v/v) ethanol, and 1% (v/v) trifluoroacetic acid on a 600 μm AnchorChip target (Bruker). An external mass calibration was done based on Protein Calibration Standard II (Bruker). The full sequences of the proteins are listed in Table S2. SDS-PAGE and MALDI-TOF results are displayed in Figure 3.7. SDS-PAGE shows well defined and thick bands at the top of the gel. Proteins containing the C-block and, to a lesser extent, variants of the S block containing ~ 12 mol-% negatively charged amino acids do not run as fast as equally heavy marker proteins, because (1) they contain a large amount of light amino acids (and thus a longer chain corresponds to a specific Mw) and (2) they have a high content of hydrophilic amino acids that bind to SDS poorly.[16] MALDI-TOF revealed a good agreement with the expected theoretical molecular weight of the proteins (Protein C_4-B^{K12} MW_{Theo} = 38,407.8 Da and MW_{Exp} = 38,608 Da), C_4-S_2-B^{K12} MW_{Theo} = 40192.6 Da and MW_{Exp} = 40,315 Da), C_4-S_4-B^{K12} (MW_{Theo} = 41331.8 Da and MW_{Exp} = 41,458 Da), C_4-S_{10}-B^{K12} (MW_{Theo} = 44,749.2 Da and MW_{Exp} = 45,160 Da) and C_4-S_{14}-B^{K12} (MW_{Theo} = 47027.5 Da). The small increment in the size of the proteins could be due to extra EA repeats left after removing the α-mating sequence required to export the expressed protein into the medium.[16]

3.6.2. Fluorescence Correlation Spectroscopy. The hydrodynamic diameter of the polymer brush around DNA due to complexation with the C_4-B^{K12} polymer was determined using Fluorescence Correlation Spectroscopy. Complexes were formed between a short linear fluorescently-labelled dsDNA (348 bp) and C_4-B^{K12}. The DNA fragment has a contour length of ~118 nm, which is about twice the persistence length (~50 nm) of dsDNA, such that these fragments diffuse as rigid rods in solution. Any decrease in the translational diffusion constant as a a consequence of complexation with the C_4-B^{K12} protein can then be attributed to an increase of the effective hydrodynamic diameter of the complexes, due to the presence of the polymer brush.

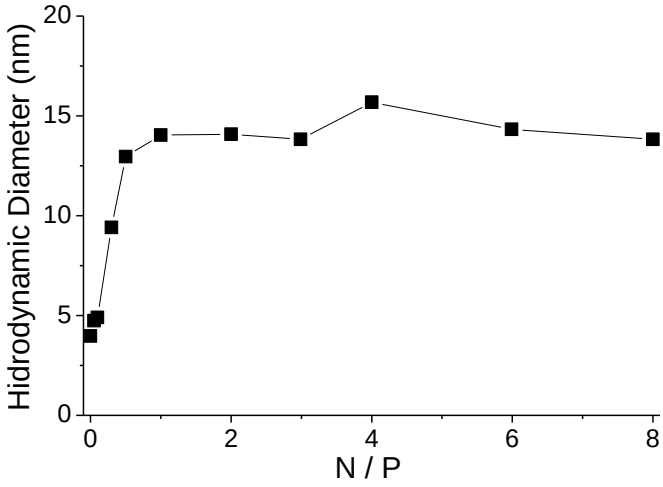


Figure 3.8. Hydrodynamic diameter *d* of protein brush around dsDNA when coated with C_4-B^{K12} diblock polymer.

Production of the fluorescently labeled dsDNA. Fluorescent dsDNA was produced by PCR using a Fw primer labelled with the photo stable fluorescent label Alexa-488 (Eurogentec, Belgium). The PCR produced dsDNA fragment was purified through agarose gel electrophoresis. The band with the fluorescent dsDNA was cut out the gel and recovered with a GeneJET Gel Extraction Kit (Thermo Scientific). The concentration of the DNA was determined using the nanodrop at 260nm. The fluorescence due to Alexa-488 dye at this wavelength was negligible (<4%).

Fluorescence Correlation Spectroscopy. A confocal microscope type Confocor2/LSM510 microscope (Carl Zeiss, Jena, Germany) was used. For calibration purposes a 2 µg/L Rhodamine 110 solution (diffusion coefficient 4.40E-10 m^2/s) in 10 mM sodium phosphate buffer pH 7.4 was used with laser power of 10%. This concentration is low enough to observe predominantly single molecules passing through the confocal volume. 200µL of sample were prepared with a labelled-DNA 348bp concentration of ~3ng/µL in 10mM of sodium phosphate buffer pH7.4 and increasing concentrations of the protein. The samples were

incubated for about 40 minutes before measuring. The samples were deposited in 400µL cuvettes (Lab-Tek Chambered #1.0 Borosilicate Cover Glass System, 155411, 8 chambers). The samples were excited with an argon laser at 488nm (5% laser power). A dichroic mirror of 488/633 and a pinhole of 70µm were used. Each sample is measured for 30 seconds and the measurement is repeated 20 times, without time intervals in between the repeats. The fitting of the autocorrelation functions was done using a FFS data processor (Scientific Software Technologies Centre, 64bit, product version 2.3). Fitted translational diffusion constants D_t were related to the effective diameter d of the rod-like complexes as described before:[23]

$$D_t = \frac{Ak_BT}{3\pi\eta L} \qquad (3.1)$$

$$A - \ln\left(\frac{L}{d}\right) + 0.312 + \frac{0.565}{(L/d)} - \frac{0.1}{(L/d)^2} \qquad (3.2)$$

where L is the length of the template dsDNA, k_B is the Boltzmann constant, T is absolute temperature, and η is the viscosity of the medium.

From the experimental diffusion constant for dsDNA alone a bare diameter d≈3.0 nm was found, close to the accepted value (2.4 nm) for dsDNA in solution. For increasing coverage of the DNA with the C_4-B^{K12} protein, the diameter of the complexes increases to a saturation value of d≈15 nm (Figure 3.8).

3.6.3. Atomic Force Microscopy. Samples were analyzed using a Digital Instruments NanoScope V equipped with a silicon nitride probe (Bruker, MA, USA) with a spring constant of 0.4-0.35 N/m in ScanAsyst™ imaging mode. Images were recorded between 0.488-0.977 Hz and 384-1024 samples/line. Image processing, height and diameter measurement was done with NanoScope Analysis 1.20 software. A first order flattening was used for all the images. Contour length measurements were performed using ImageJ software. Samples were prepared by mixing freshly prepared protein stock solutions in water (0.5 – 2 mg/mL) with DNA stock solution with 10x phosphate buffer (100 mM, pH 7.4, prepared from

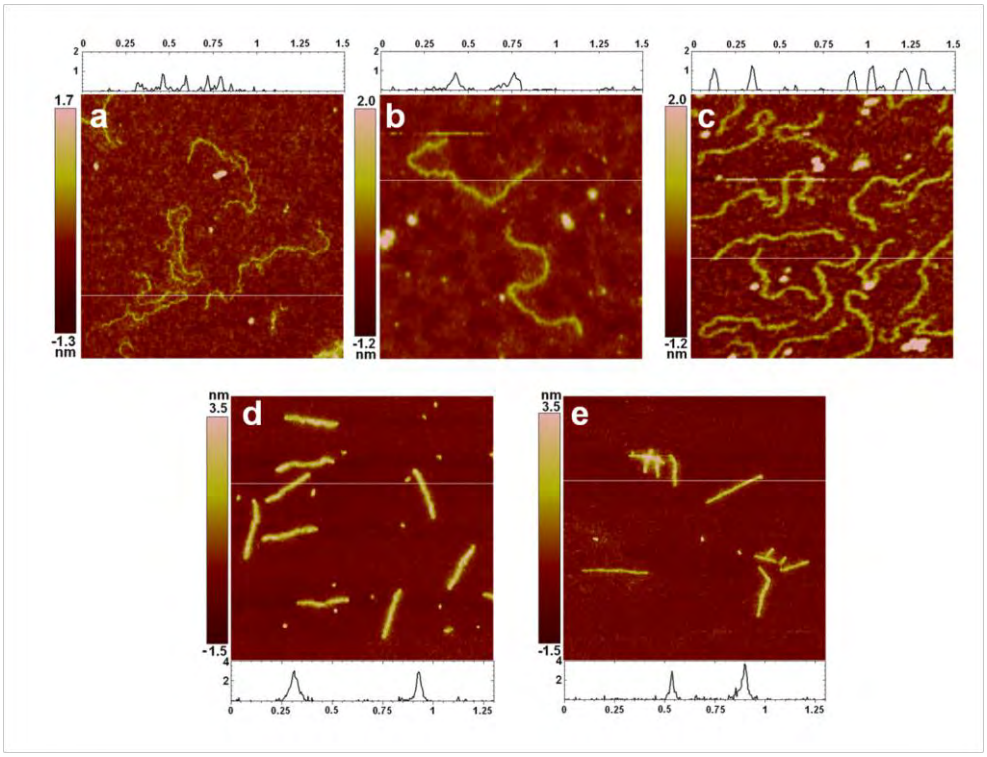


Figure 3.9. Additional AFM images of complexes of dsDNA with C_4-S_n-B^{K12} polymers. a) C_4-B^{K12} + linear dsDNA 3.7 kbp, N/P = 1, incubation time 90 min ($L_{complex}$ = 1245 nm, L_{bare} = 1258 nm); b) C_4-S_2-B^{K12} + linear dsDNA 3.7 kbp, N/P = 2, incubation time 90 min ($L_{complex}$ = 1305 nm, L_{bare} = 1258 nm, c) C_4-S_4-B^{K12} + linear dsDNA 2.7 kbp, N/P = 8, incubation time 90 min ($L_{complex}$ = 885 nm, L_{bare} = 913 nm), d) C_4-S_{10}-B^{K12} + linear dsDNA 2.5 kbp, N/P = 7, incubation time 24h ($L_{complex}$ = 315, nm L_{bare} = 850 nm) and e) C_4-S_{14}-B^{K12} + linear dsDNA 2.5 kbp, N/P = 7, incubation time 24h. ($L_{complex}$ = 304 nm, L_{bare} = 850 nm). Complexes were deposited on ultraflat silica wafers and dried before imaging.

$NaH_2PO_4 \cdot H_2O$ and $Na_2HPO_4 \cdot 2H_2O$ of analytical grade purchased from Merk), 10x digested and dephosphorylated purchased from Thermo Scientific). The concentration of DNA was determined using a Nanodrop ND-1000 spectrophotometer (Thermo Scientific). Protein-DNA samples for AFM imaging were prepared usually in the following way: 3-5 µL of sample ([DNA] = 1 µg/mL) were deposited onto a clean silicon surface. After 2 minutes the surface was rinsed with 1 mL of Milli-Q water to remove salts and non-absorbed particles, followed by soaking up of excess water using a tissue and slow drying under a N_2 stream.

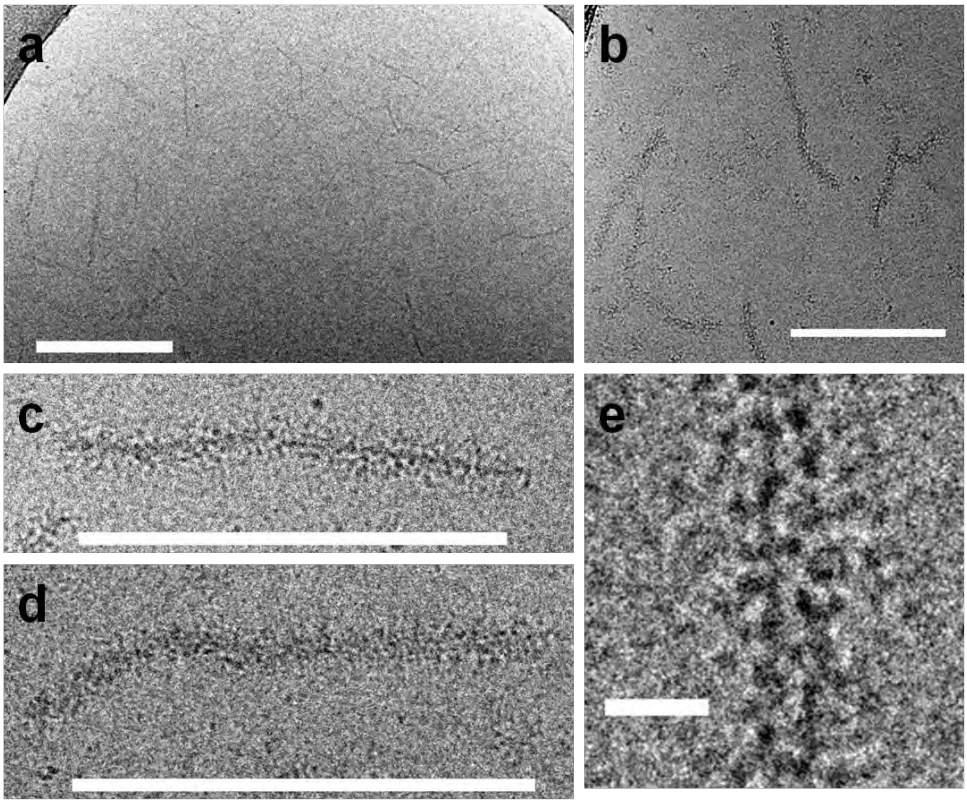


Figure 3.10. Cryo-TEM images of C_4-S_{10}-B^{K12} complexed with 2.5 kbp linear dsDNA**.** DNA concentration 30 µg/ml, charge ratio N/P = 7. a) Overview of sample on TEM grid. Scale bar 500 nm. b) scale bar 300 nm. c) and d) Single VLPs, scale bar 300 nm. e) Scale bar 25 nm.

Additional images of complexes of linear dsDNA and the protein-based polymers are shown in Figure 3.9. Heights of the complexes increase with the number of self-assembly units S, from ~1 nm for C_4-B^{K12}, to ~1.3 nm for C_4-S_2-B^{K12} and C_4-S_4-B^{K12} and finally to ~3 nm for C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12}.Dithiothreitol (1 mM, analytical grade, Sigma) and MQ-water to the desired volume and charge N/P ratio (molar ratio between positively charged NH_2 groups (N) from binding block "B^{K12}" to negatively charged PO_3 groups of the DNA template (P)). Linear dsDNA stock were either NoLimits™ DNA fragments (0.3, 1.2, 2.5 and 8.0 kbp, 0.5 µg/mL, Thermo Scientific, Waltham, MA) or 3.7 kbp linearized plasmid DNA (prepared by

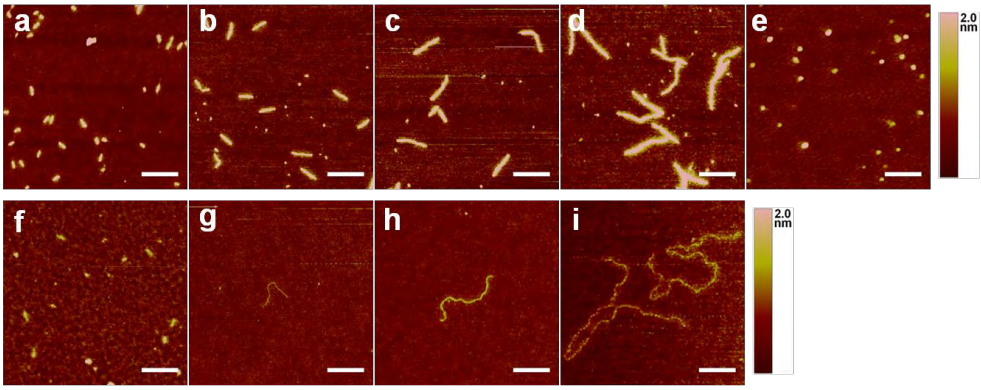


Figure 3.11. Dependence of Virus-like particle size on length of DNA template. Top row, AFM images of complexes of C_4-S_{10}-B^{K12} and linear dsDNA A) 0.3 kbp, B) 1.2 kbp, C) 2.5 kbp, D) 8.0 kbp, E) C-S_{10}-B protein only control. Bottom row, AFM images of complexes of C_4-B^{K12} diblock with linear dsDNA. F) 0.3 kbp, G) 1.2 kbp, H) 2.5 kbp, I) 8.0 kbp. Scale bar is 300 nm.

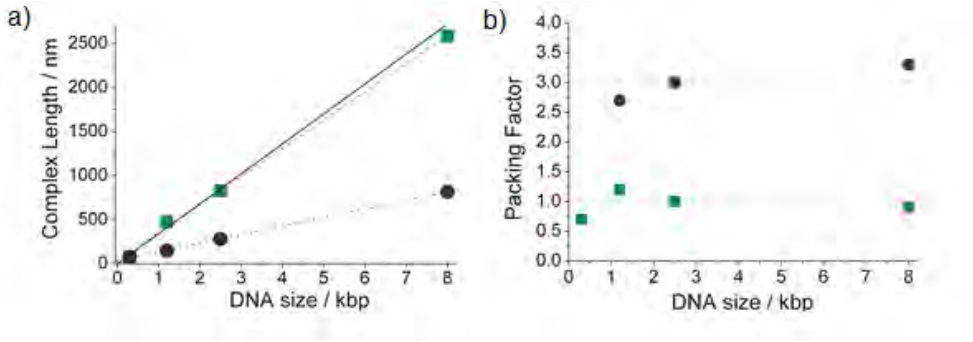


Figure 3.12. Average contour lengths of DNA-polymer complexes versus length of DNA template in bp. a) Green squares: C_4-B^{K12} diblock polymer, Blue circles: C_4-S_{10}-B^{K12} triblock polymer. Dotted lines are linear fit used to determine the average packing factor, black solid line is the contour length of the template DNA. b) Packing factor of complexes (contour length of complex/contour length of template DNA) versus length of DNA template. Green squares: C-B diblock polymer, blue circles: C_4-S_{10}-B^{K12} triblock polymer. Dotted lines indicate the average packing factor.

digesting pDNA from bacterial source, miniprep kit, Thermo Scientific and purified with a GeneJET Gel Extraction and DNA Cleanup Micro Kit, Thermo Scientific) or 2.7 kbp linearized plasmid DNA (pPUC/BamHI, digested and dephosphorylated purchased from Thermo Scientific). The concentration of DNA was determined using a Nanodrop ND-1000 spectrophotometer (Thermo

Scientific). Preparation of protein-DNA samples for AFM imaging were prepared usually in the following way: 3-5 μL of sample ([DNA] = 1 μg/mL) were deposited onto a clean silicon surface. After 2 minutes the surface was rinsed with 1 mL of Milli-Q water to remove salts and non-absorbed particles, followed by soaking up of excess water using a tissue and slow drying under a N_2 stream. Additional images of complexes of linear dsDNA and the protein-based polymers are shown in Fig. S3. Heights of the complexes increase with the number of self-assembly units S, from ~1 nm for C-B, to ~1.3 nm for C-S_2-B and C-S_4-B and finally to ~3 nm for C-S_{10}-B and C-S_{14}-B.

3.6.4. Cryo-TEM Images of Polymer + DNA Complexes. Samples ([DNA] = 30 μg/mL) were prepared using a vitrification robot (FEI Vitrobot Mark III) using grids R2/2 (Quantifoil Micro Tools GmbH, Jena, Germany) previously surface plasma treated (Cressington 208 carbon coater). The cryo-TEM characterizations were performed on a CryoTitan (FEI) equipped with a field emission gun operating at 300 kV and a postcolumn Gatan energy filter. Images were recorded using a 2k×2k Gatan CCD camera.

3.6.5. Dependence of Virus-like Particle Size on Length of DNA Template. Linear dsDNA [1 μg/mL] of different lengths (0.3, 1.2, 2.5 and 8.0 kbp) was incubated for 1h at N/P = 10 with C_4-S_{10}-B^{K12} and C_4-B^{K12} proteins and complexes were imaged using AFM (Figure 3.11a-d and Figure 3.11f-i). At conditions used for the AFM experiment, the C_4-S_{10}-B^{K12} protein-only control shows short micelles with a length of around 40 nm (Figure 3.11e).

Lengths of the complexes were measured from the AFM images. For the longest DNA (8.0 kbp DNA) nucleation of VLP formation in most cases occurs at multiple sites such that the final complexes have a branched structure. For this case, the lengths of all arms were summed up. Average lengths of complexes are plotted against the theoretical contour length of the bare DNA template in Figure 3.12a. The corresponding packing factors (length of complexes/contour length of DNA template) are shown in Figure 3.12b, as a function of the contour length of the DNA template.

3.6.6. Static Light Scattering Polymer + DNA Complexes. Static light scattering experiments were performed using an ALV-125 goniometer, combined with a 300 mW Cobolt Samba-300 DPSS laser operating at a wavelength of λ = 532 nm, an ALV optical fiber with a diameter of 50 mm, and an ALV/SO Single Photon Detector. Temperature was maintained at 25 ± 0.1 °C using a Haake F8-C35 thermostatic bath. Measurements were taken at multiple scattering angles θ ranging from 20° to 130°, corresponding to scattering vectors q given by:

$$q = \frac{4\pi n_s}{\lambda}\sin(\theta/2) \qquad (3.3)$$

where n_s = 1.333 is the refractive index of the aqueous solvent. Absolute scattering intensities (Rayleigh ratio's) R_θ were calculated from the relative scattered intensities of the sample (I_{sample}), solvent ($I_{solvent}$) and toluene reference($I_{toluene}$) according to:

$$R_\theta = \frac{I_{sample}(\theta) - I_{solvent}(\theta)}{I_{toluene}(\theta)} \frac{n_s^2}{n_{tol}^2} R_{toluene} \qquad (3.4)$$

The absolute scattering of the reference toluene is $R_{toluene}$ = 2.10 x 10^{-3} m^{-1} and the toluene refractive index is $n_{toluene}$ = 1.496. At sufficiently low concentrations, particle interactions can be neglected. Then, in the limit $qR_g << 1$, the absolute scattering intensity for a solution with objects of molar mass M and gyration radius R_g at a weight concentration C is given by

$$R_\theta = K_R CM(1 - \tfrac{1}{3}(qR_g)^2 + \ldots) \qquad (3.5)$$

where the Rayleigh constant K_R is given by

$$K_R = \frac{4\pi^2 n_s^2}{\lambda N_{Av}}\left(\frac{dn}{dc}\right)^2 \qquad (3.6)$$

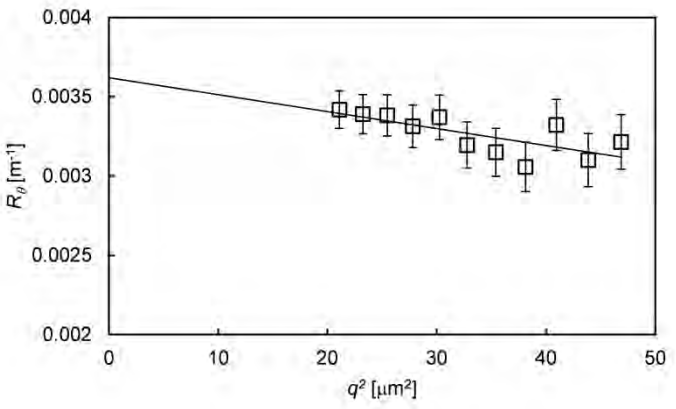


Figure 3.11. Static light scattering of C_4-S_{10}-B^{K12} complexed with 2.5kb linear dsDNA, c_{DNA} = 0.535nM, N/P = 3. Absolute intensity R_θ versus square of the magnitude of the wavevector q. Symbols: experimental data, solid line, fit to Eq. (S8). The fit corresponds to a number of proteins per VLP of N = 400±10, and a gyration radius R_g = 95 ±7nm.

where *dn/dc* is the refractive index increment of the scattering objects. Strictly speaking, in order to determine solution molar masses and radii of gyration an extrapolation to zero concentration is required, but for self-assembling objects this is problematic since dilution may affect the composition and hence the molar mass and gyration radius of the complexes. Here we use a single low concentration to estimate the molar mass of C_4-S_{10}-B^{K12} + DNA complexes at one specific solution condition. The linear DNA that was used had a length of 2.5kb. Composition of the samples was C_{DNA} = 0.83 μg/ml or c_{DNA} = 0.535nM at a protein/DNA ratio (in terms of binding block charges over DNA phosphate charges) of N/P = 3. The absolute scattering as a function of scattering wavevector is shown in Figure 3.11. It was verified that the scattering of a control sample with the same concentration of protein but in the absence of DNA gave a very low scattering intensity, comparable to the scattering of the solvent, and negligible with respect to the scattering of the complexes. Therefore, the scattering is dominated by that of the complexes. The number concentration of complexes is equal to the number concentration of DNA templates, i.e. c_{DNA}. On the other hand, the mass of the complex is dominated by the mass of the proteins in the complex, i.e.

$$M_{complex} \approx NM_0 \qquad (3.7)$$

where M_0 = 45 kg/mol is the molar mass of the C_4-S_{10}-B^{K12} protein-based polymer. Therefore, from the light scattering data, we extract the gyration radius R_g and number of proteins N per complex, using

$$R_\theta \approx K_R c_{DNA}^2 M_0^2 N^2 (1 - \tfrac{1}{3}(qR_g)^2 + \ldots) \qquad (3.8)$$

Since the mass of the complexes is dominated by that of the proteins in the complex, the relevant value for the refractive index increment is that of C_4-S_{10}-B^{K12}. Typical values for proteins are in the range dn/dc = 0.16-0.18 $cm^3\ g^{-1}$ and here we use a value of dn/dc = 0.17 $cm^3\ g^{-1}$. From the solid curve shown in Figure 3.11 we deduce the following estimates: R_g = 95 ± 7nm and N = 400 ± 10.

Assuming full charge stoichiometry of the complexes, the predicted number of proteins per complex is N = 417 (one protein per 6bp of dsDNA). Hence the experimental estimate shows that the complex stoichiometry is close to that expected for complete charge neutralization.

Furthermore, the gyration radius of a rod with length L and diameter d, is given by

$$R_g^2 = \frac{d^2}{8} + \frac{L^2}{12} \qquad (3.9)$$

From AFM the average length of the VLPs is 1/3 of the contour length, or L = 283 nm for 2.5kb DNA. Assuming a diameter d = 15nm as determined for complexes of DNA coated with the C-B diblock polymer using FCS (Figure 3.8), the expected radius of gyration for the VLPs would be $R_g \approx$ 82 nm, which is close to the value estimated from the light scattering experiment.

3.6.7. Virus-like Particles Formed with ssDNA. Linear ssDNA derived from M13mp18 virus (7249 nt, Thermo scientific) was complexed with C_4-S_{10}-B^{K12}

([DNA] = 1μg/mL, incubated time 26h, N/P = 11) and imaged using AFM (Figure 3.12). As for the case of 8kb dsDNA, nucleation of VLPs formation in many cases occurred at a few sites per molecule such that the final structures are typically branched. For each complex, the sum of the length of all branches was measured. The average total encapsulated length of was found to be 413 ± 72 nm. This corresponds to a packing factor of 6.1, since the contour length of M13mp18 ssDNA is 2516 nm.

3.6.8. Self-assembly of Protein-based Polymers in the Absence of DNA Template. Dynamic Light scattering on dilute solutions of both C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12} (in the absence of nucleic acid template) shows coexistence of free polymers (hydrodynamic radius approx 8 nm) and small aggregates (hydrodynamic radius 40-80nm). Above a critical concentration, both the fraction of aggregates and their size increase sharply. For C_4-S_{10}-B^{K12} the critical concentration is approx 80 μM (or 3.6 mg/ml), whereas for C_4-S_{14}-B^{K12} it is around 12 μM (or 0.56 mg/ml). AFM and cryo-TEM images of C_4-S_{10}-B^{K12} are shown in Figure 3.13b. (AFM, 5 mg/ml, incubated for 2 days) and Figure 3.13c (cryo-TEM, 2.5 mg/ml, incubated for 7 days). The images clearly show that at high concentrations, the C_4-S_{10}-B^{K12} polymers also tend to form rod-like assemblies in the absence of DNA template. AFM and cryo-TEM images were used to construct histograms of the length distributions of these rod- like assemblies (N = 100 samples for cryo-EM, N = 201 samples for AFM, image analysis was done using *ImageJ*). As shown in Figure 3.13a (upper panel), the rod-like assemblies have a typical length of around 100 nm.

VLPs were typically studied at an excess of C_4-S_n-B^{K12} polymers over DNA template (in terms of the charge ratio N/P) such that it may be expected that VLPs should coexist with protein-only assemblies. Conditions for AFM were such that the excess polymer concentration was always much lower than the critical concentration for C_4-S_{10}-B^{K12} self-assembly referred to above. For this case, AFM imaging shows that VLPs coexist with occasional smaller protein-only structures.

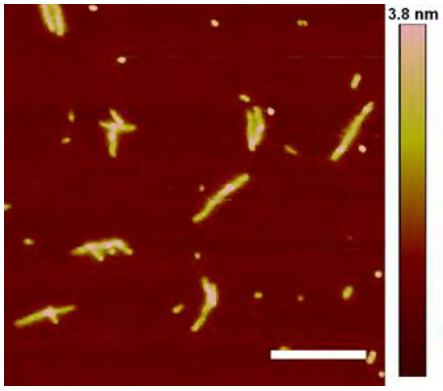


Figure 3.12. AFM image of VLPs formed by complexation of ssDNA from M13 virus with C_4-S_{10}-B^{K12} polymer. Scale bar 500nm. The average total encapsulated length for the M13 ssDNA was found to be 413 ± 72 nm. This corresponds to a packing factor of 6.1, since the contour length of M13 ssDNA is 2,516 nm.

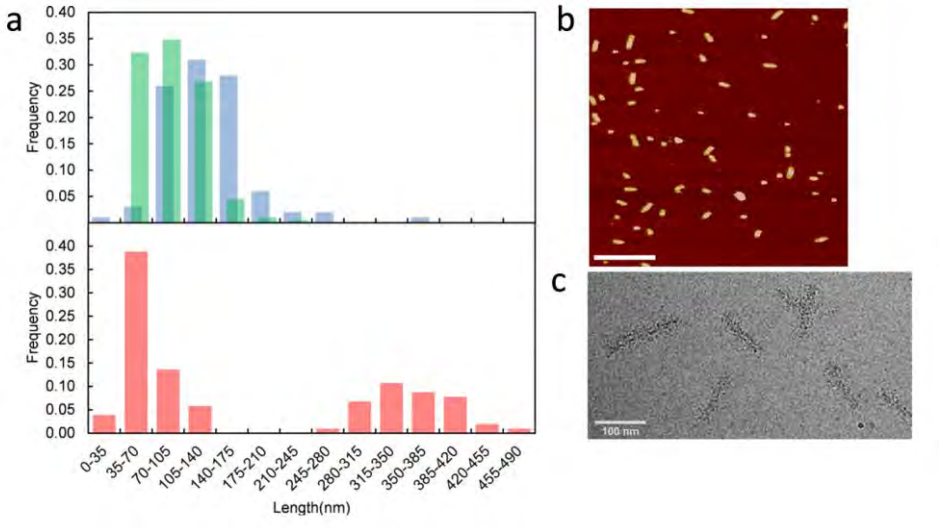


Figure 3.13. Self-assembly of C_4-S_{10}-B^{K12} in the presence and absence of DNA template. a) Top panel: length distributions of rod-like protein-only assemblies. Blue: from cryo-TEM images for sample at concentration of 2.5 mg/mL, incubation time 7 days. Green: from AFM images for sample at concentration of 5 mg/mL, incubation time 2 days. Bottom panel: length distributions from cryo-TEM for sample C_4-S_{10}-B^{K12} + linear dsDNA 2.5 kbp. DNA concentration 30 μg/ml, charge ratio N/P = 7 (corresponding to a C_4-S_{10}-B^{K12} concentration of 2.5 mg/ml). Representative cryo-TEM images of this sample are shown in Figure 3.2f and 3.2g and in Figure 3.10b) Representative AFM image of protein-only assemblies scale bar 600 nm c) representative cryo-TEM image of protein-only assemblies, scale bar 100 nm.

On the other hand, cryo-TEM imaging of VLPs was done at much higher concentrations, and as expected, shows a clear coexistence of VLPs with a significant fraction of shorter protein-only rods. The two species are easily distinguished by their length, as is also evident from the bi-disperse length-distribution obtained from cryo-TEM images of complexes of 2.5kb linear dsDNA with the C_4-S_{10}-B^{K12} polymer shown in Figure 3.13a (lower panel).

3.6.9. Protection Against Enzymatic DNA Degradation by Protein-based Polymer. Protein-DNA complexes were incubated for 1h at room temperature with pDNA (2,640 bp) [DNA] = 20.5 μg/mL (N/P = 8.5). After, approximately 1μL of the enzyme DNAse I (RNAse free , Thermo Scientific) 0.055 U was added to 35.5 μL complexes dissolved in reaction DNAse I buffer (100 mM Tris-HCl, pH 7.5, 25 mM $MgCl_2$, 1 mM $CaCl_2$) for a final [DNA] = 17.9 μg/mL and immediately incubated at 25ºC using a thermo block. Aliquots of 3.5 μL were taken at different times and mixed with 3.5 μL of EDTA 50 mM. After adding loading buffer (6x) they were electrophoresed in agarose gel 1% at 100V for 45 min. DNA bands were visualized using ethidium bromide. The agarose gels showing the bands are shown in figure 3.3. No intact naked pDNA is present anymore after 1 minute of incubation, whereas when complexed with C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12} polymers, a significant fraction of DNA was still intact after 60 minutes of incubation.

3.6.10. Transfection by DNA-polymer Complexes.

Cell culture conditions. HeLa cells were grown as monolayers in RPMI-1640 medium, supplemented with 10% fetal calf serum (fcs) and 200 mM L-glutamine, at 37°C, in a humidified atmosphere containing 5% CO_2. Cells were passaged after 2 days or when at 80-90% sub-confluency.

Transfection. Proteins C_4-B^{K12}, C_4-S_2-B^{K12}, C_4-S_4-B^{K12} and C_4-S_{10}-B^{K12} were incubated with plasmid DNA encoding for yellow fluorescent protein (YFP-pDNA), at N/P = 6, for 1 hour to form the relevant complexes. Complexes (60 ng pDNA) were then incubated for 2 hours, in a 24-well plate, with HeLa cells ($4x10^5$ cells/well/ml) at a final 10 μM concentration, in RPMI-1640 medium supplemented with 10% fcs. Equivalent concentration of free YFP-pDNA and

complexes formed with PEI (branched, 25 kDa, N/P = 6) and Lipofectamine 2000, containing the same amount of YFP-pDNA as protein complexes, were used as controls. After incubation, cells were washed twice with HBS and fresh culture medium was added.

Expression of YFP was assessed after 2 days by flow cytometry, using a Becton Dickinson FACScalibur flow cytometer and fluorescein detector. Data were analyzed using Weasel (WEHI, version 30114win). Results are shown in Figure 3.4.

3.6.11. Theory of Nucleated Linear Self-assembly of Virus-like Particles. The assembly model that we invoke to rationalize our experimental data for the kinetics of VLP growth is identical to the one that we recently proposed for the co-assembly of the coat proteins (or coat protein assemblies) and the ssRNA of tobacco mosaic virus.[20] It hinges on a pathway characterized by the sequential addition of protein subunits starting from a so-called origin of assembly sequence (OAS) on the single-stranded RNA of the virus.[2] The first subunit, which binds to the first of a linear array of q binding sites, pays a free energy penalty upon adsorption due to conformational switching, $h>0$, whilst all adsorbed subunits gain a binding free energy, $g<0$, driven by coulomb interactions, hydrophobic interactions and hydrogen bonding between different moieties on the subunits and the ssRNA. The conformational switching of the second, third, etc. protein unit is presumed to be catalysed by the first one by way of allosteric effect, so no free energy penalty is required for those, this can be considered as absorbed in the net binding free energy, g. In addition, the bound protein subunits gain a free energy, $\varepsilon < 0$, per bond due to favorable protein-protein interactions. Here, all free energies are in units of thermal energy, k_BT.

The combination of sequential adsorption and protein-protein interactions makes this model much more cooperative than the well-known McGhee-von Hippel model that allows random adsorption of the proteins units on the one-dimensional template.[24,25] Based on our model, which is a variant of Kittel's molecular "zipper" model,[26] we set up our kinetic (reaction) equations with on and off rates that are related to each other due the condition that under conditions of

steady state the equilibrium (Boltzmann) distribution of proteins on the templates must be obtained. We, in principle, allow the on rate of the very first protein bound to be different from the subsequent ones. For simplicity we presume in the present paper, there to be only one rate constant, K_+ so the first and subsequent on rates are set to be equal. For a discussion why this is a sensible approximation, we refer to our previous work.[20] There we also discuss the importance of knowing "zero time", i.e. the time it takes before the first data are recorded following the quench or mixing of components. As it is not known exactly, this introduces an additional parameter: the time the first data are taken, τ_0.

As already alluded to, the quasi one-dimensional model presumes that there are q binding sites for the protein subunits. Two important model parameters are the mass action variable $S = \varphi_p K$ where φ_p is the mole fraction of proteins in the solution and $K = \exp(-g - \varepsilon)$ is a dimensionless association constant, and the cooperativity parameter $\sigma = \exp(\varepsilon - h)$. Another one is the stoichiometric ratio, $\gamma = q\rho_T / \varphi_p$, where ρ_T is the mole fraction template molecules. For TMV the template is ssRNA, whilst in the present case it is dsDNA. The stoichiometric ratio enforces mass conservation. If $\gamma << 1$ there is excess protein in the solution, which is numerically the most straightforward to deal with, and is the limit investigated in detail in ref. 20. We shall invoke this limit here too, even though not strictly valid in our experiments as γ = 0.324 (for N/P = 3). However, for small times, the limit of vanishing stoichiometry always applies as little of the background solution is then depleted by adsorption onto the template. For TMV we found the model to accurately describe the time evolution of the length of ssRNA covered by coat proteins.[20] The model also describes the assembly kinetics data on our artificial virus reasonably well. The full comparison (for all incubation times that were are analyzed) is shown in Figure 3.16. The figure in the main text (Figure 3.6) shows these same results for a selected number of incubation times.

To obtain the fit corresponding to Figure 3.16, we have calculated the average coverage per dsDNA molecule from the experimental data at the first and last time step of the time measurement series. The template length q is set to 417, because on

each dsDNA stretch of 5,000 negative charges (2,500 bp) 417 proteins with 12 positive charges each can adsorb to yield a neutral assembly. From these parameters we can calculate the value for the mass action variable for various co-operativities σ and numerically compute the assembly as a function of time using the starting and the end coverage as input. In the final step, we binned the idealized numerical data in the same way as the experimental data, and compared the two. Note that the theory does not account for the formation of micellar structures at low lengths due to excess addition of protein; we therefore neglected this fraction for the fit. Furthermore, any aggregate that exceeded 245 nm was presumed to be a fully covered DNA strand and thus all data above 210 was binned together. As mentioned in the main text, the fit leads to estimated parameter values of $h-\varepsilon \approx 5.3k_BT$ for the effective penalty for nucleation (or a cooperativity parameter $\sigma = 0.005$) , and $g+\varepsilon \approx -18k_BT$ for the effective binding free energy (or a dimensionless association constant $K = 9.7\cdot10^7$).

That agreement between theory and experiment is not perfect should not be a surprise. Indeed, the assembly of our artificial virus is slightly different from that of TMV. First, there is no origin-of-assembly domain encoded on the DNA, although for the C_4-S_{10}-B^{K12} polymer and 2.5kb linear dsDNA as a template nucleation of VLP growth almost always starts at one end of the dsDNA, mimicking an origin of assembly. Second, prior to the sequential binding onto the DNA, characterized in the AFM images by a shortening and significant widening of the chain, random adsorption of a thin layer of proteins seems to take place. The latter plausibly is of the Langmuir type and does neither not lead to any conformational changes of the proteins, nor to a shortening of the chain. In order to gauge the importance of this second effect, we have performed a preliminary theoretical analysis of the competition between the zipper- and Langmuir-type adsorption in a simple statistical mechanical model (unpublished), and found the following. (i) Even if the binding strength of the Langmuir-type is smaller than that in the molecular zipper-type adsorption, the former starts at smaller values of the mass action variable $S = \varphi_p K$, in essence because the latter is much more co-operative. (ii) Still, when zipper adsorption takes place it displaces the proteins that have been in the

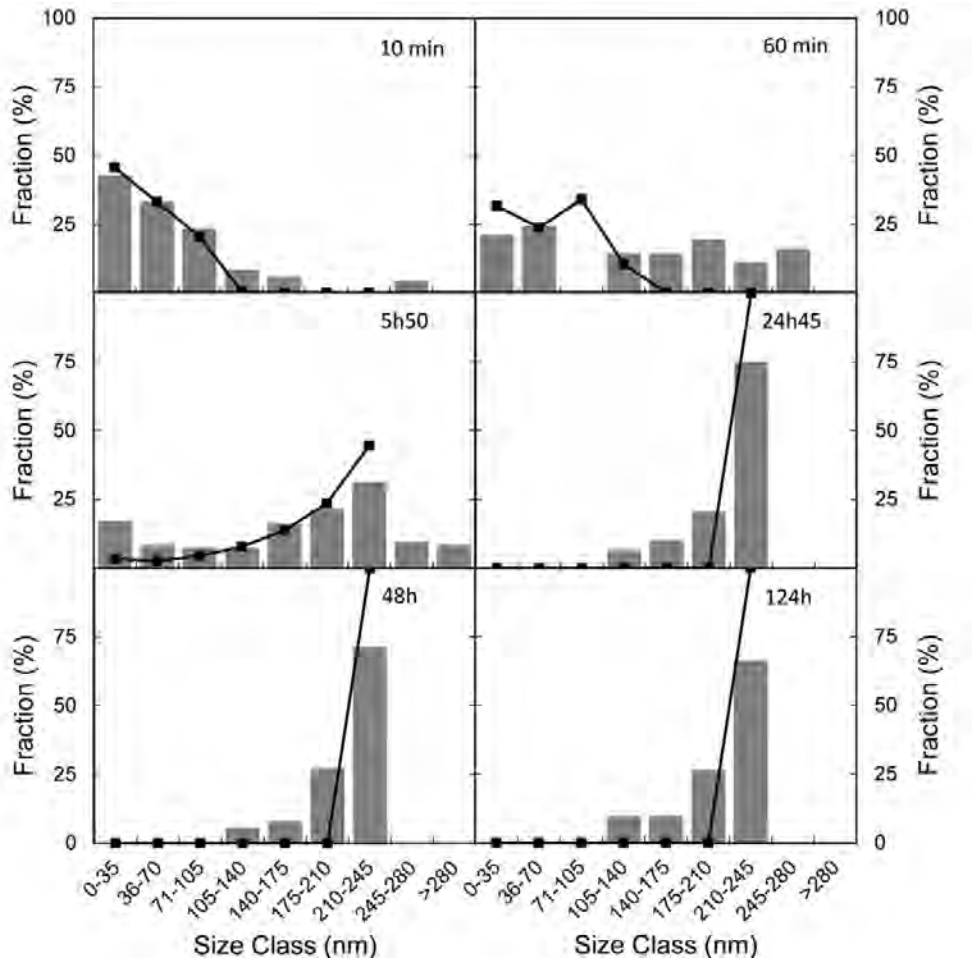


Figure 3.16. Kinetics of VLP formation. Linear 2.5kb dsDNA with C_4-S_{10}-B^{K12} polymers at a charge ratio N/P = 3, [DNA] = 1 μg/mL. Incubation times are indicated in the panels. Experimental data (grey bars) are obtained by analyzing AFM images of complexes incubated for the indicated times. Fits to the nucleated linear self-assembly model[20] (black squares connected by black solid lines) have been obtained as described in the text.

Langmuir adsorption mode. (iii) All in all, the Langmuir adsorption does not shift the onset of the zipper adsorption significantly, if the binding free energy in the zipper mode is more than a few times the thermal energy larger than that in the Langmuir mode. Considering the large cooperativity that we observe of the former, this seems more than likely to be the case. Any influence of displacing the proteins

adsorbed in the Langmuir mode, we absorb in the aforementioned zero time parameter, .

3.6.12. Electrophoresis Mobility Shift Assay. DNA-protein samples ([DNA] = 15 ng/μL) incubated for 1.5h were electrophoresed in a 1% agarose gel (95 min / 60V) using 1x TAE buffer (pH 8). Bands were visualized with ethidium bromide.

3.6.13. References

31. Werten, M.W.T, Moers, A.P.H.A, Vong, T., Zuilhof, H., van Hest, J.C.M., de Wolf, F.A. Biosynthesis of an amphiphilic silk-like polymer. *Biomacromolecules* **9**, 1705-1711 (2008).

32. DeRouchey, J., Walker, G.F., Wagner, E., Rädler, J.O. Decorated Rods: A "Bottom-Up" Self-Assembly of Monomolecular DNA Complexes. *J. Phys. Chem. B* **110**, 4548-4554 (2006).

33. McGhee, J.D., and von Hippel, P.H. Theoretical Aspects of DNA-Protein Interactions: Co-operative and Non-co-operative Binding of Large Ligands to a One-dimensional Homogeneous Lattice. *J. Mol. Biol.* **86,** 469-489 (1974).

34. Janssen, P. G. A., Jabbari-Farouji, S., Surin, M., Vila, X., Gielen, J. C., de Greef, T. F. A., Vos, M. R. J., Bomans, P. H. H., Sommerdijk, N.A. J. M., Christianen, P. C. M., Leclere, P., Lazzaroni, R., van der Schoot, P., Meijer, E. W., and Schenning, A. P. H. J. Insights into Templated Supramolecular Polymerization: Binding of Naphthalene Derivatives to ssDNA Templates of Different Lengths. *J. Am. Chem. Soc.* **131**, 1222–1231 (2009).

35. Kittel, C. Phase Transition of a Molecular Zipper, *Am. J. Phys.* **37** , 917–920 (1969).

Chapter 4

Amplified stretch of bottlebrush-coated DNA in nanofluidic channels[1]

The effect of a cationic-neutral diblock polypeptide on the conformation of single DNA molecules confined in rectangular nanochannels is investigated with fluorescence microscopy. An enhanced stretch along the channel is observed with increased binding of the cationic block of the polypeptide to DNA. A maximum stretch of 85% of the contour length can be achieved inside a channel with across-sectional diameter of 200 nm and at a twofold excess of polypeptide with respect to DNA charge. With site-specific fluorescence labelling, it is demonstrated that this maximum stretch is sufficient to map large-scale genomic organization. Monte Carlo computer simulation shows that the amplification of the stretch inside the nanochannels is due to an increase in bending rigidity and thickness of bottlebrush coated DNA. The persistence lengths and widths deduced from the nanochannel data agree with what has been estimated from the analysis of atomic force microscopy images of complexes.

---

1 This chapter is based on: Ce Zhang, Armando Hernandez-Garcia, Kai Jiang, Zongying Gong, Durgarao Guttula, Siow Yee Ng, Piravi P. Malar, Jeroen A. van Kan, Liang Dai, Patrick S. Doyle, Renko de Vries, and Johan R. C. van der Maarel. Amplified stretch of bottlebrush-coated DNA in nanofluidic channels, Nucl. Acids Res. Accepted, 2013. DOI:10.1093/nar/gkt783.

4.1 Introduction

Advances in nanofabrication have made it possible to fabricate quasi one-dimensional devices with cross-sectional diameters on the order of tens to hundreds of nanometers. These nanochannels serve as a platform for studying, among others, single DNA molecules.[1-8] Furthermore, confinement in a nanospace result in significant modification of certain biophysical phenomena, such as the knotting probability of circular DNA and the effect of macromolecular crowding on the conformation and folding of DNA.[9-12] Nanochannels are also important for single-molecule technologies for mapping of large-scale genomic organization, including restriction enzyme cutting, nick labelling, and denaturation mapping.[13-18] Advantages of the nanochannel platform are that the DNA molecules are in an equilibrium conformation, high throughput can be achieved by using parallel arrays of channels, and integration with lab-on-chip devices. Future devices might involve non-equilibrium translocation of the DNA molecule through a pore or along a sensor, a process which also benefits from a nanochannel platform.[19] The molecules of interest are usually visualized with fluorescence microscopy.

The resolution of the genome mapping technologies relies on the ability to stretch DNA to a length close to its contour length. Stretching of single DNA molecules to more than 60% of the contour length can be achieved by confinement in channels with cross-sectional diameters below 100 nm.[3, 14, 15] Alternatively, DNA molecules can be stretched by working at very low ionic strength of less than 1 mM.[4, 5, 6] At such low ionic strength, the persistence length of the duplex exceeds 100 nm, and is significantly larger than the physiological value of around 50 nm. Following the latter strategy of amplifying the stretch by an increase in bending rigidity, we here propose to apply a polymer coating to the double-stranded DNA molecule. The advantage of this approach is that wider channels can be used to achieve a similar stretch. In particular, the less stringent channel diameter allows the use of nanofluidic devices produced by soft lithography and made of elastomer such as polydimethylsiloxane (PDMS). Elastomer chips have some advantages, including the possibility of obtaining

multiple replicas by casting on a master stamp.[3] Furthermore, a wider nanofluidic channel system allows in situ changes in environmental solution conditions, which is required for DNA binding assays. Using a novel device based on a cross-channel configuration, recently we have shown that DNA molecules trapped inside wider channels (250 nm) remain accessible for binding proteins and/or ligands.[20] The wider channels also allow loading of DNA molecules by electrophoresis.

Recently, a cationic-neutral diblock polypeptide has been developed that qualifies as a suitable candidate for coating of DNA.[21] The cationic, DNA binding block is relatively short and consist of 12 lysine residues (B^{K12}). The B^{K12} block is coupled to 4 repeats of a 100 amino acid long polypeptide (C_4).[22] The C_4 block has an amino acid composition similar to that of collagen. Furthermore, it is hydrophilic, net electroneutral, and behaves as a flexible polymer in aqueous solution. The entire C_4-B^{K12} diblock polypeptide is produced using recombinant DNA technology by large-scale expression with high yields in yeast. The diblock copolymers are monodisperse with a total molecular weight of 38.4 kDa. The polypeptide concentration is conveniently expressed as the N/P ratio that is the number of positively charged amino groups on the lysine B^{K12} binding block divided by the number of phosphate groups on the DNA. As shown in previous work, complexation of DNA with C_4-B^{K12} does not result in aggregates consisting of multiple DNA molecules.[21] Instead, single DNA molecules are uniformly coated by a dense bottlebrush as illustrated in Figure 4.1. We have verified that DNA metabolism is not inhibited by coating with the polypeptide (see Materials and Methods). Two effects are expected to enhance stretching of the DNA–C_4-B^{K12} complex in nanochannels: a larger cross-sectional diameter and a larger bending rigidity or persistence length. These effects will be gauged from nanofluidics experiments, atomic force microscopy, and Monte Carlo simulation.

4.2 Materials and Methods

4.2.1. Sample Preparation. C_4-B^{K12} polypeptide (Mw = 38.4 kDa) was expressed in the yeast *Pichia Pastoris* as an exocellular protein and purified by NH_4SO_4

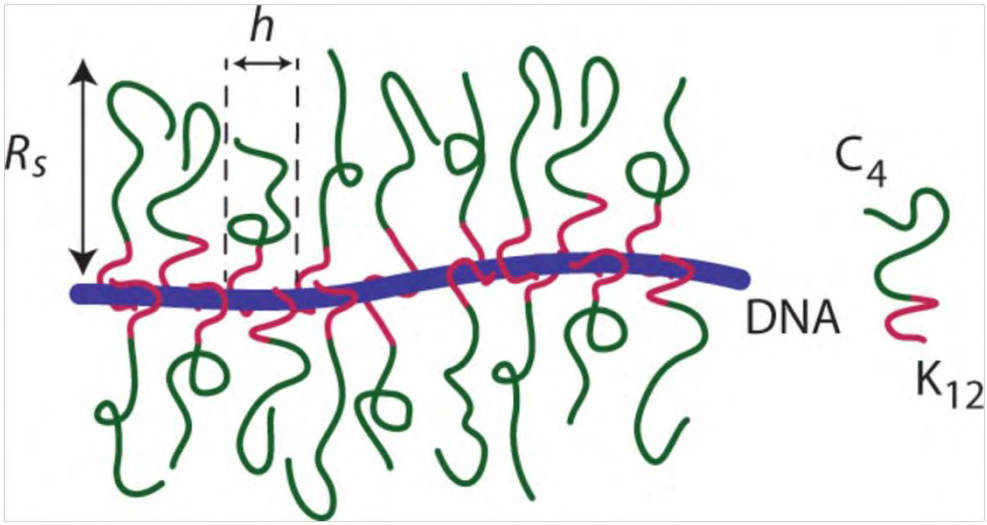


Figure 4.1. Illustration of a bottlebrush formed by binding of a diblock polypeptide copolymer to DNA. The binding block of the copolymer contains 12 cationic lysine residues.

precipitation of the supernatant.[21, 22] The purity and absence of degradation of the polypeptides have been verified with SDS-PAGE and MALDI-TOF. In order to verify the effect of polypeptide coating on DNA metabolism, we have done a digestion assay.[21] As shown by the image of the gel in Figure 4.2, bare DNA is degraded by DNAaseI in less than one minute. In the same condition, polypeptide coated DNA with N/P = 7.5 is degraded in about an hour. DNA metabolism is hence slowed down, but not inhibited by the polypeptide coating. T4-DNA was purchased from Nippon Gene, Tokyo and used without further purification. λ-DNA, *Nb.BbvCI* nicking endonuclease, and VentR (exo-) DNA polymerase were purchased from New England Biolabs, Ipswich, MA. YOYO-1 and ChromaTide Alexa Fluor 546-14- dUTP were purchased from Invitrogen, Carlsbad, CA. λ-DNA was sequence specifically stained with Alexa Fluor 546.[15] T4-DNA was stained with YOYO-1 with an intercalation ratio of 100 base-pairs per dye molecule. For nanofluidics, samples were prepared by dialyzing solutions of T4-DNA against 10 mM Tris/HCl and 1 mM EDTA in micro-dialyzers. The Tris/HCl concentration is 10 mM Tris adjusted with HCl to pH 8.5, i.e. 2.9 mM TrisCl and 7.1 mM Tris. The ionic strength of the buffer was calculated with

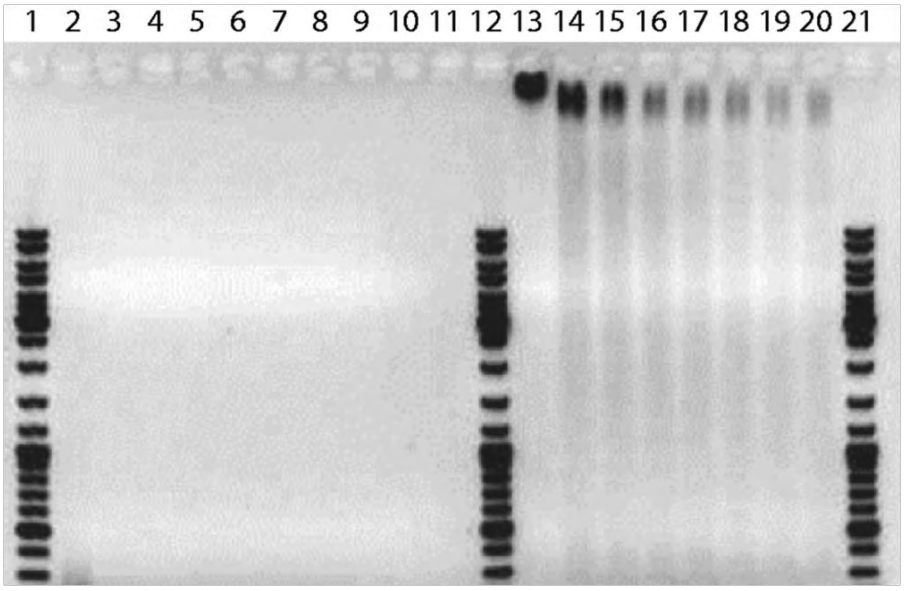


Figure 4.2. Degradation of pMTL23 (3.7 kb) by DNAaseI. The plasmid and DNAseI concentrations are 7.2 nM and 8.8 U/ml, respectively. Lanes 1, 12 and 21: DNA molar weight marker. Lanes 2–11: reaction product of bare DNA after 1, 5, 8, 15, 25, 35, 45, 60 min of incubation, respectively. Lanes 13–20: as lanes 2–11, but for C_4-B^{K12}-coated DNA with N/P = 7.5.

the Davies equation for estimating the activity coefficients of the ions and a dissociation constant pK = 8.08 for Tris. The polypeptide was dissolved in the same buffer. Solutions of polypeptide and DNA were subsequently mixed in equal volumes and incubated for 24 h at 277 K. The final DNA concentration is 3 mg/L. No anti-photo bleaching agent was used. Atomic force microscopy was done with pUC18 plasmid (2686 bp). pUC18 was purchased and linearized with HindIII restriction endonuclease (New England Biolabs, Ipswich, MA). Solutions of plasmid and polypeptide were mixed in equal volumes to a final concentration of 0.2 mg of DNA/L and incubated for 24 h at 277 K.

4.2.2. Chip Fabrication. The nanofluidic devices were fabricated by replication in PDMS of patterned master stamps.[5, 23] The stamps were made in HSQ resist (Dow Corning, Midland, MI) using a lithography process with proton beam writing.[24, 25] The 60 ± 5, 100 ± 5, 200 ± 5, 250 ± 5, and 300 ± 5 nm heights of

the positive channel structures on the stamps were measured with atomic force microscopy (Dimension 3000, Veeco, Woodbury, NY). The stamps were replicated in PDMS followed by curing with a curing agent (Sylgard, Dow Corning) at 338 K for 24 h.[26] After both substrates were plasma oxidized (Harrick, Ossining, NY), the PDMS replicas were sealed with glass slides. The widths of the channels in the replicas were measured with atomic force microscopy and the values agreed with those obtained from the scanning electron microscopy images of the master stamps. The nanochannels have a length of 60 μm and rectangular cross-sections of 100 x 60, 250 x 100, 200 x 200, 250 x 200, 250 x 250, 250 x 300, 300 x 300, and 500 x 300 nm^2. A fresh chip was made for every experiment.

To check for possible adhesion of polypeptide to PDMS, we have prepared a flat PDMS chip. The chip was plasma oxidized and covered with glass slides. A square 2 x 2 mm^2 window of the PDMS surface was left uncovered. The chip was exposed to a solution of polypeptide (1 g/L) for the typical duration of a nanofluidic experiment (30 min). The glass slides were removed, the chip was N_2-dried, and the surface was imaged with atomic force microscopy. We did not observe a difference in surface roughness or a transition in height level from the previously covered to the exposed PDMS surface area of the chip. Accordingly, there is no appreciable adhesion of polypeptide to PDMS.

4.2.3. Fluorescence Imaging. The stained DNA molecules dispersed in the relevant solution were loaded into one of the two reservoirs connected by the nanochannels. The DNA molecules were subsequently driven into the channels by electrophoresis. For this purpose, two platinum electrodes were immersed in the reservoirs and connected to an electrophoresis power supply with a relatively low voltage in the range 0.1–10 V (Keithley, Cleveland, Ohio). Once the DNA molecules were localized inside the nanochannels, the electric field was switched off and the molecules were allowed to relax to their equilibrium state for at least 60 s. The stained DNA molecules were visualized with a Nikon Eclipse Ti inverted fluorescence microscope equipped with a 200 W metal halide lamp, a filter set, and a 100x oil immersion objective. A UV light shutter controlled the exposure time.

Images were collected with an electron multiplying charge coupled device (EMCCD) camera (Andor iXon X3) and the extension of the DNA molecules inside the channels was measured with IMAGEJ software.

4.2.4. DNA Combing. Polystyrene (Mw = 280 kDa) is dissolved in toluene (Fisher Scientific, Pittsburgh, PA) at a concentration of 100 g/L. Glass cover slips were cleaned by sonication in 70% ethanol for 30 min. The cover slips were spin-coated with the solution of polystyrene for 30 s at 2000 rpm. A 5 μL droplet was spotted onto the cover slip and sheared with the help of a pipette tip along the surface. The DNA molecules were visualized with fluorescence microscopy.[27]

4.2.5. Atomic Force Microscopy. All imaging experiments were carried out at room temperature in air with a Dimension 3000 atomic force microscope, Veeco, Woodbury, NY. Images were acquired in the tapping mode with silicon (Si) cantilevers (spring constant of 20–100 N/m) and operated below their resonance frequency (typically 230– 410 kHz). The images were flattened, and the contrast and brightness were adjusted for optimum viewing conditions. A 20 μL droplet was spotted onto a fleshly cleaved mica or silica surface. Mica was used for bare DNA, whereas polypeptide coated DNA was also adsorbed on silica. After 10 min to allow for DNA adsorption onto the surface, the specimens were developed by immersing them in ultrapure water for 30 min followed by drying in a stream of N_2 gas.

4.2.6. Monte Carlo Simulation. In the Monte Carlo protocol, the DNA chain is modelled as a string of (N+1) beads, which are connected by N inextensible bonds of length lb.[28, 29] The simulation model consists of bending energy between adjacent bonds and two interaction terms. The bending rigidity is set to reproduce persistence length of 50 nm. The interactions are hard sphere repulsion between DNA beads and hard-wall repulsion between the beads and the wall. If the center of a bead is beyond the channel wall, the potential becomes infinitely large and the configuration is rejected. The effective channel diameter is hence the real diameter minus the diameter of the bead. The contour length of the DNA chain was fixed at 8 μm. The diameter of the bead was set equal to the bond length lb, which is equivalent to the chain width w. Accordingly, for chains with a fixed

contour length of 8 μm, the number of beads is 1601, 1068 or 801 for w = 5, 7:5, or 10 nm, respectively. In each Monte Carlo cycle, we carried out one crankshaft and one reptation move. The simulation started from a random conformation and usually reached equilibrium after 107 cycles. In the production run, we generated 1010 cycles and recorded the conformation every other 105 cycles. For each conformation, we calculated the extension as the maximum span of the DNA molecule along the channel axis. Finally, the extension was averaged over the ensemble of 105 conformations. Results for different persistence lengths and widths were obtained from the scale invariance of the Monte Carlo results for the relative extension. This implies that the maximum span divided by the contour length of a chain with persistence length P, width w, contour length L, and confined inside a channel of diameter D is the same as the one of a chain with scaled parameters αP, αw, and αL inside a channel of diameter αD. We have verified that effects of finite contour length on the relative stretch are unimportant for the relevant range of channel diameters.

4.3 Results and Discussion

4.3.1. Amplification of Stretch. Our nanofluidic devices are made of PDMS casted on a high quality master stamp, which was obtained by proton beam writing and UV lithography.[5, 23, 24] T4-DNA (166 kbp) was stained with a ratio of one YOYO-1 molecule per 100 base pairs. With this intercalation ratio, the YOYO- 1 corrected contour length of T4-DNA amounts 57 μm. Furthermore, the concomitant minimal reduction in DNA charge has no effect on the electrostatic binding of the cationic block of the polypeptide on DNA. The stained molecules were subsequently incubated in solutions of various concentrations of polypeptide for at least 24 h. The molecules are accordingly coated with various amounts of C_4-B^{K12}, as indicated by the lysine to DNA N/P ratio (the DNA concentration is 3 mg/L).

In the nanofluidic experiment, the coated and stained DNA molecules were electrophoresed into an array of 60 μm long and rectangular nanochannels and

imaged with fluorescence microscopy. A montage of images of single T4- DNA molecules in 10 mM Tris/HCl (2.9 mM TrisCl, 7.1 mM Tris, pH 8.5) and confined in channels with a cross-section of 250 x 250 nm^2 is shown in Figure 4.3A. The images refer to well-equilibrated conformations. After the electric field has been switched off, the molecules relax to their equilibrium state within 60 s. Video imaging was started 5–10 min after the molecules were brought into the channels and lasted for another 10 min. With increasing C_4-B^{K12} to DNA ratio, the equilibrated molecules stretch in the longitudinal direction of the channel. Notice that the stretch of the bottlebrush with a twofold excess of polypeptide to DNA charge (N/P = 2.0) reaches about 85% of the contour length. An excess of polypeptide is needed to achieve maximum stretch due to the dynamic equilibrium of the binding process (law of mass action). A relative stretch of 85% has also been reported for DNA in narrow, 45-nm channels as well as for DNA at low ionic strength of 0.12 mM in 250-nm channels.[6, 17] An advantage of the present technology is the less stringent conditions in terms of channel diameter and ionic strength of the supporting medium.

We have measured the extension of the T4-DNA molecules confined in the nanochannels of various cross-sectional diameters. With increasing N/P ratio, it becomes more difficult to electrophorese DNA molecules into channels of smaller cross-sectional diameter. For the highest employed N/P ratio (2.0), we could not insert a large enough number of molecules into channels with a diameter less than 200 nm. For each experimental condition, we have used a fresh PDMS replica and measured around 50 molecules. The distributions in extension are close to Gaussian. Examples pertaining to different N/P ratios are shown in Figure 4.3B. DNA fragments can easily be discerned, because their extensions fall below the values pertaining to the intact molecules. For the cutoff, we have used the mean value minus two time s the standard deviation. Resolution broadening can be neglected, because the optical resolution of 200 nm is one order of magnitude smaller than the variance. The relative extensions Lk = L, that is the mean stretch divided by the contour length of the DNA molecule, are set out in Figure 4.4. With decreasing channel diameter and/or increasing C_4-B^{K12} to DNA ratio, the relative

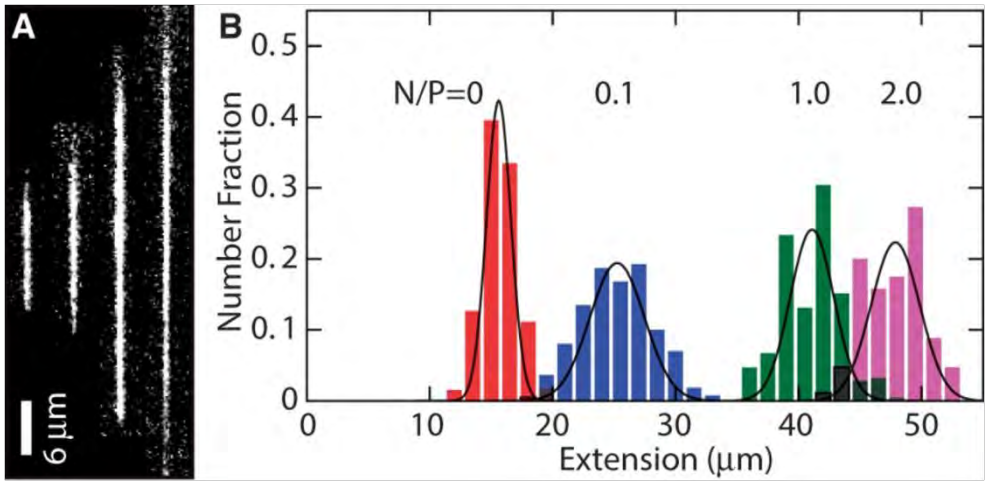


Figure 4.3. (A) Montage of fluorescence images of T4-DNA in 250 × 250 nm^2 channels and in 10 mM Tris/HCl (pH 8.5). From left to right, N/P = 0 (polypeptide-free), 0.1, 1.0 and 2.0. The scale bar denotes 6 µm and the YOYO-1 intercalation ratio is 100 bp per dye molecule. (B) Distribution in extension of a population of ~50 T4-DNA molecules with indicated N/P ratio. Gaussian fits give mean extensions of = 16 ± 2, 25 ± 3, 41 ± 3 and 48 ± 3 µm for N/P = 0, 0.1, 1.0 and 2.0, respectively.

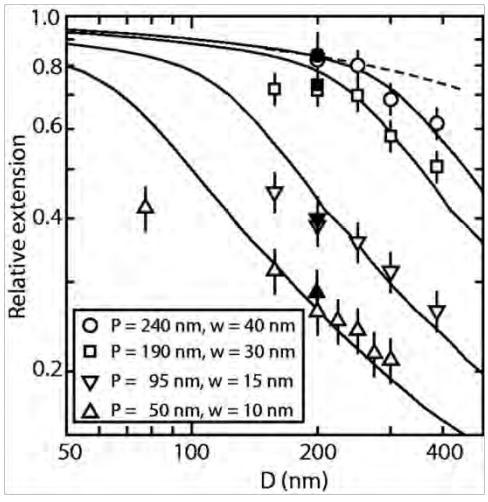


Figure 4.4. Relative extension of T4-DNA (open symbols) and λ-DNA (closed symbols) in 10 mM Tris/HCl versus channel diameter *D*. The N/P ratios are 0 (△), 0.1 (▽), 1.0 (□) and 2.0 (○). The solid curves represent Monte Carlo results with noted values of persistence length *P* and width *w*. The dashed curve represents deflection theory for narrow channels.

extension increases. For N/P = 0 (polypeptide-free) and 0.1, the extensions are in the range 0.2 – 0.5 times the contour length. In the case of N/P ratios of 1.0 and 2.0, the relative extensions are in the range 0.5 – 0.85. Furthermore, the extensions level off at a plateau value for smaller channel diameters and higher N/P ratios. The increase in relative extension is due to an increase in bending rigidity and cross-sectional diameter of the bottlebrush coated DNA. These effects will be gauged below with atomic force microscopy. We will however first demonstrate that large-scale genomic organization can be mapped with our channel platform in tandem with polypeptide coating.

4.3.2. Large Scale Genome Mapping. The aim of large scale genome mapping is not to determine the genetic code at the single base level, but rather to provide a map for assembling the genome at the larger scale and to sort out large scale variations such as insertions, inversions, and translocations. For the demonstration of genome mapping, we have site-specifically labelled nicks of λ-DNA (48.5 kbp) with Alexa Fluor 546.[15, 17] The YOYO-1 and Alexa labelled DNA was coated with C_4-B^{K12} and electrophoresed into an array of nanochannels with cross-sectional diameters of 150x 250 nm^2. From an analysis of the YOYO-1 fluorescence, we have verified that the relative extensions of λ-DNA are the same as for T4-DNA (see Figure 4.4). There are four resolvable Alexa labelling sites at 8, 18.3 (average of two sites at 18.1 and 18.5 kbp), 31.3 (average of 30.9, 31.2, and 31.8 kbp), and 35.8 kbp. Furthermore, the sticky ends of the λ-DNA molecule can also be labelled. As can be seen in panel A of Figure 4.5, individual Alexa labels are discernible with an N/P ratio of 2.0. Individual labels can also be discerned for a stoichiometric ratio of N/P = 1.0, albeit with decreased separation. In the case of polypeptide-free DNA, individual Alexa labelled sites cannot be discerned. This confirms that an excess of polypeptide to DNA charge is needed to achieve the highest possible stretch. For comparison, in panel B of Figure 4.5 we have displayed fluorescence images of combed, bare DNA molecules. In the combing experiment, the molecules are stretched to almost their full contour length. In both the nanofluidic and combing experiments, occasionally some Alexa labels are missing (in particular the ones at the end). Out of a pool of 100 molecules, 2, 10, and 40 % of the molecules show 6, 5

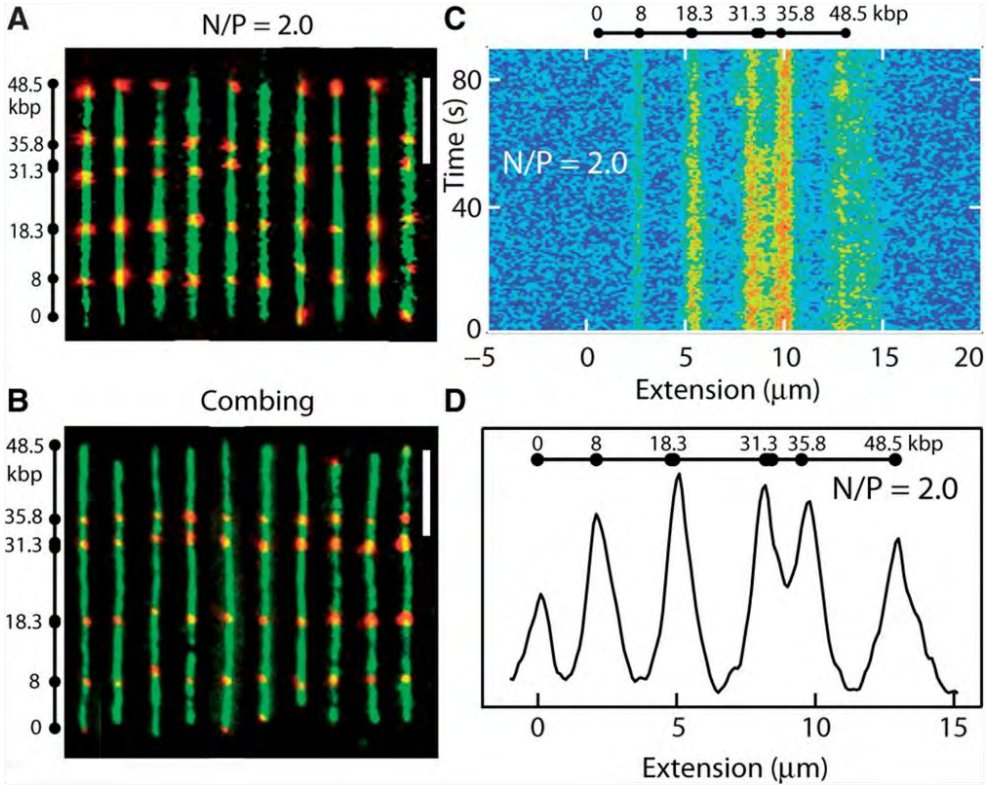


Figure 4.5. (**A**) Montage of fluorescence images of λ-DNA. The images obtained with YOYO-1 (green) and Alexa Fluor 541 (red) staining are superposed. The molecules are confined in 150 × 250 nm^2 channels with N/P = 2.0. Labelling sites are noted. The scale bar denotes 5 μm. (**B**) As in panel (A), but for combed λ-DNA. (**C**) Time dependence of the Alexa Fluor intensity profile pertaining to N/P = 2.0 and in a 150 × 250 nm^2 channel. The exposure time is 500 ms. (**D**) Average Alexa profile over a pool of 50 molecules with N/P = 2.0 and in 150 × 250 nm^2 channels.

and 4 discernible labels, respectively. Since we did not observe a difference in labeling efficiency for combed (bare) and channel-confined (coated) molecules, missing labels are due to incomplete labeling rather than caused by the coating procedure.

In the case of maximum stretch for N/P = 2.0 in 150x250 nm^2 channels, we have analyzed the Alexa fluorescence profiles of a pool of 50 DNA molecules. An example of the time-dependence of the Alexa fluorescence is shown in panel C of

Figure 4.5. On the time scale of the experiment (90 s), the molecules are trapped inside the channel and there is no net displacement. Single molecule Alexa profiles were aligned and corrected for a minor variation in overall stretch of less than 5%. The locations of corresponding peaks for different molecules are in register. As shown by the fluorescence images and the average profile in panels A and D of Figure 4.5, respectively, the locations of the labels along the molecule are in good agreement with the locations of the nicking sites. Accordingly, the molecules are uniformly stretched along the direction of the channel without ever folding back. The averaged Alexa fluorescence intensity profile confirms a relative stretch of about 85% for N/P = 2.0, as determined by the YOYO-1 fluorescence. The resolution in peak position is around 2 kbp, which is similar to the one reported for bare DNA in 45-nm channels.[17]

4.3.3. Persistence Length and Width. The persistence lengths were obtained from analysis of atomic force microscopy images of bottlebrush coated DNA on silica. For this purpose, linearized pUC18 plasmid (2686 bp) was used at a concentration of 0.2 mg/L. The C_4 blocks of the diblock polypeptide weakly adsorb to silica such that no additions to the buffer are necessary to promote adhesion. The C_4-block mediated adsorption of bottlebrush coated DNA is relatively weak, since it is found that the molecules are more easily removed by flushing with water as compared to DNA molecules that have been adsorbed in the presence of either MgCl2 or high molecular weight polylysine. Accordingly, the bottlebrush is not kinetically trapped in a 3D conformation.[30] For the reference case of bare DNA, molecules were bound to a mica surface by the use of a buffer comprising 10 mM $MgCl_2$. Excess polypeptide and $MgCl_2$ were removed from the interface by immersion of the specimens in ultrapure water for 30 min. During this development time, the molecules equilibrate on the surface in a 2D conformation. Subsequently, all specimens were N_2-dried. A series of typical images for increasing C_4-B^{K12} to DNA ratios are shown in Figure 4.6A-D. The relatively small plasmids are imaged in their entirety with a field of view of 3 x 3 μm^2.We have also imaged polypeptide coated λ-DNA with N/P = 2.0. However, as shown in panels E and F of Figure 4.6, only segments of these relatively long molecules can be visualized.

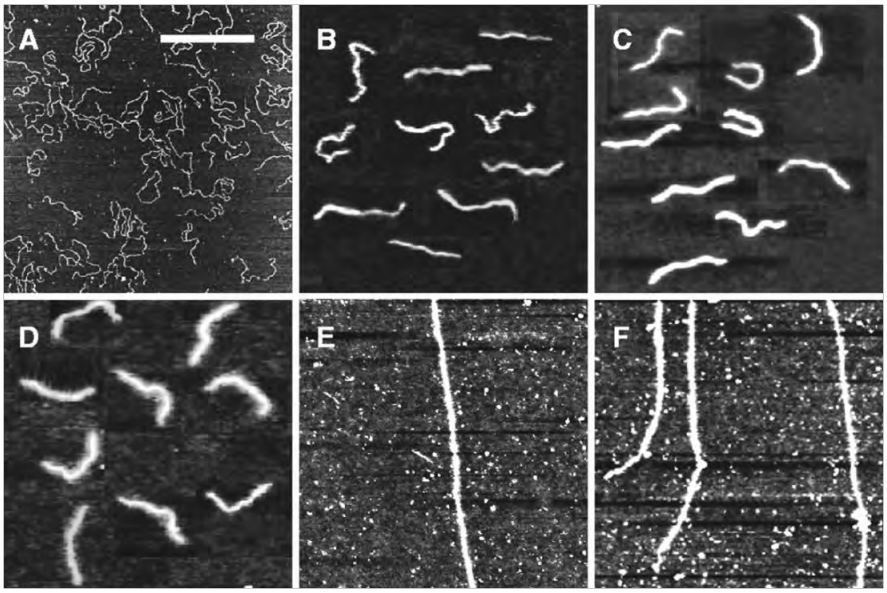


Figure 4.6. (A–D) Tapping mode atomic force microscopy images of linearized pUC18 DNA complexed with increasing amounts of polypeptide. The N/P ratios are 0, 0.1, 1.0 and 2.0, in panels A, B, C and D, respectively. (E, F) As in panels (A–D), but for DNA with N/P = 2.0. The scale bar denotes 1 μm. Panels B–D are a montage.

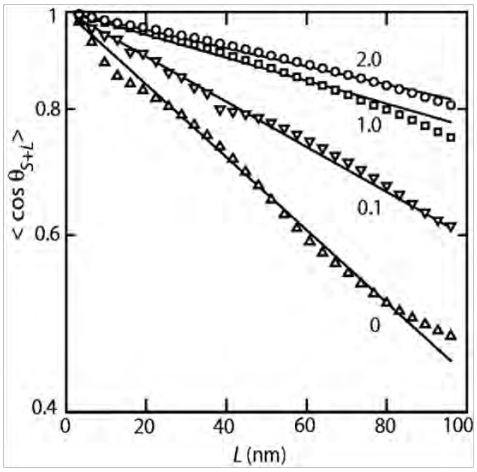


Figure 4.7. Orientation correlation of the tangent vectors at a pair of points separated by distance L along the contour. The lines represent exponential fits, and the polypeptide-to-DNA N/P ratios are noted. The fitted values of the persistence lengths $P = 60 \pm 5$, 95 ± 5, 190 ± 10 and 240 ± 10 nm for N/P = 0, 0.1, 1.0 and 2.0, respectively.

Nevertheless, their appearance in terms of thickness and rigidity is similar to the one of the shorter plasmid molecules.

To obtain the persistence length from the atomic force microscopy images, we traced the centerline of a population of 10–30 individual pUC18 molecules, with the smaller number pertaining to the highest polypeptide concentrations. These were used to obtain the tangent vector correlation function $\langle \mathrm{Cos}\theta_{s,s+L} \rangle$, where θ is the angle between tangent vectors at points s and s+L, by averaging s along the contour. For weakly adsorbed DNA molecules equilibrated in 2D conformation, this correlation follows:

$$\langle \mathrm{Cos}\theta_{s,s+L} \rangle = \exp(-L/(2P)) \qquad (4.1)$$

Experimental and fitted tangent correlation functions are shown in Figure 4.7.With increasing C_4-B^{K12} to DNA ratio, fitted values of the persistence length P increase from 60 nm (bare DNA) to 240 nm at the highest polypeptide concentration.

The value obtained for bare DNA agrees with the commonly accepted value of around 50 nm.[31] We have measured the length of the molecules by tracing the contour. No change in contour length induced by the binding of the polypeptide on DNA was observed.

The atomic force microscopy images show an increase in thickness of the bottlebrushes with increased polypeptide concentration. We analyzed the cross-sectional profiles taken at 10 different and randomly chosen positions of the molecules. Gaussian fits to these profiles gives variances of 10, 15, 20, and 30 nm for N/P = 0, 0.1, 1.0, and 2.0, respectively. These values should however be taken as indicative of the thickness of the brushes, because the profiles are broadened by the width of the tip and the brushes are dried and spread out on the mica surface.

The experimental values of the stretch can be compared to Monte Carlo results that we have obtained for the stretching of nanoconfined semiflexible polymers varying in persistence length P and effective thickness w. The computational results are also displayed in Figure 4.4. In comparing the experimental values to the Monte Carlo simulations, we have used the values of the persistence length P as obtained from atomic force microscopy, without any adjustment. Only for bare DNA, a slightly better fit was obtained with the nominal persistence length of 50 nm. For each N/P, the width w used in the Monte Carlo simulations was adjusted to obtain the best agreement with the experimental stretching data. The values of w obtained in this way show an increase with increased polypeptide coating and are in reasonable agreement with the variances of the cross-sectional profiles obtained from atomic force microscopy.

4.3.4. Theoretical Considerations. For D < 250 nm and N/P = 2.0, the data are close to the prediction of Odijk's deflection model. Here, the bottlebrush is undulating along the channel by deflections from the wall.[34] For larger values of D, the data fall below the mark set by deflection theory due to back-folding and looping of the undulating wormlike bottlebrush conformation.[28]

Existing theories support our findings of increasing effective thickness and persistence lengths for bottlebrush coated DNA.[21] Assuming that saturated binding corresponds to full charge neutralization of the DNA phosphates by the lysines on the B^{K12} binding block, a single C_4-B^{K12} molecule has a binding site size of 6 base pairs, and saturated binding corresponds to a distance h between bound polypeptides along the DNA contour of 2 nm. Recall that this requires an excess of polypeptide to DNA charge (NP > 1), due to the dynamic equilibrium of the binding process. Side chain stretching in bottlebrushes is controlled by the dimensionless parameter

$$\Gamma = Nl/h \qquad (4.2)$$

where N is the number of segments in a side chain, l is the length of a side chain segment, and h is the distance between grafted side chains along the main chain. If

Rs is the mean square end-to-end distance of the stretched side chain (in the bottlebrush) and R is the mean-square end-to-end distance of an unstretched side chain (in free solution), the scaling prediction for the degree of side chain stretching takes the form[32]

$$Rs = R\,\alpha\,\Gamma^{1/4} \qquad (4.3)$$

Assuming l = 0.4 nm and N = 400, we find $\Gamma^{1/4} \approx 3.0$ for the charge neutral complex with h = 2 nm. Taking the hydrodynamic radius R_H = 6 nm for the chain in free solution as a measure for R,[21] the scaling estimate suggests that the side chains stretch to a value on the order of 20 nm. A precise value of Rs cannot be given, since the proportionality factor in Eq. 3 is unknown. Nevertheless, a maximum cross-sectional diameter of the bottlebrush of around 40 nm at saturated binding is consistent with our experimental estimate for N/P = 2.0. The persistence length P of bottlebrush polymers can be estimated based on scaling theory that was verified by self-consistent field calculations[21, 33]

$$P = P_0 + \mu N^2 l^3 / h^2 \qquad (4.4)$$

with a small numerical constant μ = 0.02. For h = 2 nm and the commonly accepted value of P_0 = 50 nm for the DNA main chain intrinsic stiffness, the above theory predicts P = 150 nm, or a 3-fold increase as compared to bare DNA. Hence, theoretical estimates confirm the expected order of magnitude of the stiffening effect, even though the predicted increase is lower than the 4 to 5-fold increase at N/P = 2.0 that we have estimated from the analysis of the atomic force microscopy images and the extensions of the complexes confined inside the nanochannels.

4.4 Conclusions

We have shown that we can amplify the stretch of DNA confined in nanochannels through a coating of the duplex with a diblock polypeptide. For channels with cross-sectional diameters of around 200 nm we can achieve a

uniform stretch of about 85% of the contour length. Due to the less stringent cross-sectional dimension of a few hundred nm, the channels are relatively easy to make out of elastomer with soft lithography. Furthermore, the DNA molecules can be loaded within the nanochannel using electrophoresis. Monte Carlo computer simulation shows that the stretch is due to an increase in bending rigidity and thickness of the bottlebrush coated DNA. The persistence lengths and widths deduced from the stretch agree with what has been observed with atomic force microscopy. The highest degree of stretching corresponds to the deflection limit, in which the bottlebrush undulates within the nanochannel without ever folding back. We have demonstrated that the amplification of the stretch in tandem with site-specific fluorescence labeling allows the investigation of large scale genomic organization of λ-phage DNA. The limiting factor of the nanochannel platform is the length of the channels. With present day technology it is possible to fabricate linear channels with a length of around a few mm, so that DNA molecules of a few Mbp can be stretched. For longer DNAs, one has to use fragments and/or nanofluidic devices with a back-folded channel layout.

4.5 References

1. Reisner, W., Morton, K.J., Riehn, R., Wang, Y.M., Yu, Z., Rosen, M.,Sturm, J.C., Chou, S.Y., Frey, E. and Austin, R.H. (2005) Statics anddynamics of single DNA molecules confined in nanochannels. Phys. Rev.Lett., 94, 196101.

2. Mannion, J.T., Reccius, C.H., Cross, J.D. and Craighead, H.G. (2006) Conformational analysis of single DNA molecules undergoing entropically induced motion in nanochannels. Biophys J., 90, 4538–4545.

3. Reisner, W., Beech, J.P., Larsen, N.B., Flyvbjerg, H., Kristensen, A. and Tegenfeldt, J. O. (2007) Nanoconfinement-enhanced conformational response of single DNA molecules to changes in ionic environment. Phys. Rev. Lett., 99, 058302.

4. Jo, K., Dhingra, D.M., Odijk, T., de Pablo, J.J., Graham, M.D, Runnheim, R., Forrest, D., Schwartz, D.C. (2007) A single-molecule barcoding system using nanoslits for DNA analysis. Proc. Natl. Acad. Sci. U.S.A., 104, 2673–2678.

5. Zhang, C., Zhang, F., van Kan, J.A. and van der Maarel, J.R.C. (2008) Effects of electrostatic screening on the conformation of single DNA molecules confined in a nanochannel. J. Chem. Phys., 128, 225109.

6. Kim, Y., Kim, K.S., Kounovsky, K.L., Chang, R., Jung, G.Y., dePablo, J.J., Jo, K. and Schwartz, D.C. (2011) Nanochannel confinement: DNA stretch approaching full contour length. Lab Chip, 11, 1721–1729.

7. Persson, F., Utko, P., Reisner, W., Larsen, N.B., Kristensen, A. (2009) Confinement spectroscopy: probing single DNA molecules with tapered nanochannels. Nano Lett., 9, 1382–1385.

8. Su, T. Das, S.K., Xiao, M. and Purohit, P.K. (2011) Transition between two regimes describing internal fluctuation of DNA in a nanochannel. PLoS One, 6, e16890.

9. Dai, L., van der Maarel, J.R.C. and Doyle, P.S. (2012) Effect of nanoslit confinement on the knotting probability of circular DNA. ACS Macro Lett., 1, 732–736.

10. Zhang, C., Shao, P.G., van Kan, J.A. and van der Maarel, J.R.C. (2009) Macromolecular crowding induced elongation and compaction of single DNA molecules confined in a nanochannel. Proc. Natl. Acad. Sci. U.S.A., 106, 16651–16656.

11. Jones, J.J., van der Maarel, J.R.C. and Doyle, P.S. (2011) Effect of nanochannel geometry on DNA structure in the presence of macromolecular crowding agent. Nano Lett., 11, 5047–5053.

12. Zhang, C., Gong, Z.Y., Guttula, D., Malar, P.P., van Kan, J.A., Doyle, P.S. and van der Maarel, J.R.C. (2012) Nanouidic compaction of DNA by like-charged protein. J. Phys. Chem. B, 116, 3031–3036.

13. Riehn, R., Lu, M., Wang, Y.M., Lim, S.F., Cox, E.C. and Austin R.H. (2005) Restriction mapping in nanofluidic devices. Proc. Natl. Acad. Sci. U.S.A., 102, 10012–10016.

14. Levy, S.L. and Craighead, H.G. (2010) DNA manipulation, sorting, and mapping in nanofluidic systems. Chem. Soc. Rev., 39, 1133–1152.

15. Das, S.K., Austin, M.D., Akana, M.C., Deshpande, P., Cao, H. and Xiao, M. (2010) Single molecule linear analysis of DNA in nano-channel labelled with sequence specific fluorescent probes. Nucleic Acids Res., 38, e177.

16. Reisner, W., Larsen, N.B., Silahtaroglu, A., Kristensen, A., Tommerup, N., Tegenfeldt, J.O. and Flyvbjerg H. (2010) Single-molecule denaturation mapping of DNA in nanofluidic channels. Proc. Natl. Acad. Sci. U.S.A. 107, 13294–13299.

17. Lam, E.T., Hastie, A., Lin, C., Ehrlich, D., Das, S.K., Austin, M.D., Deshpande, P., Cao, H., Nagarajan, N., Xiao, M. and Kwok, P.Y. (2012) Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nat. Biotechnol., 30, 771–776.

18. Reisner,W., Pedersen, J.N. and Austin, R.H. (2012) DNA Confinement in nanochannels: physics and biological applications. Rep. Prog. Phys., 75, 106601.

19. Min, S.K., Kim, W.Y., Cho, Y. and Kim, K.S. (2011) Fast DNA sequencing with a graphene-based nanochannel device. Nature Nanotech. 6, 162–165.

20. Zhang, C., Jiang, K., Liu, F., Doyle, P.S., van Kan, J.A. and van der Maarel, J.R.C. (2013) A nanofluidic device for single molecule studies with in situ control of environmental solution conditions. Lab Chip, 13, 2821–2826.

21. Hernandez-Garcia, A., Werten, M.W., Stuart, M.C., de Wolf, F.A. and de Vries, R. (2012) Coating of single DNA molecules by genetically engineered protein diblock copolymers. Small, 8, 3491–3501.

22. Werten, M.W., Wisselink, W.H., Jansen-van den Bosch, T.J., de Bruin, E.C. and de Wolf, F.A. (2001) Secreted production of a custom-designed, highly hydrophilic gelatin in Pichia pastoris. Protein Eng., 14, 447–454.

23. van Kan, J.A., Zhang, C., Malar, P.P. and van der Maarel, J.R.C. (2012) High throughput fabrication of disposable nanofluidic lab-on-chip devices for single molecule studies. Biomicrofluidics, 6, 036502.

24. van Kan, J.A., Bettiol, A.A. and Watt, F. (2006) Proton beam writing of three-dimensional nanostructures in hydrogen silsesquioxane. Nano Lett., 6, 579–582.

25. van Kan, J.A., Bettiol, A.A. and Watt, F. (2003) Three-dimensional nanolithography using proton beam writing. Appl. Phys. Lett., 83, 1629–1631.

26. Shao, P.G., van Kan, J.A., Ansari, K., Bettiol, A.A. and Watt, F. (2007) Poly (dimethyl siloxane) micro/nanostructure replication using proton beam written masters. Nucl. Instrum. Methods Phys. Res. B, 260, 479–482.

27. Allemand, J.F., Bensimon, D., Jullien, L., Bensimon, A. and Croquette, V. (1997) pH-dependent specific binding and combing of DNA. Biophys. J., 73, 2064–2070.

28. Dai, L., Ng, S.Y., Doyle, P.S. and van der Maarel, J.R.C. (2012) Conformation model of back-folding and looping of a single DNA molecule confined inside a nanochannel. ACS Macro Lett., 1, 1046–1050.

29. Dai, L., Jones, J.J., van der Maarel, J.R.C. and Doyle, P.S. (2012) A systematic study of DNA conformation in slit-like confinement. Soft Matter, 8, 2972–2982.

30. Rivetti, C., Guthold, M. and Bustamante, C. (1996) Scanning force microscopy of DNA deposited onto mica: Equilibrium versus kinetic trapping studied by statistical polymer chain analysis. J. Mol. Biol., 264, 919–932.

31. Wiggins, P. A., van der Heijden, T., Herrero, F. M., Spakowitz, A., Phillips, R., Widom, J., Dekker, C. and Nelson, P. C. (2006) High flexibility of DNA on short length scales probed by atomic force microscopy. Nat. Nanotechnol., 1, 137–141.

32. Hsu, H.-P., Paul, W. and Binder K. (2011) Understanding the multiple length scales describing the structure of bottle brush polymers by Monte Carlo simulation methods. Macromol. Theory Simul., 20, 510–525.

33. Feuz, L., Leermakers, F.A.M., Textor, M. and Borisov, O. (2005) Bending rigidity and induced persistence length of molecular bottle brushes: a selfconsistent field theory. Macromolecules, 38, 8891–8901.

34. Odijk, T. (1983) On the statistics and dynamics of confined or entangled stiff polymers. Macromolecules, 16, 1340–1344.

Chapter 5

Protein polymer incorporating a DNA binding domain:

Coating, stiffening and delivery of single DNA molecules[1]

Emerging DNA based nanotechnologies would benefit from the ability to modulate the physical properties of individual DNA molecules. Whereas previous approaches have mostly focused on using (block co-)polymers, we here introduce natural DNA-binding proteins connected to long hydrophilic blocks as generic agents to modulate DNA physical properties. A DNA binding domain (7kDa) is linked to a hydrophilic (73kDa) random coil polypeptide and produced in yeast. This protein polymer coat individual single and double stranded DNA molecules with high specificity, stiffening both to a considerable extent and remarkably, prevents hybridization in single-stranded DNA. We show that the protein polymer is non-toxic and non-hemolytic and show a transfection efficiency that is comparable to current non-viral gene delivery standards. Incorporating natural DNA-binding domains into protein-based polymers is a versatile way to modulate physical properties of individual DNA molecules that can ultimately find potential applications in both emerging DNA nanotechnologies and non-viral gene transfer.

---

[1] This chapter is based on: Armando Hernandez-Garcia, Marco Favretto, Roland Brock, Frits A. de Wolf, and Renko de Vries. Protein polymer incorporating a DNA binding domain: Coating, stiffening and delivery of single DNA molecules. Submitted, 2013.

5.1 Introduction

New DNA nanotechnologies based on DNA origami[1-4] and single molecule DNA imaging[5] or sequencing strategies such as optical mapping,[6-9] and nanopore sequencing[10, 11] increasingly rely on control of the physical-chemical properties of individual DNA molecules: mechanical properties, interactions with nanoscale environments,[12] etc. While there is some room for control via tuning of solution conditions, it may be expected that much higher levels of control can be obtained by developing dedicated DNA binders that modulate specific DNA properties. Polymers developed as non-viral delivery agents for nucleic acids can also be considered to be dedicated DNA binders that modulate specific DNA properties. A general toolbox of DNA binders that modulate physical properties of individual DNA molecules may therefore be expected to be useful for a wide range of technologies.

With this in mind, we have recently designed, produced and characterized a recombinant, protein-based polymer that coats single double stranded DNA molecules and significantly increases its persistence lengths.[13] The polymer coating protects against enzymatic degradation without making the DNA completely inaccessible to strong (sequence specific) binders. It is composed of two polypeptide domains and its sequence is abbreviated as C_4-B^{K12}. The C_4 domain is a tetramer of a previously published collagen-inspired sequence C, a 98 amino acid polypeptide that is highly hydrophilic and forms random coils over a wide range of solution conditions. As a DNA binding domain, a simple stretch of 12 lysines was used (K_{12}). The enhanced DNA stiffness provided by the C_4-B^{K12} polymer coating has already been shown to be useful for nanochannel-based optical mapping of DNA, where it allows for full stretching of DNA in rather wide nanochannels (250x250nm)[8]. In addition, as we show here the C_4-B^{K12} polymer is an effective non-viral transfer agent for single plasmid DNA molecules.

However, as a binding domain, the oligolysine domain is quite unspecific. Virtually all anticipated applications of DNA binders that modulate specific DNA properties will involve complicated background solutions composed of

biopolymers other than DNA and the oligolysine domain will also bind to those molecules. The same of course holds for most synthetic polycationic blocks currently used for non-viral gene transfer such as Poly-Lys, PEI or PDMAEMA. A logical next step is therefore to replace non-specific polycationic binding blocks by nucleic acid-specific binding domains. Here we choose the 7kDa DNA-binding protein Sso7d of the hyperthermophilic archaeabacterium *Sulfolobus solfataricus* as a nucleic acid, but non sequence specific, binding domain.[14-19] The binding characteristics of Sso7d to both double- and single stranded DNA have been very well characterized. It is a highly stable and easily produced small protein that has also been used by other groups as a basis for further engineering: sequence unspecific DNA binding by Sso7d has been used to improve processivity of DNA polymerases[20] and as a structural scaffold to generate highly stable binding proteins.[21, 22]

We here produce and study the properties of a fusion of Sso7d with a very long shielding domain: an octamer of the previously mentioned 98 amino acid C-domain. The C_8-B^{Sso7d} protein-based polymer is produced with high yields by secreted expression in the yeast *Pichia pastoris*. We study DNA stiffening by coating with C_8-B^{Sso7d}, for both double stranded- and single stranded DNA, but also non-viral gene transfer. Our study highlights how engineered protein-based polymers can be used in fusions with natural folded domains to construct protein-based polymers with new abilities to modulate the physical properties of DNA for a variety of applications, including, for example, in synthetic biology.

5.2 Materials and Methods

5.2.1. Materials. Linear 8.0 and 2.5 kbp dsDNA were purchased from Thermo Scientific (Waltham, MA, USA). M13mp18 single stranded DNA 7249 nt (ssDNA) was purchased from New England Biolabs (Ipswich, MA, USA) and T4-DNA were purchased from Nippon Gene (Tokyo, Japan) and used without further purification, supercoiled 4.0 and 2.7 kbp pDNA were recovered from recombinant *E. coli* by using the GeneJet plasmid Miniprep kit from Thermo Scientific. pDNA-

YFP was produced recombinantly. YOYO-1 was purchased from Invitrogen (Life Technologies, Carlsbad, CA, USA). Restriction enzymes were purchased from New England Biolabs or from Thermo Scientific. Proteins C_4B^{K12} and C_4 protein were produced and purified according following previously published methods.[13, 23] Resazurin was purchased from Sigma (St. Louis, MO, USA). Fetal Calf Serum (FCS) and Roswell Park Memorial Institute medium 1640 (RPMI-1640) were purchased from Gibco (Life Technologies) and HEPES Buffered Saline solution (HBS) was prepared locally. 25 kDa branched polyethyleneimine (PEI), and Lipofectamine 2000 were purchased from Sigma and Invitrogen, respectively.

5.2.2. Cloning of the Protein and Strains. Two double stranded oligonucleotides, previously annealed from complementary codon-optimized oligonucleotides (produced by Eurogentec, Belgium, see Supporting Information Table S1 for DNA sequences), encoded the full amino acid sequence of the Sso7d binding domain (B^{Sso7d}) used in this study. The vector containing the DNA coding for "C_8" was prepared in the follow way: a fragment C_4, obtained from the plasmid pMTL23-C4 by double digestion with the restriction enzymes *Van*91I/*Dra*III, was ligated into the plasmid pMTL23-C4 previously digested with *Van*91I to obtain pMTL23-C8. The plasmid pMTL23-C_8B^{Sso7d} was obtained by ligating the two annealed double stranded oligonucleotides coding for B^{Sso7d} into the vector pMTL23-C8 previously digested with restriction enzymes *Van*91I and *Eco*RI. The fragment encoding the C_8B^{Sso7d} protein-based polymer was released through digestion of plasmid pMTL23-C_8B^{Sso7d} with *Xho*I/*Eco*RI and ligated into the correspondingly *Xho*I/*Eco*RI digested *P. pastoris* expression vector pPIC9 (Invitrogen). The resulting plasmid pPIC9-C_8B^{Sso7d} was linearized with *Sal*I and electroporated into *P. pastoris* strain GS115 (Invitrogen). The plasmid integrates into the genome through homologous recombination at the *his4* locus providing normal growth on methanol. The presence of the genes was verified by polymerase chain reaction.

5.2.3. Biosynthesis of Hybrid Protein. The fermentation was similar to the previously described method.[13] Fed-batch fermentations using minimal basal salts medium were performed in 2.5-L Bioflo 3000 fermentors (New Brunswick

Scientific, Edison, NJ). The methanol fed-batch phase for protein production lasted 2–3 days. A homemade semiconductor gas sensor–controller was used to monitor the methanol level in the off-gas and to maintain a constant level of 0.2% (w/v) methanol in the broth. The pH was maintained at 3.0 throughout the fermentation by addition of ammonium hydroxide. At the end of the fermentation, the cells were separated from the broth by centrifugation for 15 min at 10 000 $\times g$ (room temperature or 4 °C) in an SLA-3000 rotor (Thermo Scientific, Waltham, MA), and the supernatant was microfiltered (Pall Corporation, Port Washington, NY) and immediately stored at 4 °C for subsequent purification.

5.2.4. Protein Purification. All centrifugation was done for 30 min at 20 000 $\times g$ at 4 °C, interchangeably in a Sorvall SLA-1500 or SLA-3000 rotor (Thermo Scientific, Waltham, MA). First, medium salts were removed from the cell-free broth by adjustment of the pH to 8.0 with NaOH, followed by centrifugation. Subsequently, the protein-based polymer was selectively precipitated from the solution by adding ammonium sulfate to a saturation of 45%, incubating overnight at 4°C, and subsequent centrifugation. The pellet was resuspended in an equal volume (relative to the cell-free broth) of Milli-Q water and precipitation was repeated once at 4°C, using an overnight incubation. The pellet was resuspended in 0.2 volumes (relative to the cell-free broth) of Milli-Q water and sodium chloride and acetone were added to a final concentration of 50 mM and 40% (v/v), respectively. After centrifugation the acetone concentration of the supernatant was raised to 80% (v/v), and the solution was centrifuged to precipitate the pure protein-based polymer. The pellet was dried overnight, resuspended in Milli-Q water, extensively desalted by dialysis against Milli-Q water using a using Spectra/Por 7 tubing (Spectrum Laboratories) with a 1 kDa molecular weight cutoff and lyophilized.

5.2.5. SDS-PAGE and MALDI-TOF. SDS-PAGE was carried out using the NuPAGE Novex system (Invitrogen, Carlsbad, CA) with 10% Bis-Tris gels, 2-(*N*-morpholino)ethanesulfonic acid (MES)–SDS as running buffer, and SeeBlue Plus2 prestained molecular mass markers. Gels were stained with Coomassie SimplyBlue

SafeStain (Invitrogen). MALDI-TOF mass spectrometry was carried out in an Ultraflex mass spectrometer (Bruker, Billerica, MA). Protein samples were prepared by the dried droplet method. The matrix was made up of 2.5-dihydroxyacetophenone (5 mg mL^{-1}), diammonium hydrogen citrate (1.5 mg mL^{-1}), 25% (v/v) ethanol, and 1% (v/v) trifluoroacetic acid on an AnchorChip target (600 μm, Bruker). An external mass calibration was done based on Protein Calibration Standard II (Bruker).

5.2.6. Preparation of Protein-DNA Complexes. Protein-DNA complexes dissolved in 10mM sodium phosphate buffer, pH 7.4, were prepared by mixing of pipetted portions of DNA stock solution, protein stock solution and 100 mM phosphate buffer, pH 7.4, in Milli-Q water. Mixtures were vortexed for a 1 minute. Volumes of the mixed portions of stock DNA and protein solutions were varied according to their initial concentration and the desired final protein/DNA-bp (ptn/bp) ratio. Protein stock solution was prepared just before use by dissolving a weighted amount of lyophilized protein in Milli-Q water. DNA stock solution corresponded to the commercially acquired or in-house prepared DNA.[16]

5.2.7. Electrophoretic Mobility Shift Assay (EMSA). Aliquots of 2.64 kb pDNA (60 ng/uL) were mixed with different volumes of a C_8-B^{Sso7d} solution (0.05 o 0.2 g L^{-1}) both dissolved in 1x Tris–acetate–EDTA (TAE) buffer (pH 8) for a final volume of 10 μL and a pDNA concentration of 6 ng/uL. After incubation for 60 min at room temperature, the mixtures were mixed with 6x loading buffer and 8-6 uL of the final mixture were subjected to electrophoresis in an agarose gel (1%) for at least 60 min at 95 V using 1x TAE buffer (pH 8). Bands were visualized using ethidium bromide.

5.2.8. Light Scattering. Light scattering measurements were performed with a Zetasizer NanoZS apparatus (Malvern Instruments, UK) equipped with a 4 mW He–Ne ion laser at a wavelength of 633 nm. DNA–protein complexes were formed at room temperature (final DNA concentration was 50 mg L^{-1}) dissolved in sodium phosphate buffer (pH 7.4, 10 mm). Stock protein solution was previously filtered (0.2 μm filters) for all LS experiments. The samples were continuously measured

after preparation and only stable measurements were taken into consideration. The intensity of light scattered by the DNA–protein complexes was determined from an average of six autocorrelation measurements carried out in the course of 30s at 25 °C using a scattering angle of 12.8° and/or 173°. The effective hydrodynamic radius was automatically obtained from the apparent diffusion coefficient calculated by the apparatus.

5.2.9. Static Light Scattering. The determination of molecular weight of complexes was done using the absolute scattering of complexes of dsDNA 8.0 kbp at concentration of 50 ng/μL at a scattering angle of 173º. The molecular weight was calculated according to the following equation:

$$M_w = R_\Theta / C * K_R \qquad (5.1)$$

Where is molecular weight of the complexes, is the Rayleigh scattering, is concentration and is the scattering constant. The later were calculated according to the following equations:

$$R_\Theta = (I_{samples} - I_{solvent} / I_{toluene}) * (n^2_{solvent} / n^2_{toluene}) * R_{toluene} \qquad (5.2)$$

$$K_R = 4\pi^2 n^2_{solvent}(dn/dC)^2 / \lambda^4 N_A \qquad (5.3)$$

Where $n^2_{solvent}$ is the square refraction index of water, $n^2_{toluene}$ the square refraction index of toluene reference. $R_{toluene}$ is the absolute scattering of toluene reference. The change in the refraction index, (dn/dC), was fixed to 0.00017 m^^3 kg^^-1 as seen for other collagen-like proteins. λ is the wavelength of the laser used and N_A the Avogadro's number.

5.2.10. Circular Dichroism. CD measurements were performed on a Jasco J-715 spectropolarimeter at room temperature. The spectra were recorded between 190 and 260 nm with a resolution of 0.2 nm and a scanning speed of 1 nm/s. Each measurement was performed in a 1 mm quartz cuvette. The final concentration of C_8-B^{Sso7d} was 0.25 mg/mL dissolved in sodium phosphate (10 mM, pH 7.4). Control protein C_4 was also dissolved in same buffer at the same concentration. Linear

dsDNA 2.5 kbp was used for complexing sample at ptn/bp ratio of 0.188. Samples were incubated at least for one hour at room temperature. Molar ellipticity [θ] was calculated according to the next equation:

$$[\Theta] = \Theta \times 100 \times M / C \times l \times n \quad (5.4)$$

Where the ellipticity Θ is in degrees, M is molecular mass of the protein, C is concentration in mg/mL, l is optical path length in cm and n is number of residues of protein.

5.2.11. Atomic Force Microscopy. Approximately 3–5 μL DNA–protein-based polymer complex solution was added to clean hydrophilic silicon (1×1 cm) and left for 2 min. Then it was rinsed with filtered Milli-Q water (1 mL) to remove salts and non absorbed particles, followed by soaking up of excess water using a tissue and slow drying under a N_2 stream. Samples were analyzed using a Digital Instruments NanoScope V equipped with a silicon nitride probe (Bruker, MA, USA) with a spring constant of 0.4 Nm^{-1} in ScanAsyst™ imaging mode. Images were recorded with >0.965 Hz and 1024 samples per line. Image processing was done with NanoScope Analysis 1.20 software. Contour length and long axis length measurements were performed with ImageJ software.

5.2.12. Fluorescence Microscopy Imaging. Protein-coated single T4-DNA molecules were stained at room temperature with the intercalating fluorescent dye YOYO-1 in 10 mM Tris-HCl, pH 8.0 (intercalation ratio of one every 25bp). The samples were incubated at least 30 min and the final T4-DNA concentration was ~5 μg/mL. The fluorescent protein-DNA was imaged with a Nikon Eclipse Ti inverted fluorescence microscope equipped with a 200 W metal halide lamp, a filter set and a 100x oil immersion objective. The exposure time was controlled using an UV light shutter and the images were collected with an electron multiplying charge-coupled device (EMCCD) camera (Andor iXon X3). The diffusion coefficient and hydrodynamic radius in bulk of naked T4-DNA and coated with C_4-K_{12} (0.834 ptn/bp) and C_8B^{Sso7d} (0.5 ptn/bp) were calculated from processing videos of at least one-minute duration using previously reported algorithms in

MatLab.[23, 24]. The samples were incubated at least 30 min and the final T4-DNA concentration was ~5 μg/mL.

5.2.13. DNA Protection Test. pDNA 2.6 kbp (concentration of 19.7 ng/uL) was complexed with C_8-B^{Sso7d} (concentration 0.44 mg/mL) in TAE buffer (pH 8) for 1h at room temperature (0.188 ptn/bp). Then 1μL of the enzyme DNAse I (RNAse free , Thermo Scientific) 0.055 U was added to 35.5 μL complexes dissolved in reaction DNAse I buffer (100 mM Tris-HCl, pH 7.5, 25 mM $MgCl_2$, 1 mM $CaCl_2$) for a final [DNA] = 17.2 μg/mL and immediately incubated at 25ºC using a thermo block. Aliquots of 3.5 μL were taken at different times and mixed with 3.5 μL of EDTA 50 mM. After addition of loading buffer (6x) the sample was electrophoresed in agarose gel 1% at 100V for 45 min. DNA bands were visualized using ethidium bromide.

5.2.14. Hemolytic Activity of Polyplexes. Experiments were performed with red blood cells (RBCs) of blood group 0, Rhesus-positive donors that had been collected and processed according to standard Dutch blood bank protocols, including leukoreduction and storage in saline-adenine-glucose-mannitol. RBCs were washed and resuspended in complete Ringer solution (125 mM NaCl, 5 mM KCl, 1 mM $MgSO_4$, 32 mM HEPES, 5 mM glucose and 1 mM $CaCl_2$). All experiments with RBCs were performed at room temperature. In an eppendorf tube, 3 x 10^7 RBCs were suspended in 300 μL of Ringer buffer and complexes were added to the cell suspension with a total protein concentration of 20 μM. After 2 hours of incubation, RBCs were centrifuged (3000 RPM for 3 minutes) and the supernatant (containing the free hemoglobin) was collected. Non-treated cells were used as a negative control, while burst or hemolysis of RBC's induced by incubation in Milli-Q water was used as a positive control. An aliquot of 60 μl of the collected supernatant was transferred to a 96-well plate and diluted to 300 μl with water. The absorbance of each sample was subsequently measured at 405 nm.

The degree of hemolysis was calculated as follows:

$$\% \, Hemolysis = (Abs_{sample} - Abs_{negative\ control} / Abs_{positive\ control}) \, x \, 100 \qquad (5.5)$$

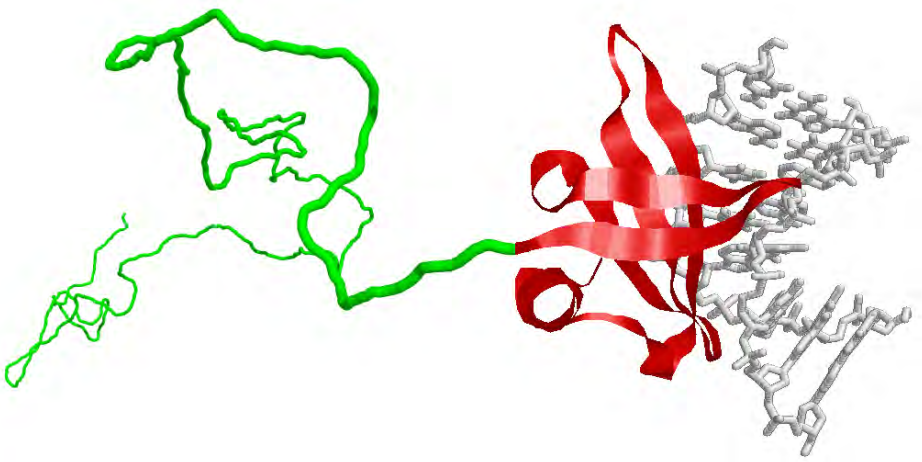
5.2.15. Acute Toxicity Induced by Polyplexes Towards Host Cells. The impact of complexes on HeLa cell viability was assessed using the resazurin assay. Briefly, 8 x 10^3 cells/well were seeded in a 96 well plate. After 24 hours, complexes were added to a protein concentration of 1, 2, 10 and 20 µM for an incubation of 2 hours. Cells were washed and resazurin (100 µg/ml in RPMI-1640 culture medium supplemented with 10% FCS was added. Non-treated cells were used as a negative control, culture medium was used as a blank. After 4 hours, readings of fluorescence intensity (Ex: 540/25, Em: 620/40) were taken on a BioTek Synergy 2 plate reader. Cell viability was calculated as follows:

$$\%\ Cell\ viability = fluorescence_{sample} - fluorescence_{blanck} / fluorescence_{control} - fluorescence_{blank} \times 100 \quad (5.6)$$

5.2.16. Transfection. HeLa cells were grown as monolayers in RPMI-1640 medium, supplemented with 10% FCS and 200 mM L-glutamine, at 37°C, in a humidified atmosphere containing 5% CO_2. Cells were passaged after 2 days, or when at 80-90% sub-confluency. C_8B^{Sso7d} and C_4B^{K12} were incubated during 1 hour with pDNA coding for yellow fluorescent protein (YFP-pDNA) at 0.75 ptn/bp and 1 ptn/bp, respectively, to form the protein-DNA complexes in water. Then, 60 ng of pDNA (complexed with protein at a final protein concentration of 10 µM) were incubated during 2 hours with HeLa cells in a 24-well plate ($4x10^5$ cells/well/ml) in RPMI-1640 medium supplemented with 10% FCS. Equivalent concentrations of free YFP-pDNA and of YFP-pDNA complexed with PEI (at an amino-to-phosphate ratio of 6:1), or with Lipofectamine 2000 were used in control experiments. After incubation with the YFP-pDNA-protein complexes, the cells were washed twice with HBS, and fresh culture medium was added. Synthesis of the YFP was assessed after 2 days by flow cytometry, using a Becton Dickinson FACScalibur flow cytometer and fluorescein detector. Data were analyzed using Weasel software (WEHI, version 30114win).

5.3 Results and Discussion

5.3.1. Design and Production. The sequence of the C_8-B^{Sso7d} protein-based polymer as produced and a cartoon of the structure of the protein-based polymer bound to DNA is shown in Figure 5.1, where we have used the published X-ray crystallographic structure of B^{Sso7d} bound to short dsDNA.[16] Note the large asymmetry of the polymer in terms of the domain lengths (63 amino acids for the B^{Sso7d} binding domain and 797 amino acids for the hydrophilic C_8 domain). The 860 amino acid polypeptide has a predicted molar mass of 80,372 Da. The C_8-B^{Sso7d} protein based polymer was successfully produced by secreted recombinant fermentation in Pichia pastoris at an approximate yield (purified protein to volume of filtered, cell-free medium) of 0.72 g/L. The production of the DNA binding domain link to an unstructured polypeptide was well tolerated by the P. Pastoris cells suggesting that production of other large engineered protein-based polymers is attainable. The reason could come from a possible protection provided by the large, unstructured and very hydrophilic C_8 domain. Bulk purification using ammonium sulfate precipitation was sufficient to separate the protein polymers from most other proteins secreted in the extracellular medium, as shown by SDS-PAGE in Figure 5.2a. Multimers of the hydrophilic C-domain are known to poorly bind SDS and hence move with anomalously low speeds in SDS-PAGE.[25] Indeed, the apparent molar mass of the purified polypeptides as deduced from SDS-PAGE would be around 190kDa. The molar mass of the purified proteins was analyzed by MALDI-TOF mass spectrometry. The spectrum shown in Figure 5.2b shows three peaks that correspond the M^+ (80,210 Da), M^{2+} (40,190 Da) and M^{3+} (26,645 Da) ions. The molar mass of the protein estimated from the peak positions is 80.1±0.2kDa, which agrees with the predicted molar mass of 80,372Da to within the experimental accuracy. Furthermore, the existence of the protein in a monomer state in solutions of the purified biosynthesized protein was corroborated by DLS by detection of a single molecule population with a hydrodynamic radius of 7.92 nm, which is in accord with the predicted value for a protein polymer of this size.[26]



GPP(GEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQGN PGKNGQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPA)$_8$GGATVKFKYKGEEKEV DISKIKKVWRVGKMISFTYDEGGGKTGRGAVSEKDAPKELLQMLEKQKK*

Figure 5.1. - Schematic representation of the structure of the protein-based polymer C_8-B^{Sso7d} bound to double stranded DNA (dsDNA). In green: the hydrophilic unstructured C_8 polypeptide, in red: the B^{Sso7d} DNA-binding domain, in grey: (dsDNA). The structure of B^{Sso7d} bound to (dsDNA) is taken from the X-ray crystallographic structure of B^{Sso7d},[16] (pdb: 1BNZ). Below is shown the amino acid sequence of C_8-B^{Sso7d} protein based polymer. In green: the hydrophilic unstructured C_8 polypeptide, in red: the B^{Sso7d} DNA-binding domain.

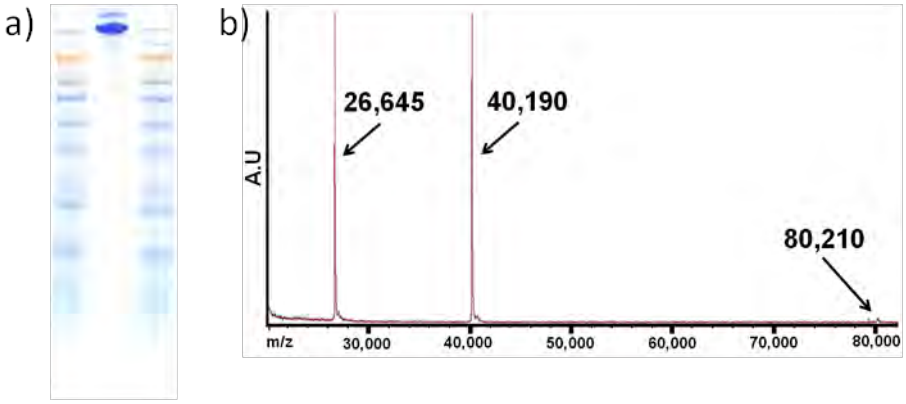


Figure 5.2. - Molecular characterization of purified C_8-B^{Sso7d}. a) SDS-PAGE. b) MALDI-TOF showing three peaks corresponding to M^+ (80,210 Da), M^{2+} (40,190 Da) and M^{3+} (26,645 Da).

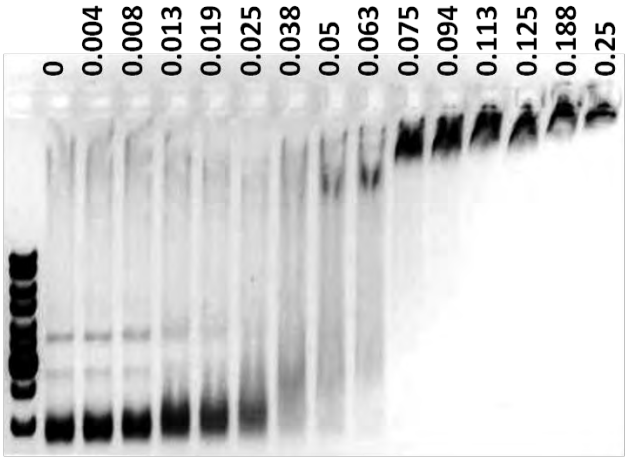


Figure 5.3. Electrophoretic Mobility Shift Assay for 2,6kb pDNA complexed with C_8-B^{Sso7d}. Each lane has ~40 ng of DNA. Protein/DNA bp ratio is shown at the top. Left lane shows the molecular markers.

5.3.2. DNA Binding Properties* of *C_8-B^{Sso7d} Protein-based Polymer. In order to probe the DNA binding properties of C_8-B^{Sso7d} we studied the mobility shift of complexes of C_8-B^{Sso7d} with a 2.6kbp pDNA in agarose gel. As is clear from Figure 5.3, the mobility of complexes starts to be reduced at ratio of 0.013 proteins per base pair (ptn/bp), whereas the mobility shift more or less saturates at around 0.25 ptn/bp. Even higher protein concentrations do not lead to a further reduction of the mobility (see Appendix Figure 5.11). The latter effect was also found by light scattering studies (see later). Structural and crystallographic studies report that B^{Sso7d} binds to dsDNA every 4bp or 0.25 ptn/bp.[16, 27] Here we found that dsDNA is indeed fully saturated at those values. Hence, the long C_8 domain does not seem to lead to a much reduced binding stoichiometry due possibly to the high effective binding affinity and specificity that B^{Sso7d} has for dsDNA (>0.25 ptn/bp). In comparison, using an unspecific cationic binding domain required to add a large excess of protein (> 0.667 ptn/bp) to reach a coating degree of ~80%.[13] The reduction of unspecific binding between the colloidal domain (net charge of -10 at pH of 7.4) and the binding domain could also play an important role.

Table 5.1. Absolute Molar mass of C_8-B^{Sso7d} /DNA complexes and binding stoichiometry estimated from Static Light Scattering.

Experimental ptn/bp ratio	MW x10^7 (g/mol)	Aggregation Number	Effective ptn/bp ratio
0.06	3.57	383	0.05
0.12	8.44	990	0.13
0.25	19.9	2414	0.30
0.25	22.2	2707	0.34

Next, absolute molar masses of C_8-B^{Sso7d} DNA complexes were estimated using static light scattering at a fixed low scattering angle. From these scattering intensities, absolute molar masses of complexes can be deduced. Strictly speaking the molar mass should be calculated from the scattering extrapolated to zero scattering angle and zero concentration. We here report an estimate of the solution molar mass by instead using the scattering intensity at a single low scattering angle and at very low concentrations of complexes. Also, we assume that scattering is dominated by the protein-DNA complexes and that the contribution of the excess free proteins to the light scattering is negligible. This leads to the estimates shown in Table 5.1, which confirm again that at saturation, the density of proteins is about 0.25 protein per bp.

5.3.3. Secondary Structure of B^{Sso7d} Domain in the Context of C_8-B^{Sso7d} Polymer. The highly specific binding by B^{Sso7d} domain towards dsDNA with high binding affinity relies completely in the presence of a folded state. To check that the B^{Sso7d} domain is properly folded in the context of C_8-B^{Sso7d} polymer, circular dichroism spectroscopy was done (figure 5.4). Molar ellipticity from solutions of C_8-B^{Sso7d} was compared with that of a reference tetramer C_4. C_8-B^{Sso7d} and C_4 both show a deep minimum at 200nm, characteristic associated to a random coil secondary structure (figure 5.4a). This is to be expected since the random coil C_8 dominates the CD spectrum in view of its much larger mass.[25, 28] The difference spectrum should correspond to the spectrum of the B^{Sso7d} domain. Since the mass of the C_8 domain is

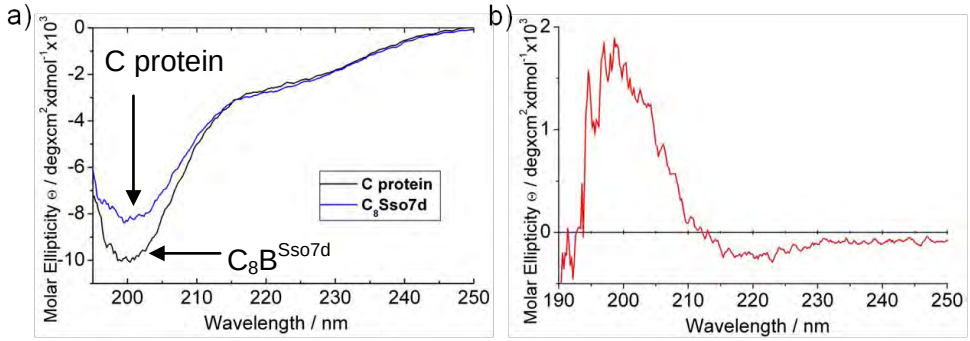


Figure 5.4. Possible influence of C_8 domain on secondary structure of B^{Sso7d} domain**.** a) CD spectrum of C_8-B^{Sso7d} and C_4 control. b) Spectrum pertaining to B^{Sso7d} domain alone obtained by subtraction of C_8 domain ellipticity.

more than 10 times larger than that of the B^{Sso7d}, we cannot determine this very accurately. The difference spectrum for C_8-B^{Sso7d} bound to DNA and the C_4 reference polymer is shown in Figure 5.4b. Despite the noise, it is clear that the difference spectrum has a positive ellipticity at 200 nm and a slightly negative ellipticity at the 220 nm, which is related to the presence of five β-sheet strands[16] and is very similar to a previously published CD spectrum for B^{Sso7d} in solution.[29, 30] Therefore, not only the B^{Sso7d} DNA binding properties are quite insensitive to the attachment of the very large C_8 domain, but also its secondary structure appears not to be affected very strongly. Indeed the B^{Sso7d} domain is very stable even to truncation of 8 residues at the C-terminus.[30]

5.3.4. AFM Imaging of Complexes of C_8-B^{Sso7d} with Linear and Supercoiled DNA. Controlled coating of individual DNA molecules is a pre-requisite for certain applications that require modulation of physical properties of DNA molecules. Then, next we investigated the structures of complexes of C_8-B^{Sso7d} polymers with both linearized and supercoiled DNA. To highlight the differences introduced by using a DNA -specific binding domain as compared to a simple cationic oligolysine binding block, we compare with complexes formed with the C_4-B^{K12} diblock polymer that we have studied previously. AFM images of complexes of C_8-B^{Sso7d} with single molecules of 8kb linear double stranded DNA are qualitatively similar

to those we have previously published for the C_4-B^{K12} diblock: monodisperse with regular and homogeneous length and height. Upon increasing the protein to DNA ratio, the DNA molecules get progressively coated, without any indication of intramolecular aggregation as shown in the representative images of Figure 5.5a-c. It is possible to recognize the brush-like nature of the complexes in detail images (figure 5.5d). At a more quantitative level, it appears that the C_8-B^{Sso7d} complexes are somewhat higher than the C_4-B^{K12} complexes (Table 5.2). This most likely reflects both the larger and globular binding domain, and the longer side chains. Quantitative analysis also reveals that contour lengths of C_8-B^{Sso7d} complexes are significantly shorter than that of the bare DNA (Table 5.2). It has been reported that Ss07d exerts kinks that reduce the effective contour length of dsDNA complexes in between 10-20%.[31]

For 4kb supercoiled pDNA, complexation with C_8-B^{Sso7d} seems to lead to global opening up of plectonemic supercoils (Figure 5.6). The complexes have a contour length of 1,395 ± 29 nm, which is just slightly larger than the theoretical contour length of the bare pDNA of 1360nm. B^{Sso7d} binding has been reported to lead to DNA unwinding by placing its triple beta-sheet domain across the minor groove.[19, 32] At a fixed linking number[33] this will be compensated by the introduction of positive supercoiling that balances part or the entire original negative supercoiling. The clear shortening of the apparent contour length of complexes that was observed for linear DNA, is absent in the case of supercoiled pDNA. In contrast to complexation with C_8-B^{Sso7d}, the C_4-B^{K12} diblock leads to a tightening of plectonemic supercoils, possibly via bridging of the opposing sides of a plectonemic supercoil by the binding domain (Figure 5.6b). As a consequence, single thick contours are observed for C_4-B^{K12} pDNA complexes, that appear as flexible rod-like structures with a contour length that is close to half that of the bare contour length of the pDNA, as we have also reported before.[13] For more images of the complexes of pDNA with both proteins, we refer to the Appendix Figure 5.12.

5.3.5. AFM Imaging of Complexes of C_8-B^{Sso7d} with Single-stranded DNA. We have studied complexes of both C_8-B^{Sso7d} and C_4-B^{K12} with circular ssDNA from

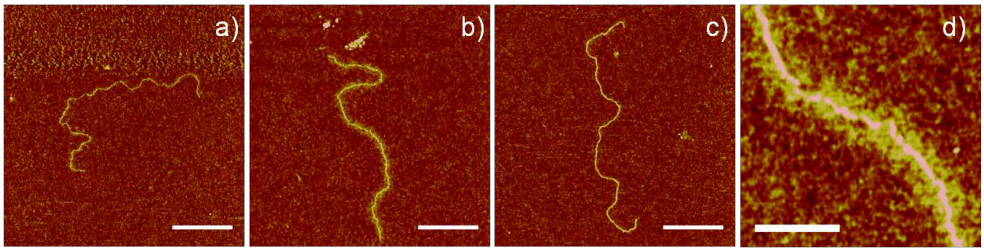


Figure 5.5. Coating single molecules of linear dsDNA with C_8-B^{Sso7d}. Atomic force microscopy images of protein complexes formed with 8.0 kbp DNA at different protein per bp ratios. (a) 0.062 ptn/bp, (b) 0.125 ptn/bp, (c) 0.25 ptn/bp. Scale bars are 500nm. (d) Bottle brush can be appreciated in zoomed in section of (b) image. Scale bar is 150 nm.

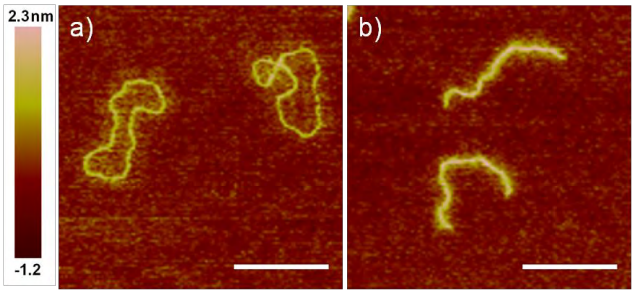


Figure 5.6. Opening up- and tightening plectonemically supercoiled pDNA single molecules of by coating them with a) C_8-B^{Sso7d} and b) C_4-B^{K12}. Atomic force microscopy images of protein complexes of pDNA 4.0 kbp formed with C_8-B^{Sso7d} (0.375 ptn/nt) and b) C_4-B^{K12} at 0.833 ptn/bp. Scale bars are 400 nm.

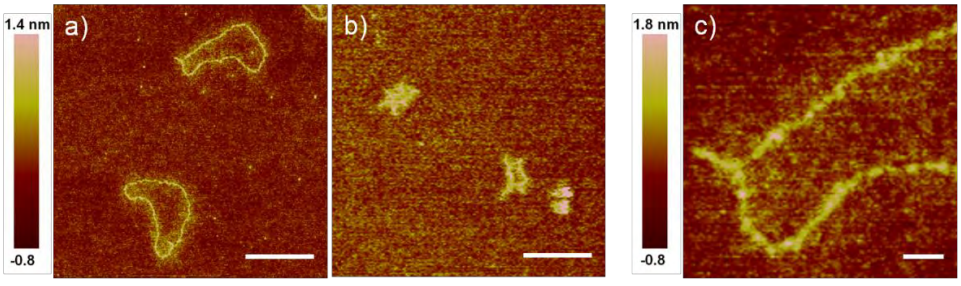


Figure 5.7. Disentangling and stretching single circular ssDNA molecules by coating them with C_8-B^{Sso7d}. Atomic force microscopy images of complexes formed between circular ssDNA from M13mp18 virus and a) C_8-B^{Sso7d} (1.45 ptn/bp) and b) C_4-B^{K12} at 0.792 ptn/bp. Scale bars are 500nm. c) Zoomed section of C_8-B^{Sso7d} + ssDNA complex from (a). Scale bar is 50 nm.

Table 5.2. Height and contour lengths of protein coated 8kbp dsDNA.

	C_8-B^{Sso7d}						C_4-B^{K12}
Proteins / base pair	0	0.06	0.13	0.19	0.25	0.49	0.26
Height (nm)	ND	0.61±0.19	0.94±0.18	1.09±0.29	1.22±0.26	1.21±0.2	0.95±0.19
Contour length (nm)	2,720*	ND	2,382±83	ND	2,551±65	2,439±36	2,720±32

*Theoretical naked contour length, ND: not determined

M13mp18 virus (7,249 nt) since Sso7d is reported to also bind to ssDNA.[34] Stretching ssDNAs could have several technological implications for high resolution of genomic studies.[9] The protein-ssDNA complexes formed again have distinctly different complex morphologies for the two proteins (figure 5.7). We find that the C_8-B^{Sso7d} protein nearly completely prevents the formation of secondary structures due to base pairing of ssDNA and therefore disentangles and stretch the structure (figure 5.7a). In most images, a single short piece of an apparent duplex segment remains that could correspond to a local (nearly) palindrome sequence with particularly strong base pairing. The C_8-B^{Sso7d} + ssDNA complexes have a contour length of 1,326 ± 95 nm, which is 53.8% of the calculated contour length of naked ssDNA (2,465 nm, assuming a length of 0.34 nm per nt). In clear contrast, the formation of secondary structures by base pairing of the circular single stranded M13mp18 virus DNA is not prevented when it is complexed with the diblock C_4-B^{K12} resulting in complexes with a collapsed and branched appearance (Figure 5.7b). Interestingly, zooming in the C_8-B^{Sso7d} + ssDNA complexes show a beaded appearance that is reminiscent of the "pearl-necklace" configurations found for complexes with flexible polyanions and C_4-B^{K12}.[35] The degree of stretching of ssDNA reached by C_8-B^{Sso7d} and not with C_4-B^{K12} supports the idea that using a natural DNA-binding domain that avoid intramolecular bridging offer advantages to stretch single DNA molecules which could be used effectively for genomic mapping purposes.[9]

5.3.6. Increase of Coil Sizes of T4-DNA upon Binding C_8-B^{Sso7d}: Bottle-brush Stiffening. The formation of a protein bottle-brush around the coated nucleic acid

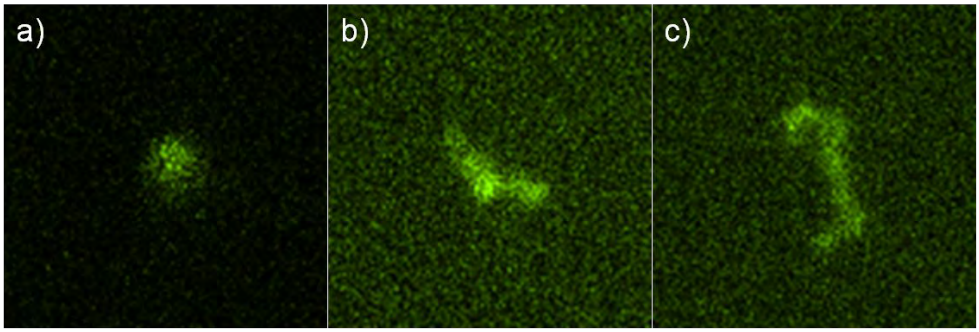


Figure 5.8. Increase of coil sizes of T4-DNA upon binding C_8-B^{Sso7d}. Snapshots taken from fluorescence microscope videos of a) naked T4-DNA, b) coated with C_4-B^{K12} and c) coated with C_8-B^{Sso7d}.

Table 5.3. Bulk diffusion parameters of protein coated T4-DNA.

	T4-DNA	T4-DNA + C_4-B^{K12}	T4-DNA + C_8-B^{Sso7d}
Diffusion coefficient ($x10^{-13}$, m^2/s)	3.50 ± ($9.0x10^{-2}$)	2.42 ± ($4.3x10^{-2}$)	1.73 ± ($7.1x10^{-2}$)
Hydrodynamic radius (μm)	0.696 ± ($0.018x10^{-2}$)	1.007 ± ($1.8x10^{-2}$)	1.411 ± ($0.058x10^{-2}$)

would have an increasing effect in the stiffening of the coated-DNA. To probe this effect we intercalate fluorescent YOYO-1 in T4-DNA to image and track single molecules when coated with C_8-B^{Sso7d} and C_4-B^{K12} (figure 5.8). With based on those images we quantified the diffusion constants and hydrodynamic sizes of the complexes when diffusing in bulk to compare the protein-coating effects. The diffusion coefficients and hydrodynamic radius (R_H) of T4-DNA naked and coated with C_8-B^{Sso7d} and C_4-B^{K12} is shown in table 5.3. When T4-DNA is coated with C_8-B^{Sso7d}, DNA molecule looks extended and its contour can be distinguish clearly (figure 5.8a). In the videos it looks that diffuse slower. C_4-B^{K12} has similar effect but less pronounced (figure 5.8b). Naked T4-DNA looks more compact (figure 5.8c) than when protein coated, rotates more and diffuses faster. Calculated diffusion coefficients for protein coated T4-DNA are lower than when is naked, and being

Figure 5.9. - Enzymatic protection assay of DNA complexed with C_8-B^{Sso7d}. DNA-protein complexes (0.188 ptn/bp) were incubated at different times with DNAse I. Aliquots of reaction samples stopped with EDTA were ran in an agarose gel to detect the degradation of DNA (stained with EtBr). Top numbers are incubation time in minutes.

with C_8-B^{Sso7d} the lowest. This is reflected in a two-folded increase in R_H of the complexes with C_8-B^{Sso7d} in comparison with the naked T4-DNA. The increase in R_H indicates that the protein is stretching the single T4-DNA molecule upon forming the coat. The C_8-B^{Sso7d} + T4-DNA appeared more rigid and larger in comparison to naked T4-DNA and when coated with C_4-B^{K12}. In all the cases no aggregation was detected. The effect when coating T4-DNA molecules with C_8-B^{Sso7d} is stronger than with C_4-B^{K12}. This has implications for example in resolving fluorescent markers along the DNA chain when is stretched. Further work to calculate the increments in persistence length both in bulk and inside nanochannels are required. However, from previous estimations DNA coated with C_4-B^{K12} when inserted in nanochannels had an effective persistence length of 200 nm, which suggests that coating DNA with C_8-B^{Sso7d} should be higher and with promising applications in genomic mapping.

5.3.7. Enzymatic Accessibility of Coated DNA. Accessibility to DNA coated by molecules such as proteins or nucleotides is relevant for diverse possible applications. For non-viral gene delivery is required to have a compromise balance

between protection and ease of unpacking. Also, for optical mapping is required that DNA gets in contact and hybridizes with fluorescent markers. As a way to estimate the DNA accessibility, we proceed to evaluate the ability of the C_8-B^{Sso7d} protein to decrease the enzymatic accessibility to the DNA. Solutions containing DNA- C_8-B^{Sso7d} complexes were incubated with unspecific nuclease DNAseI. At different incubation times aliquots of the sample were taken and run in an agarose gel (Figure 5.9). The band corresponding to the protein-DNA complexes was noticeable for about 10 minutes of incubation with the enzyme, whereas naked DNA was completely digested after 2 minutes of incubation with the enzyme. This could be associated to the fact that the protein forms a large brush around the DNA and that the B^{Sso7d} binds to DNA with high affinity, reducing the exchange ratio with surrounding buffer.

5.3.8. Non-viral Gene Delivery: Toxicity, Biocompatibility and Transfection Efficiency. In order to evaluate the potential application of C_8-B^{Sso7d} and C_4-B^{K12} as DNA nanocarriers we proceed to test in first instance their safety profiles. We measure cell viability and cell hemolysis at very high protein concentrations (Figure 5.10). For both proteins, the effect in the cell viability followed a very similar trend. It was reduced in about 15% at concentration of 1μM and slightly decreased to 27.6% for 20 μM in comparison to untreated cells. Hemolysis was reduced 5% for C_8-B^{Sso7d} while for C_4-B^{K12} was 6.9%. The slightly higher hemolysis caused by C_4-B^{K12} may be related to the presence of a cationic segment involved in the DNA binding. In general these results indicate that proteins present low hemolysis and acute toxicity. Biocompatibility and a good safety profile are obligated requisites for biomedical application of the proteins.

Finally, transfection studies of HeLa cells were carried out. The measurement of fluorescence by the expression of a reporter protein was correlated with the delivery into the nucleus of the pDNA cargo encoding for that reporter protein.

The transfection capabilities of C_8-B^{Sso7d} and C_4-B^{K12} are reported in figure 5.10c. Naked pDNA used as control and commercial standards such as PEI and Lipofectamine 2000 are also included. It can be appreciated that both proteins, C_8-

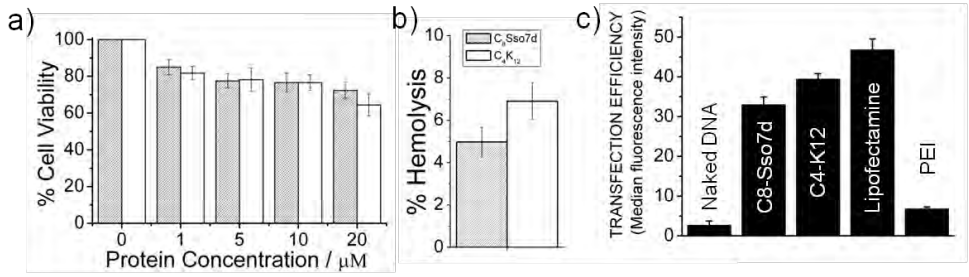


Figure 5.10. Toxicity and transfection activity of protein C_8B^{Sso7d} complexing DNA compared with protein C_4B^{K12} complexes and standard gene-delivery agents. (a) Effects in cell acute toxicity of protein-DNA complexes with increasing protein concentrations. (b) Effects in hemolysis of protein-DNA complexes with protein concentration of 20 μM. (c) Gene delivery efficiency of C_8B^{Sso7d}, C_4B^{K12} and standard gene-delivery agents reported as measured fluorescence by the YFP encoded in the delivered pDNA.

B^{Sso7d} and C_4-B^{K12}, performed much better than naked DNA and the commercial standard for polymer-based non-viral gene delivery nanocarrier. Both proteins performed almost with similar levels to Lipofectamine 2000 and statistically they did not have any significant difference with it. It is interesting to notice that our designed protein had gene delivery efficiency superior than PEI and similar to Lipofectamine 2000 without any optimization process. The transfection capacity could come from a proper balance between the protection and stability given by the protein coating and detachment and release of the pDNA cargo inside the cell. This suggests that both proteins C_8-B^{Sso7d} and C_4-B^{K12} could be used to build "artificial viral" nanocarriers for gene delivery.[36] The positive results prompt us to evaluate the mechanistic details of the transfection capacity for further improvements of the protein design in order to use them as nanocarriers for gene and nucleic acid delivery.

5.4 Concluding Remarks

The study shows that we can harness the high specificity and affinity of natural occurring DNA binding domains for engineering protein-based polymers with controlled binding and coating of single DNA molecules. Its combination with a

long unstructured domain is suitable not only for imaging[5] but also for modulating their stretching and delivery. B^{Sso7d} DNA-binding domain presents numerous advantages such as facilitated formation of stable protein-DNA complexes with lower protein concentrations. It also reduces intra-molecular bridging in ssDNA and supercoiled pDNA. As the DNA binding is not based on electrostatics is expected to decrease undesired molecular cross-talk with charged surfaces or other negative polyelectrolytes. All this contrasts largely with unstructured and unspecific cationic segments. Besides this, B^{Sso7d} has convenient physical chemical properties such as high thermal (up to 90°C), acid and chemical stability that can find a place in practical implications for example to increase thermal stability of protein-DNA complex.[16] It is also possible to choose other DNA-binding domains with particular sequence binding specificity, stoichiometry, and affinity from the large repertoire of different characterized binding domains that exist.[37, 38]

Protein C_8-B^{Sso7d} could be used to give multivalency activity to DNA coated complexes and achieve by this way molecular recognition with high affinity in biological contexts.[39] This would open the doors to functionalize in a controlled manner DNA-based structures such as origami and nano-robots for delivery based on DNA.[1, 3, 4] Furthermore, it would be possible to engineer a protein with high affinity for specific nucleic acid sequences to locate functional domains or particles conjugated to the proteins at particular places of the DNA. Moreover, modulating physical configurations and interactions of DNA by coating them with binders gives the opportunity to design artificial transcription factors for regulating genetic circuits in cells which is relevant for synthetic biology.[40, 41]

In conclusion, we have designed and produced by recombinant means an artificial highly asymmetric protein-based polymer made up of a properly folded natural occurring DNA binding domain (B^{Sso7d}) linked to a large colloidal domain (C_8). The protein C_8-B^{Sso7d} binds with high affinity and high specificity in a predicted stoichiometry to DNA, in opposite to proteins carrying a cationic segment as DNA binding domain. C_8-B^{Sso7d} fully coats single DNA molecules of different sizes (up to several dozens of kbp) and kinds (linear dsDNA, circular

ssDNA, supercoil pDNA), reduces largely inter- and intramolecular bridging of DNAs and forms stable complexes. This leads to superior disentangling and stretching of the DNA molecule, particularly demonstrated for ssDNA and large T4-DNA, in opposite to our previously designed protein C_4-B^{K12}. Therefore C_8-B^{Sso7d} has a huge potential to be harnessed as adjuvant biomaterial for genetic mapping and sequencing technologies of DNA. Furthermore, C_8-B^{Sso7d} presented an acceptable safety profile and significantly transfected cells. It is therefore a promising biomaterial to be used as nanocarrier in gene delivery. Incorporating natural occurring DNA binding domains is an alternative to polycationic segments that lack of high specificity and affinity for DNA. Due to its recombinant origin C_8-B^{Sso7d} can be further optimized by tuning precisely the length of the domains and by incorporation of bioactive domains. In general, developing protein polymer biomaterials that offer precise control of the physical chemical properties and interactions of single DNA molecules can pave the way for the advance of DNA-related nanotechnology, gene delivery and synthetic biology.

5.5 References

1. Andersen, E. S.; Dong, M.; Nielsen, M. M.; Jahn, K.; Subramani, R.; Mamdouh, W.; Golas, M. M.; Sander, B.; Stark, H.; Oliveira, C. L. P.; Pedersen, J. S.; Birkedal, V.; Besenbacher, F.; Gothelf, K. V.; Kjems, J., Self-assembly of a nanoscale DNA box with a controllable lid. *Nature* 2009, *459*, 73-U75.

2. Lee, H.; Lytton-Jean, A. K. R.; Chen, Y.; Love, K. T.; Park, A. I.; Karagiannis, E. D.; Sehgal, A.; Querbes, W.; Zurenko, C. S.; Jayaraman, M.; Peng, C. G.; Charisse, K.; Borodovsky, A.; Manoharan, M.; Donahoe, J. S.; Truelove, J.; Nahrendorf, M.; Langer, R.; Anderson, D. G., Molecularly self-assembled nucleic acid nanoparticles for targeted in vivo siRNA delivery. *Nat. Nanotechnol.* 2012, *7*, 389-393.

3. Douglas, S. M.; Dietz, H.; Liedl, T.; Hogberg, B.; Graf, F.; Shih, W. M., Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* 2009, *459*, 414-418.

4. LaBean, T. H.; Li, H. Y., Constructing novel materials with DNA. *Nano Today* 2007, *2*, 26-35.

5. Hamon, L. c.; PastrÃ©, D.; Dupaigne, P.; Breton, C. L.; Cam, E. L.; PiÃ©trement, O., High-resolution AFM imaging of single-stranded DNA-binding (SSB) protein-DNA complexes. *Nucleic Acids Res.* 2007, *35*, e58.

6. Bensimon, A.; Simon, A.; Chiffaudel, A.; Croquette, V.; Heslot, F.; Bensimon, D., Alignment and sensitive detection of DNA by a moving interface. *Science* 1994, *265*, 2096-2098.

7. Lam, E. T.; Hastie, A.; Lin, C.; Ehrlich, D.; Das, S. K.; Austin, M. D.; Deshpande, P.; Cao, H.; Nagarajan, N.; Xiao, M.; Kwok, P. Y., Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nat. Biotechnol.* 2012, *30*, 771-776.

8. Zhang, C.; Hernandez-Garcia, A.; Jiang, K.; Gong, Z.; Guttula, D.; Ng, S. Y.; Malar, P. P.; van Kan, J. A.; Dai, L.; Doyle, P. S.; Vries, R. d.; van der Maarel, J. R. C., Amplified stretch of bottlebrush-coated DNA in nanofluidic channels. *Nucleic Acids Res.* 2013.

9. Michalet, X.; Ekong, R.; Fougerousse, F.; Rousseaux, S.; Schurra, C.; Hornigold, N.; vanSlegtenhorst, M.; Wolfe, J.; Povey, S.; Beckmann, J. S.; Bensimon, A., Dynamic molecular combing: Stretching the whole human genome for high-resolution studies. *Science* 1997, *277*, 1518-1523.

10. Hall, A. R.; van Dorp, S.; Lemay, S. G.; Dekker, C., Electrophoretic Force on a Protein-Coated DNA Molecule in a Solid-State Nanopore. *Nano Lett.* 2009, *9*, 4441-4445.

11. Kumar, H.; Lansac, Y.; Glaser, M. A.; Maiti, P. K., Biopolymers in nanopores: challenges and opportunities. *Soft Matter* 2011, *7*, 5898-5907.

12. Yeo, L. Y.; Chang, H. C.; Chan, P. P. Y.; Friend, J. R., Microfluidic Devices for Bioapplications. *Small* 2011, *7*, 12-48.

13. Hernandez-Garcia, A.; Werten, M. W. T.; Stuart, M. C.; de Wolf, F. A.; de Vries, R., Coating of Single DNA Molecules by Genetically Engineered Protein Diblock Copolymers. *Small* 2012, *8*, 3491-3501.

14. Baumann, H.; Knapp, S.; Lundback, T.; Ladenstein, R.; Hard, T., Solution structure and DNA-binding properties of a thermostable protein from the archaeon Sulfolobus solfataricus. *Nat. Struct. Biol.* 1994, *1*, 808-819.

15. Baumann, H.; Knapp, S.; Karshikoff, A.; Ladenstein, R.; HÃ¤rd, T., DNA-binding Surface of the Sso7d Protein from Sulfolobus solfataricus. *J. Mol. Biol.* 1995, *247*, 840-846.

16. Gao, Y. G.; Su, S. Y.; Robinson, H.; Padmanabhan, S.; Lim, L.; McCrary, B. S.; Edmondson, S. P.; Shriver, J. W.; Wang, A. H. J., The crystal structure of the hyperthermophile chromosomal protein Sso7d bound to DNA. *Nat. Struct. Biol.* 1998, *5*, 782-786.

17. Edmondson, S. P.; Shriver, J. W., DNA-binding proteins Sac7d and Sso7d from Sulfolobus. *Methods Enzymol.* 2001, *334*, 129-145.

18. Consonni, R.; Arosio, I.; Belloni, B.; Fogolari, F.; Fusi, P.; Shehi, E.; Zetta, L., Investigations of Sso7d catalytic residues by NMR titration shifts and electrostatic calculations. *Biochemistry* 2003, *42*, 1421-1429.

19. Agback, P.; Baumann, H.; Knapp, S.; Ladenstein, R.; Hard, T., Architecture of nonspecific protein-DNA interactions in the Sso7d-DNA complex. *Nat. Struct. Biol.* 1998, *5*, 579-584.

20. Wang, Y.; Prosen, D. E.; Mei, L.; Sullivan, J. C.; Finney, M.; Vander Horn, P. B., A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. *Nucleic Acids Res.* 2004, *32*, 1197-1207.

21. Gera, N.; Hill, A. B.; White, D. P.; Carbonell, R. G.; Rao, B. M., Design of pH Sensitive Binding Proteins from the Hyperthermophilic Sso7d Scaffold. *PLoS One* 2012, *7*, 14.

22. Gera, N.; Hussain, M.; Wright, R. C.; Rao, B. M., Highly Stable Binding Proteins Derived from the Hyperthermophilic Sso7d Scaffold. *J. Mol. Biol.* 2011, *409*, 601-616.

23. Kundukad, B.; van der Maarel, J. R. C., Control of the Flow Properties of DNA by Topoisomerase II and Its Targeting Inhibitor. *Biophys. J.* 2010, *99*, 1906-1915.

24. Zhu, X. Y.; Kundukad, B.; van der Maarel, J. R. C., Viscoelasticity of entangled lambda-phage DNA solutions. *J. Chem. Phys.* 2008, *129*, 6.

25. Werten, M. W. T.; Wisselink, W. H.; Jansen-van den Bosch, T. J.; de Bruin, E. C.; de Wolf, F. A., Secreted production of a custom-designed, highly hydrophilic gelatin in Pichia pastoris. *Protein Engineering* 2001, *14*, 447-454.

26. Schor, M.; Bolhuis, P. G., The self-assembly mechanism of fibril-forming silk-based block copolymers. *Phys. Chem. Chem. Phys.* 2011, *13*, 10457-10467.

27. Robinson, H.; Gao, Y. G.; McCrary, B. S.; Edmondson, S. P.; Shriver, J. W.; Wang, A. H. J., The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. *Nature* 1998, *392*, 202-205.

28. Beun, L. H.; Beaudoux, X. J.; Kleijn, J. M.; de Wolf, F. A.; Stuart, M. A. C., Self-Assembly of Silk-Collagen-like Triblock Copolymers Resembles a Supramolecular Living Polymerization. *ACS Nano* 2012, *6*, 133-140.

29. Knapp, S.; Karshikoff, A.; Berndt, K. D.; Christova, P.; Atanasov, B.; Ladenstein, R., Thermal unfolding of the DNA-binding protein Sso7d from the hyperthermophile Sulfolobus solfataricus. *J. Mol. Biol.* 1996, *264*, 1132-1144.

30. Shehi, E.; Granata, V.; Del Vecchio, P.; Barone, G.; Fusi, P.; Tortora, P.; Graziano, G., Thermal stability and DNA binding activity of a variant form of the Sso7d protein from the archeon Sulfolobus solfataricus truncated at leucine 54. *Biochemistry* 2003, *42*, 8362-8368.

31. Driessen, R. P. C.; Meng, H.; Suresh, G.; Shahapure, R.; Lanzani, G.; Priyakumar, U. D.; White, M. F.; Schiessel, H.; van Noort, J.; Dame, R. T., Crenarchaeal chromatin proteins Cren7 and Sul7 compact DNA by inducing rigid bends. *Nucleic Acids Res.* 2013, *41*, 196-205.

32. Napoli, A.; Zivanovic, Y.; Bocs, C.; Buhler, C.; Rossi, M.; Forterre, P.; Ciaramella, M., DNA bending, compaction and negative supercoiling by the architectural protein Sso7d of Sulfolobus solfataricus. *Nucleic Acids Res.* 2002, *30*, 2656-2662.

33. Bates, A. D.; Maxwell, A., *DNA Topology*. Oxford University Press: Oxford, 2005; p 216.

34. Hardy, C. D.; Martin, P. K., Biochemical characterization of DNA-binding proteins from Pyrobaculum aerophilum and Aeropyrum pernix. *Extremophiles* 2008, *12*, 235-246.

35. Golinska, M. D.; de Wolf, F.; Stuart, M. A. C.; Hernandez-Garcia, A.; de Vries, R., Pearl-necklace complexes of flexible polyanions with neutral-cationic diblock copolymers. *Soft Matter* 2013, *9*, 6406-6411.

36. Mastrobattista, E.; van der Aa, M.; Hennink, W. E.; Crommelin, D. J. A., Artificial viruses: a nanotechnological approach to gene delivery. *Nature Reviews Drug Discovery* 2006, *5*, 115-121.

37. Driessen, R. P. C.; Dame, R. T., Nucleoid-associated proteins in Crenarchaea. *Biochem. Soc. Trans.* 2011, *39*, 116-121.

38. Guo, L.; Feng, Y. G.; Zhang, Z. F.; Yao, H. W.; Luo, Y. M.; Wang, J. F.; Huang, L., Biochemical and structural characterization of Cren7, a novel chromatin protein conserved among Crenarchaea. *Nucleic Acids Res.* 2008, *36*, 1129-1137.

39. Barnard, A.; Smith, D. K., Self-Assembled Multivalency: Dynamic Ligand Arrays for High-Affinity Binding. *Angew. Chem.-Int. Edit.* 2012, *51*, 6572-6581.

40. Zampini, M.; Hayes, F., Combinatorial targeting of ribbon-helix-helix artificial transcription factors to chimeric recognition sites. *Nucleic Acids Res.* 2012, *40*, 6673-6682.

41. Toettcher, J. E.; Mock, C.; Batchelor, E.; Loewer, A.; Lahav, G., A synthetic-natural hybrid oscillator in human cells. *Proc. Natl. Acad. Sci. U. S. A.* 2012, *107*, 17047-17052.

5.6 Appendix

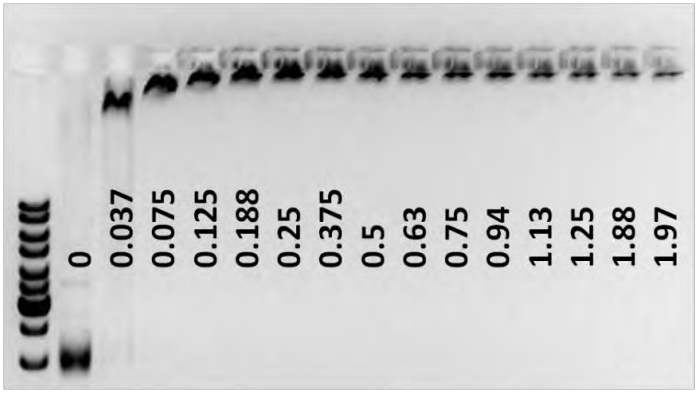


Figure 5.11. Electrophoretic Mobility Shift Assay of 2.6 kbp pDNA complexed with C_8-B^{Sso7d} at higher Proteins per base pair ratios (numbers at the middle). Left lane is the molecular markers.

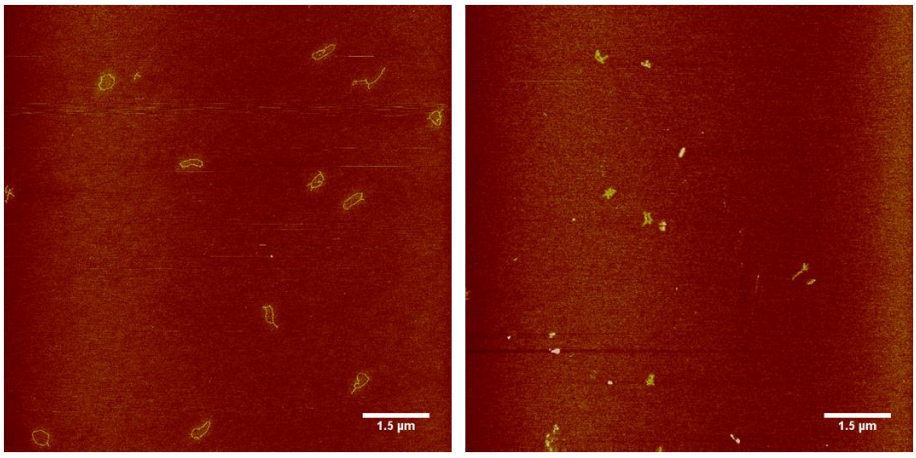


Figure 5.12. AFM images of M13 single stranded DNA (ssDNA) complexed with C_8-B^{Sso7d} (left) and C_4-B^{K12} (right).

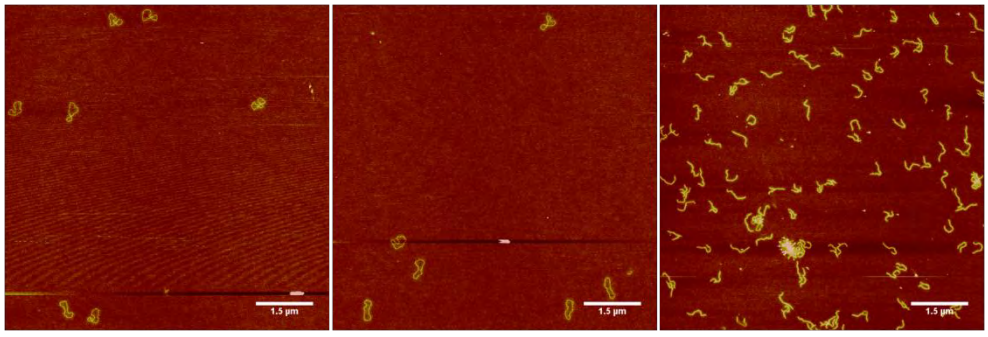


Figure 5.13. AFM images of pDNA complexed with C_8-B^{Sso7d} (left and center) and C_4-B^{K12} (right).

Chapter 6

Supramolecular nanorods from triblock protein polymers[1]

[1] This chapter is based on Armando Hernandez-Garcia, Frits A. de Wolf, Martien Cohen Stuart, and Renko de Vries. Supramolecular nanorods from genetically engineered protein-based triblock polymers. Manuscript, 2013.

6.1 Introduction

The design of peptides and proteins that self-assemble into supramolecular structures with specific, well-ordered architectures (e.g. rod-like, cages, networks, etc) is a fast-growing field.[1-8] The challenge is to "encode" the required multiple weak bonds that lead to folding and self-assembly into the primary sequence of the proteins or peptides, such that a network of non-covalent interactions is formed, based on chemical and geometric complementarity.

Using peptides and proteins to construct nanomaterials has several advantages over using traditional polymeric building blocks. The 20 natural amino acid monomers plus a number of other non-canonical amino acids can now all be used in protein design, which gives an enormous parameter space of sequenced amino-acid polymers. Short peptides can be prepared using solid state synthesis and longer polypeptides using genetic engineering, with full control over primary sequence. In addition, nature provides an extensive library of modules with well-defined secondary structures such as beta-sheets, beta- roll, alpha-helix, beta-turn, coiled coils, and others that can be incorporated as modules in larger protein designs.[9, 10] Using these structural modules makes it possible, for example, to position functional groups at precise locations on self-assembled objects, which is very difficult to achieve with self-assembling polymers prepared using synthetic chemistry. More generally speaking, it enables us to organize matter at the nanoscale with a precision much closer to that found in nature than synthetic chemical approaches. In contrast to synthetic polymer synthesis, the use of recombinant microorganisms allow for the production of long, fully sequenced and perfectly monodisperse (amino-acid polymers).[11] Control over sequence and monodispersity is crucial when aiming for the design of polymers that self-assemble into precisely controlled nanostructures, since minimal variations can drastically affect self-assembly behavior.[12] Already our ability to produce de-novo designed, self-assembling peptides and proteins has led to new routes for addressing a number of, in particular biomedical, technological problems. For example, peptide-based nano-rod structures with different functionalities have

been used for DNA packaging, energy transfer, metallic nanowires synthesis, chemical sensing, delivery vehicles, hydrogels formation, etc.[10, 13-18]

Obtaining good control over the supramolecular self-assembly of de-novo designed peptides and proteins is a challenge, since it requires a thorough understanding of both, protein folding and supramolecular self-assembly. Studies of many groups working on functional protein-based nanomaterials have helped to formulate at least qualitative design rules that relate primary sequence to self-assembly behavior. So far, most of these studies have focused on relating primary sequence with self-assembly for short peptide building blocks.[3, 12, 19-21] Work on much longer recombinant polypeptides has mostly focused on self-assembling hydrogels[22-25] with the exception of work mainly by Chilkoti and collaborators on spherical micelle-forming polypeptides based on elastine-like polymers.[26, 27]

In chapter 3 of this thesis, we have introduced de-novo designed triblock copolymers that co-assemble with DNA into artificial virus-rodlike particles.[14] These co-assembled artificial virus structures may have important future applications as nanocarriers for drug delivery, and other biomedical and nanotechnological applications. We have noted that these triblock polypeptides also self-assemble in the absence of a template, though only at higher concentrations, forming small, finite nanorods.

Here we study in more detail the self-assembly of these triblock copolymers, in the absence of a template. In order to characterize the self-assembly, we use dynamic light scattering (DLS) and atomic force microscopy (AFM) for obtaining information on their size, structure and mass, both at the ensemble level, and at the level of individual nanorods. We identify a critical concentration for the growth of nanorods and show how this critical concentration correlates with polypeptide primary sequence, in particular with the length of the middle block, which is a silk-like self-assembly block.

By studying the self-assembly of the triblocks in the absence of a template, we expect to be able to advance our understanding of self-assembly of de-novo

designed multiblock co-polypeptides in general, and possibly distill design rules that could benefit further work on the design of supramolecular assemblies composed of sequenced (poly)peptides. Our study is also relevant for understanding and designing polypeptide guided co-assembly with other entities (DNA, negative polyelectrolytes, etc.)[14, 28] since designability of co-assembly also implies control over the self-assembly of the polypeptides in the absence of a template. Finally, all of this is important to bring us closer to towards the realization of rationally designed, protein-based self-assembling systems with real biomedical- or nanotechnological applications.[29, 30]

6.2 Materials and Methods

6.2.1. Materials. Protein-based triblock copolymers C_4-S_{14}-B^{K12} (MW: 47,027.5 Da), C_4-S_{10}-B^{K12} (MW: 45,160 Da), C_4-S_2-B^{K12} (MW: 40,315 Da), C_4-S_4-B^{K12} (MW: 41,458 Da) and C-B^{K12} (MW: 38,608 Da) were produced by genetic engineering and biosynthesis in the recombinant yeast *P. pastoris* as reported in chapter 3 of this thesis.

6.2.2. Sample Preparation. Samples were prepared from lyophilized and desalted protein powder stored in sealed polypropylene tubes. Chemicals $NaH_2PO_4.H_2O$ and $Na_2HPO_4{\cdot}2H_2O$ of analytical grade purchased from Merk were used to prepare the buffer (10 mM, pH 7.4). Dithiothreitol (analytical grade, Sigma) at a concentration of 5 or 0.1mM (for DLS and AFM respectively) was present in all protein solutions to prevent the formation of disulfide bridge between protein monomers (all proteins, except C-B, have a cysteine at their N-terminus). Solutions were prepared from freshly dissolved powder proteins (1-5 minutes) and never stored for reuse. For weighing protein powders, a Mettler Toledo XS205 scale was used with an error of ± 0.01 mg. Concentrated stock solutions were prepared in eppendorf tubes using MQ-water, and diluted to the required concentration with a buffer.

6.2.3. Dynamic Light Scattering. DLS measurements were carried out using a Zetasizer NanoZS equipment from Malvern Instruments, UK. It is equipped with a 4 mW He-Ne ion laser at a wavelength of 633 nm. The measurements were usually the average of 6 runs for 30 seconds at 25°C taken every 5 minutes using scattering angle of 173°. The concentrations of the proteins were between 4.4 to 110.7 μM for C_4-S_{10}-B^{K12} and 1.1 to 42.1 μM for C_4-S_{14}-B^{K12}.

6.2.4. Atomic Force Microscopy. Samples were analyzed using a Digital Instruments NanoScope V equipped with a silicon nitride probe (Bruker, MA, USA) with a spring constant of 0.4-0.35 N/m in ScanAsyst™ imaging mode. Images were recorded between 0.488-0.977 Hz and 1024 samples/line. Image processing and height measurement was done with NanoScope Analysis 1.20 software using a first order flattening for all the images. The concentration of protein samples was varied over several orders of magnitude (between 20 nM to 110.7 μM for C_4-S_{10}-B^{K12} and 11 nM to 42.1 μM for C_4-S_{14}-B^{K12}). For imaging, the most concentrated samples were diluted with buffer to final concentrations of ~0.4 μM. Approximately 4 μL of sample was deposited onto clean ultraflat silicon surfaces. After 2 minutes of sample incubation, surfaces were rinsed with 1 mL of Milli-Q water in order to remove salts and non-absorbed particles. This was followed by removal of excess water using a tissue and slow drying under a N_2 stream. Measurements of the dimensions of nanostructures were performed using ImageJ software.

6.3 Results and Discussion

6.3.1. Protein Self-assembly Design. Triblock copolymers, "C_4-S_n-B^{K12}", were used with different lengths *n* of the silk-like self-assembly block: *n* = 2, 4, 10 and 14. Most results are for the variants C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12} that form virus-like particles in co-assembly with nucleic acids (chapter 3). The self-assembly block S_n = $(GAGAGAGQ)_n$ forms beta-strands that presumably fold into a beta-roll secondary structures. Multiple beta-rolls stack on top of each other, driven by hydrophobic interactions to form tape-like structures,[31, 32, 33-35] as illustrated in

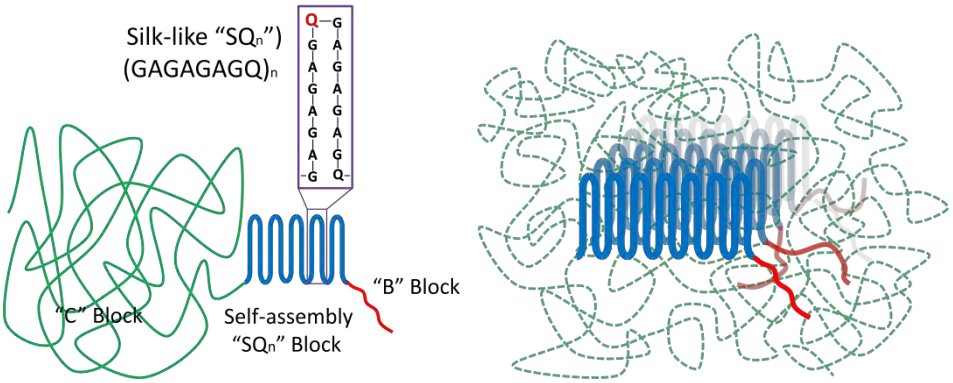


Figure 6.1. Design of self-assembling protein-based triblock copolymers C_4-S_n-B^{K12}, and cartoon of hypothetical self-assembled nanorod structures formed in solution. Left, block structures of C_4-S_n-B^{K12} protein-based polymers. The protein is made up of three different blocks. The middle one, "S_n", is the silk-like self-assembly block (blue), the "C_4" block is for colloidal stability (green) and the "B^{K12}" block is a cationic segment (red) that can bind to negatively charged templates and thus leading to co-assembly. Right, cartoon of hypothetical self-assembled nanorod structures formed by C_4-S_n-B^{K12} proteins for which the self-assembly block is sufficiently long. The core is formed by the stacked silk-like sequences together with the cationic sequence, and it is surrounded by the C_4-block corona.

Figure 6.1. The silk-like motif S= (GAGAGAGX) has been used in a range of other polymers that were designed to self-assemble into functional biomaterials.[23, 33, 35, 36] For the C_4-S_n-B^{K12} polymers we use the amino acid glutamine (Q, the amine version of the glutamic acid amino acid, which is hydrophilic and non-charged) as the corner residue X that should lead to the beta-turns of the hairpins.

Protein-based polymers with silk-like self-assembly blocks S_n=$(GAGAGAGX)_n$ with X = histidine, lysine, or glutamic acid have been studied before[23, 37] but so far, silk-like self-assembly blocks where X is a neutral but hydrophilic glutamine residue, had not been. For C_4-S_{10}-B^{K12} the self-assembly block is 84 amino acids long, for C_4-S_{14}-B^{K12} it is 116 amino acids long. The "C"-block is a very hydrophilic block that in solution adopts a random coil structure. It is 407 amino acids long, and its function is to provide colloidal stability of the self-assembled nanostructures (to prevent aggregation).[14, 31, 34, 35, 37, 38] The final block, "B" is made up of 12 lysines and was incorporated to allow for binding to negatively charged templates and hence for co-assembly (see chapters 2, 3, 7 and 8 of this thesis).

6.3.2. Kinetics of Self-assembly from Static Light Scattering. Solutions of several concentrations of C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12} were prepared and the development of their scattering intensity was measured starting immediately after dissolution, up to incubation times of a few hours. Results are shown in Figure 6.2a and 6.2b for, respectively, C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12}. First consider the results for C_4-S_{10}-B^{K12} (Figure 6.2a). For this case, scattering starts at low values and increases with a rate that increases with concentration. At the highest concentration, 110.7 μM, equilibration took over 10h. For C_4-S_{14}-B^{K12} (Figure 6.2b), scattering intensities do not start at low values but instead are high immediately, and increase only very slowly over time. Most likely, whereas the C_4-S_{10}-B^{K12} dissolves molecularly, the C_4-S_{14}-B^{K12} with its stronger drive for self-assembly does not dissolve molecularly but instead dissolves as clusters, that very slowly rearrange to adopt their final equilibrium configuration. Results for C_4-S_{10}-B^{K12} are qualitatively similar to those for the fiber-forming symmetric triblock polypeptides with very large central silk-like blocks that have been studied before. These have silk-like blocks S_{48}=$(GAGAGAGX)_{48}$, where "X" is histidine, lysine or glutamic acid. These can also be dissolved molecularly, at extreme pH values. For this case self-assembly can be initiated by a quench to a pH where the corner residue X is no longer charged. The previous data is for fiber-forming symmetric triblock polypeptides at low concentrations (1.5-9 μM). In that case, the scattering was found to level off after about 20 h.[37]

6.3.3. Concentration-dependence of self-assembly from static light scattering. In our studies on co-assembly of C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12} with nucleic acid templates it was found that in the absence of a template, the polypeptides form nanorods, but only at higher concentrations. Here, we consider the concentration dependence of the self-assembly of these polypeptides in the absence of a template in somewhat more detail. Scattering intensities of well-equilibrated samples (incubation times longer than 20h) were plotted as a function of the protein concentration, with the result shown in Figure 6.3. Clearly, both C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12} show a critical concentration at which there is an abrupt change of the slope of the intensity versus concentration curves. Critical concentrations are 20

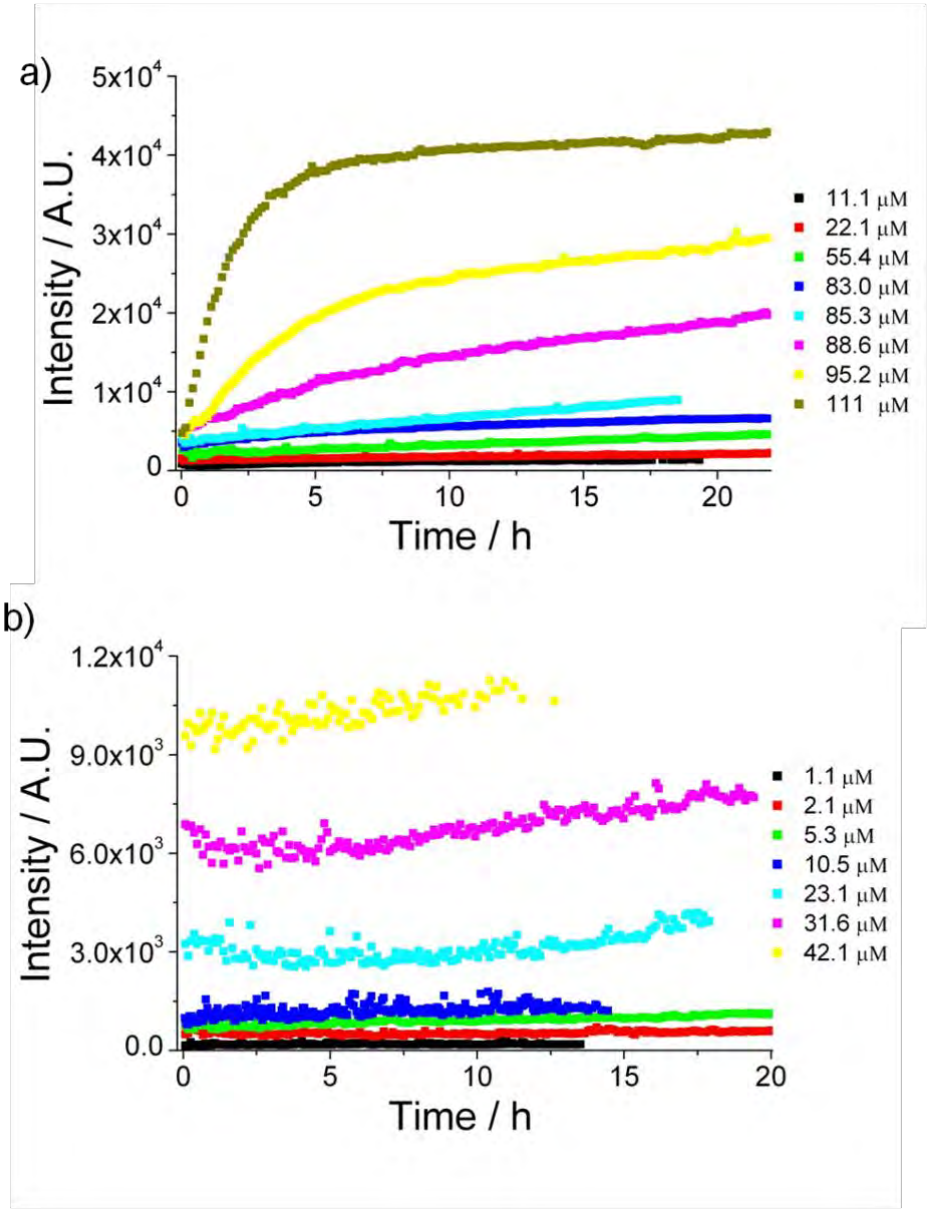


Figure 6.2. Kinetics of self-assembly of C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12} at different concentrations followed using light scattering. Scattering intensity (AU) of protein samples as a function of incubation time, for samples with different concentrations. a) C_4-S_{10}-B^{K12} and b) C_4-S_{14}-B^{K12}.

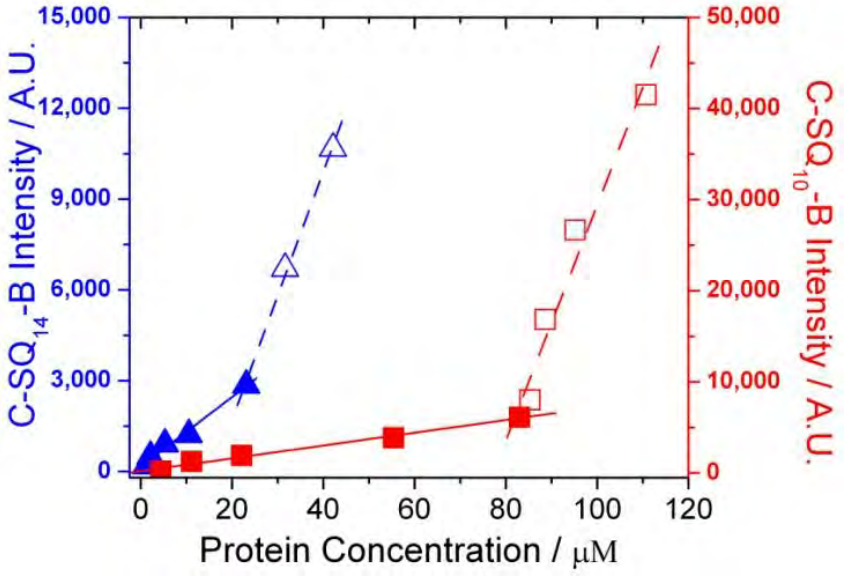


Figure 6.3. Critical concentration for the self-assembly of C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12} as detected using light scattering. Intensity of scattered light versus protein concentration for C_4-S_{14}-B^{K12} (blue ▲, Δ; samples incubated 9-11h) and for C_4-S_{10}-B^{K12} (red ▪, □; samples incubated 14-16h). Filled and open symbols are, respectively, below and above the critical concentration for self-assembly. Solid and dashed lines are linear fits respectively, above and below the critical concentration.

and 80 μM for, respectively, C_4-S_{14}-B^{K12} and C_4-S_{10}-B^{K12}. Assuming that the change in slope signals the onset of nanorod formation (we will come back to this point when discussing the AFM results), it is clear that longer self-assembly blocks favor the start of nano-rod formation at lower concentrations.Furthermore, the difference between the proteins highlights the sensitivity of the self-assembly to the primary sequence, in this case, to the length of the silk block. Studies using short undecapeptide peptides have also highlighted that small variations in the sequence of such peptides can lead to drastic changes in the critical concentration for self-assembly.[20] A similar sensitivity was for example also found for so-called lipid-like or peptide surfactants.[21, 39-41] Likewise, the critical self-assembly concentration of our triblock polypeptides is a very sensitive function of the length of the silk block.

So far we have not touched on the role of the "B" block that was included in the protein-based polymer design in order to give rise to co-assembly, but which may in fact also influence the self-assembly of the triblocks in the absence of a template. In order to test for a possible influence of electrostatic repulsion between "B" blocks, we have investigated the influence of ionic strength on the self-assembly of the triblocks. For ionic strengths up to 0.2 M NaCl, we find that the self-assembly kinetics is a bit slower and there is hardly any influence on the final scattering intensities (Appendix, Figure 6.9). Hence it appears that electrostatic repulsion interactions between the "B" blocks do not have a significant influence on the self-assembly of C_4-S_{10}-B^{K12} nanostructures. This is in clear contrast to the long hydrophilic C_4-block that has been shown to significantly oppose the folding and stacking of the silk-like blocks, as was shown in simulation studies.[31, 34]

6.3.4. Concentration Dependence of Nanostructure Sizes from Dynamic Light Scattering. An interesting question is of course whether above the critical concentration that we observe, the protein-polymers form larger objects, more objects, or both. To address this issue we have performed Dynamic Light Scattering (DLS) on well equilibrated solutions of both C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12} at a range of concentrations. It was found that typically a single diffusion peak dominates the light scattering, and this peak corresponds to the typical size range of nanorods (between 40 nm and 100 nm, depending on conditions) that was already observed in AFM. The effective hydrodynamic radius R_H corresponding to this dominant peak is plotted as a function of protein concentration in Figure 6.4a and 6.4b for respectively, C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12}. For the case of C_4-S_{10}-B^{K12} a clear change in the size of the nanorods is observed around the critical concentration of around 80μM (from about 45 to 60nm). Below the critical concentration, a small but measurable part of the scattering is still due to protein monomers with a hydrodynamic radius of around 6 nm. The monomer contribution to the total scattering quickly vanishes when approaching and passing the critical concentration. For C_4-S_{14}-B^{K12} the size of the self-assembled objects is significantly larger than for C_4-S_{10}-B^{K12}. In this case, no clear change in the size of self-assembled objects is observed at the critical concentration, and a monomer signal could not be

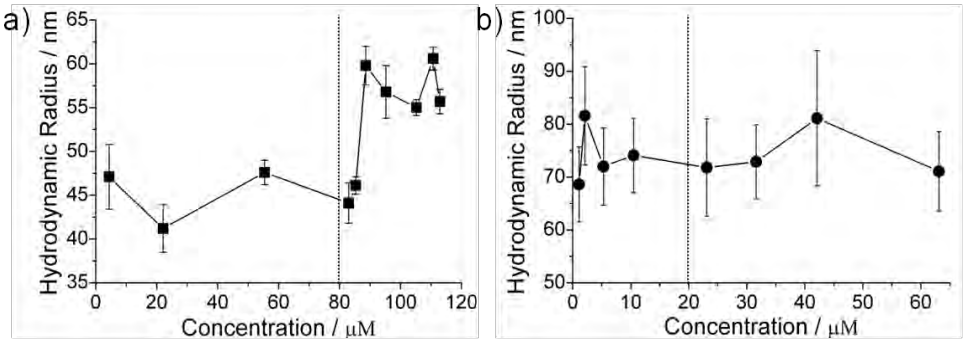


Figure 6.4. Concentration dependence of the effective hydrodynamic radius R_H of self-assembled objects of C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12}. a) [▪]C_4-S_{10}-B^{K12} b) [●]C_4-S_{14}-B^{K12}. Dotted vertical lines indicate the critical concentrations as deduced from Figure 6.3.

observed, not even at very low concentrations. This confirms the kinetic data that likewise suggest that C_4-S_{14}-B^{K12} does not dissolve molecularly. Again we find that in increasing the length of the silk block from 10 to 14, the self-assembly behavior changes significantly.

6.3.5. Kinetics of Protein Self-assembly from AFM. Initial indications for nanorod formation of C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12} came in the absence of a template came from AFM (chapter 3). These were done at a fixed protein concentration, and after complete equilibration. Here we use AFM to study details that are inaccessible to DLS: the development of the complete size-distribution of self-assembled nanorods as a function of time after dissolution, and as a function of protein concentration. By comparing C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12} we will also be able to correlate the self-assembly with the protein design. First we consider the time-dependence of self-assembly.

Self-assembly of a C_4-S_{10}-B^{K12} above its critical concentration (110.7 μM) was followed by imaging samples obtained at different times after dissolution using AFM (Figure 6.5). The slow kinetics of the process in principle allows us to visualize the entire process of self-assembly, from very small clusters to the final

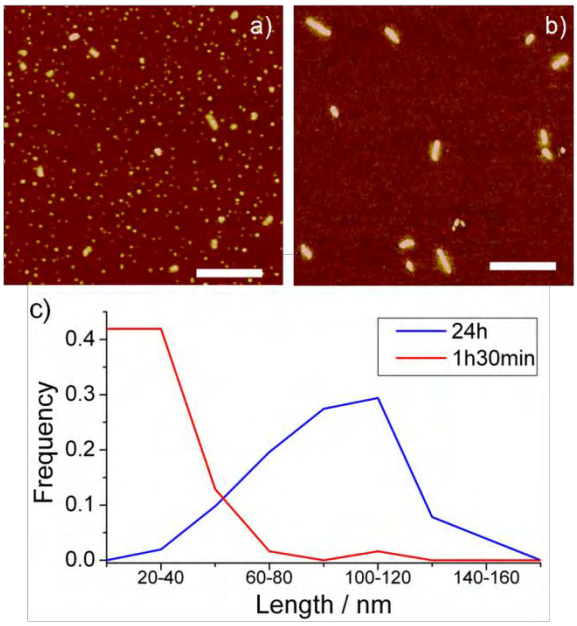


Figure 6.5. Kinetics of self-assembly of C_4-S_{10}-B^{K12} from AFM. Top, AFM images of protein assemblies formed after (a) 1h30 min and (b) 24h of incubation time of a 110.7 μM protein solution. Scale bar is 300 nm. Bottom, (c) distribution of the lengths of the C_4-S_{10}-B^{K12} assemblies formed at different times. (N=62 for image a, and N= 50 samples for b).

self-assembled nanorods. Representative examples for 2h45 and 24h, illustrating early and late stages of the self-assembly reaction, are shown in Figure 6.5a, and 5.b respectively. Size distributions corresponding to these images are shown in Figure 6.5c. In qualitative agreement with the static light scattering, in the first hours of the self-assembly reaction, we only find small clusters and a few short nano-rods. Only after a very long time (24h), all proteins have assembled into nano-rods.

Representative higher magnification images of equilibrated single nano-rods formed above the critical concentrations by C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12} are shown in Figure 6.6. They are highly regular, and apart from their length comparable to fiber structures that have been found before for the symmetric triblocks with central S_{48} blocks that have been studied before. Most likely, therefore, they have a similar structure, and consist of a central core formed by stacked silk blocks, with

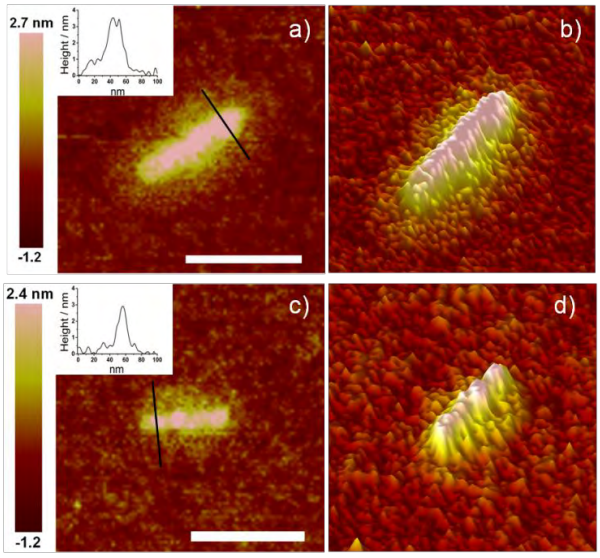


Figure 6.6. AFM images of typical self-assembled nanorod structures of designed proteins. Nanorods images of (top) C_4-S_{10}-B^{K12} formed at 110.7 μM (a) 2D and (b) 3D image. Bottom, C_4-S_{14}-B^{K12} formed at 42.1 μM (c) 2D and (d) 3D images. Insets are corresponding height profile of the section indicated as a black line on image. Bar scale is 250 nm. Samples were incubated at room temperature for 24h.

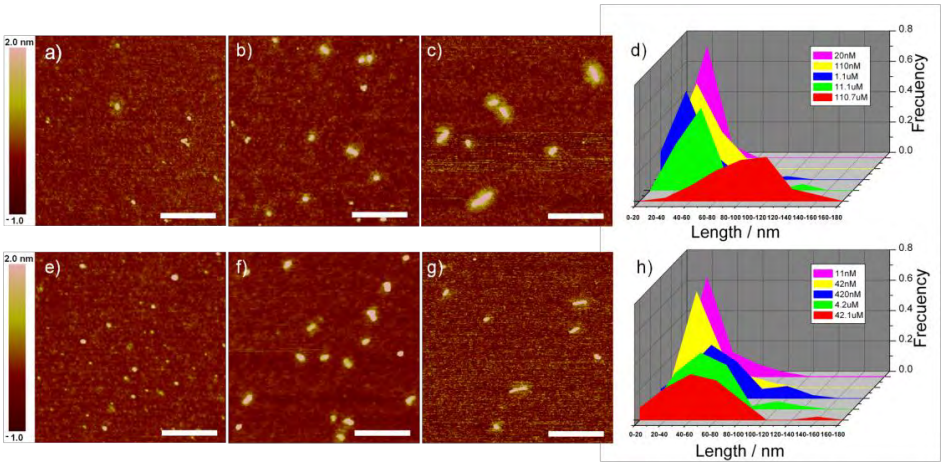


Figure 6.7. Concentration dependence of self-assembly of protein nanorods. Top row: AFM images of C_4-S_{10}-B^{K12} nanostructures at a) 20 nM, b) 1.1 μM and c) 110.7 μM. Average lengths are, respectively, 26.8, 52.5 and 91.5 nm. d) Size distributions of C_4-S_{10}-B^{K12} nanostructures for a range of protein concentration. Bottom row: AFM images of C_4-S_{14}-B^{K12} nanostructures at e) 11 nM, f) 0.42 μM and g) 42.1 μM. Average lengths are, respectively, 38.4, 58.31 and 55.3 nm. h) Size distributions of C_4-S_{14}-B^{K12} nanostructures for a range of protein concentrations. Images were acquired using *ScanAsyst* mode in air on samples that had been equilibrated (in solution) for 24h. Scale bar is 250 nm.

pendant "B" blocks, and are surrounded by a brush-like corona formed by the "C" blocks, as illustrated schematically in Figure 6.1b. The height of the nano-rods as measured by AFM is similar for both polymers, and is also similar to that measured using AFM for the much longer fibers formed by symmetric triblocks with S_{48} central blocks.[23, 33] Remarkably, the width of the C_4-S_{10}-B^{K12} nano-rods is distinctly larger than that of nano-rods formed by C_4-S_{14}-B^{K12}.

6.3.6. Concentration Dependence of Nano-rod Size Distributions from AFM. For both C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12} we first established a lower bound on the concentration for which nano-rods can be detected with AFM. Next, we quantified lengths for well equilibrated samples of C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12} (24h) over a wide range of concentrations. Selected AFM images show the strong concentration dependence of the self-assembly process for both proteins (Figure 6.7a-c and Figure 6.7e-g for, respectively, C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12}). Size distributions as a function of concentration are shown in Figure 6.7d and Figure 6.7h (for, respectively, C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12}). A comparison of the evolution of the position of the maximum of the distribution for the two proteins is shown in Figure 6.8. That figure shows that in the case of C_4-S_{10}-B^{K12} at concentration higher than 11 μM there is a big jump in the average length of the peak of the distribution. In the case of C_4-S_{14}-B^{K12} the increase happens at earlier concentration (>42 nM). Concentration dependent equilibrium lengths, and critical concentrations above which lengths start increasing more rapidly, are a common feature of many self-assembling systems that form 1D structures such as rods and tubes, including peptide-based systems.[21, 39, 41] We have also comparing the height profiles of equilibrated nano-assemblies formed at different concentrations (Appendix, Figure 6.10). It appears that the height as determined using AFM increases from about 1.0 nm at 20 nM to about 3.5 nm at 110.7 μM of C_4-S_{10}-B^{K12}. This suggests that the folding and packing of the beta-strands in the nano-assemblies may be quite concentration dependent. Possibly, at low concentrations the folding and stacking

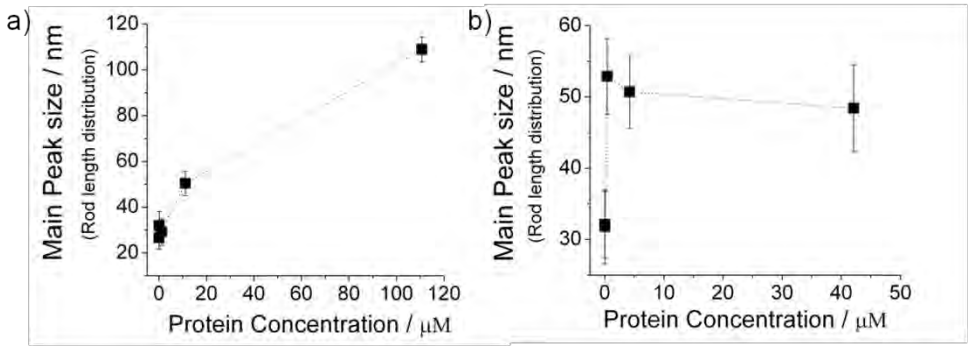


Figure 6.8. Concentration dependence of position of the peak of the size distribution (Figure 6.7d and 6.7h). A) C_4-S_{10}-B^{K12} and B) C_4-S_{14}-B^{K12}.

of the beta-strands is still highly irregular and imperfect and only at much higher concentrations the driving force for regular folding and stacking is strong enough.

6.4 Concluding Remarks

We have shown that the triblock copolymer proteins C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12} self-assemble into elongated nanorods (up to ~150 nm long at the highest concentrations) through the interaction of their silk-like blocks. Changing the length of the silk block from 10 to 14 silk blocks has quite a drastic influence on self-assembly: whereas C_4-S_{10}-B^{K12} assembles slowly from dissolved monomers, C_4-S_{14}-B^{K12} does not dissolve molecularly, and shows nanostructures immediate after dissolution. Both proteins show a critical concentration beyond which the amount and sizes of nano-rods rapidly increase with a further increase of the concentration. This critical concentration decreases rapidly with increasing length of the silk block. The most likely supramolecular structure is that of a core composed of the silk-like and the cationic B^{K12} block, surrounded by a corona composed of C_4 blocks, as shown in Figure 6.1b. The short cationic block B^{K12}, made up of 12 lysines and is crucial for co-assembly on negatively charged templates, appears not to influence self-assembly in the absence of a template. This implies that the C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12} multiblocks appear to conform to the simple picture of multiple functional blocks that can be joined to get a multi-functional polymer

without cross-talk between the different functions, which is very attractive from a polymer-design point of view.

The strong influence of the length of the silk block on self-assembly behavior is very clear from the comparison of just these two proteins. As has been demonstrated in simulation studies, an initial hydrophobic nucleation step that involves a small number of peptides that fold and assemble together, is required for triggering the self-assembly of silk-like and other peptide sequences that are similar to the silk-like blocks that we use here.[31, 32] Provided nucleation is homogeneous this implies a self-assembly that is highly sensitive to both concentration and chemical details such as the length of the silk-like block. This matches our finding of strong concentration dependence, and the fact that the critical conentration was much lower for C_4-S_{14}-B^{K12} than for C_4-S_{10}-B^{K12}.

For eventual applications, stability against changes of environmental conditions is of course crucial. It is well known that highly cooperative forms of protein self-assembly[42-44] can be hysteretic in the sense that there can be large differences in the timescales for assembly and disassembly. Although we have not studied disassembly in detail, we have found that after 4h the particle sizes of a nanorod dispersion (110 μM) diluted 500x remain constant, suggesting that disassembly can be also significantly slower than assembly for the nano-rods, which could be very advantageous for delivery applications.

While initially designed for co-assembly, one can imagine a number of applications of these self-assembled nano-rods in the absence of templates. Since the triblock copolymers are biodegradable, non-hemolytic and non-toxic (see chapter 4), the nano-rods that they form could be interesting platforms for nano-medicines, for example as a delivery systems[11, 45] that due to their elongated structure could penetrate cells.[46, 47] The outside of the "C" blocks can be modified (either chemically or at the level of the design of the protein) such that the rods become a scaffold for displaying a range of ligands.[46, 48, 49] Furthermore, we have shown that the cationic "B" block does not strongly influence self-assembly, hence similar blocks could be included without changing self-assembly. Such blocks

would be close to the nano-rod core and hence presumably well protected against proteolysis and/or cellular recognition. A potential application would therefore be to use cationic antimicrobial sequences instead of the "B" block, in order to improve their stability and circulation time in-vivo.

6.5 References

1. Zhang, S. G., Fabrication of novel biomaterials through molecular self-assembly. *Nat. Biotechnol.* 2003, *21*, 1171-1178.

2. Bromley, E. H. C.; Channon, K.; Moutevelis, E.; Woolfson, D. N., Peptide and protein building blocks for synthetic biology: From programming biomolecules to self-organized biomolecular systems. *ACS Chemical Biology* 2008, *3*, 38-50.

3. Ulijn, R. V.; Smith, A. M., Designing peptide based nanomaterials. *Chemical Society Reviews* 2008, *37*, 664-675.

4. van Hest, J. C. M.; Tirrell, D. A., Protein-based materials, toward a new level of structural control. *Chem. Commun.* 2001, 1897-1904.

5. Padilla, J. E.; Colovos, C.; Yeates, T. O., Nanohedra: Using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proceedings of the National Academy of Sciences of the United States of America* 2001, *98*, 2217-2221.

6. Rabotyagova, O. S.; Cebe, P.; Kaplan, D. L., Protein-Based Block Copolymers. *Biomacromolecules* 2011, *12*, 269-289.

7. Ringler, P.; Schulz, G. E., Self-assembly of proteins into designed networks. *Science* 2003, *302*, 106-109.

8. Sinclair, J. C.; Davies, K. M.; Venien-Bryan, C.; Noble, M. E. M., Generation of protein lattices by fusing proteins with matching rotational symmetry. *Nat. Nanotechnol.* 2011, *6*, 558-562.

9. Ryadnov, M. G.; Bella, A.; Timson, S.; Woolfson, D. N., Modular Design of Peptide Fibrillar Nano- to Microstructures. *J. Am. Chem. Soc.* 2009, *131*, 13240-+.

10. DiMarco, R. L.; Heilshorn, S. C., Multifunctional Materials through Modular Protein Engineering. *Advanced Materials* 2012, *24*, 3923-3940.

11. Garanger, E.; Lecommandoux, S., Towards Bioactive Nanovehicles Based on Protein Polymers. *Angewandte Chemie-International Edition* 2012, *51*, 3060-3062.

12. Caplan, M. R.; Schwartzfarb, E. M.; Zhang, S. G.; Kamm, R. D.; Lauffenburger, D. A., Effects of systematic variation of amino acid sequence on the mechanical properties of a self-assembling, oligopeptide biomaterial. *Journal of Biomaterials Science-Polymer Edition* 2002, *13*, 225-236.

13. Miller, R. A.; Presley, A. D.; Francis, M. B., Self-assembling light-harvesting systems from synthetically modified tobacco mosaic virus coat proteins. *J. Am. Chem. Soc.* 2007, *129*, 3104-3109.

14. Hernandez-Garcia, A.; Werten, M. W. T.; Stuart, M. C.; de Wolf, F. A.; de Vries, R., Coating of Single DNA Molecules by Genetically Engineered Protein Diblock Copolymers. *Small* 2012, *8*, 3491-3501.

15. Kyle, S.; Aggeli, A.; Ingham, E.; McPherson, M. J., Recombinant self-assembling peptides as biomaterials for tissue engineering. *Biomaterials* 2009, *31*, 9395-9405.

16. Aggeli, A.; Bell, M.; Boden, N.; Keen, J. N.; Knowles, P. F.; McLeish, T. C. B.; Pitkeathly, M.; Radford, S. E., Responsive gels formed by the spontaneous self-assembly of peptides into polymeric beta-sheet tapes. *Nature* 1997, *386*, 259-262.

17. Hartgerink, J. D.; Beniash, E.; Stupp, S. I., Self-assembly and mineralization of peptide-amphiphile nanofibers. *Science* 2001, *294*, 1684-1688.

18. Maeda, Y.; Matsui, H., Genetically engineered protein nanowires: unique features in site-specific functionalization and multi-dimensional self-assembly. *Soft Matter* 2012, *8*, 7533-7544.

19. Aggeli, A.; Bell, M.; Boden, N.; Keen, J. N.; McLeish, T. C. B.; Nyrkova, I.; Radford, S. E.; Semenov, A., Engineering of peptide beta-sheet nanotapes. *Journal of Materials Chemistry* 1997, *7*, 1135-1145.

20. Aggeli, A.; Nyrkova, I. A.; Bell, M.; Harding, R.; Carrick, L.; McLeish, T. C. B.; Semenov, A. N.; Boden, N., Hierarchical self-assembly of chiral rod-like molecules as a model for peptide beta-sheet tapes, ribbons, fibrils, and fibers. *Proceedings of the National Academy of Sciences of the United States of America* 2001, *98*, 11857-11862.

21. Yang, Y. L.; Khoe, U.; Wang, X. M.; Horii, A.; Yokoi, H.; Zhang, S. G., Designer self-assembling peptide nanomaterials. *Nano Today* 2009, *4*, 193-210.

22. Werten, M. W. T.; Teles, H.; Moers, A. P. H. A.; Wolbert, E. J. H.; Sprakel, J.; Eggink, G.; de Wolf, F. A., Precision Gels from Collagen-Inspired Triblock Copolymers. *Biomacromolecules* 2009, *10*, 1106-1113.

23. Martens, A. A.; Portale, G.; Werten, M. W. T.; de Vries, R. J.; Eggink, G.; Stuart, M. A. C.; de Wolf, F. A., Triblock Protein Copolymers Forming Supramolecular Nanotapes and pH-Responsive Gels. *Macromolecules* 2009, *42*, 1002-1009.

24. Shen, W.; Kornfield, J. A.; Tirrell, D. A., Structure and mechanical properties of artificial protein hydrogels assembled through aggregation of leucine zipper peptide domains. *Soft Matter* 2007, *3*, 99-107.

25. Petka, W. A.; Harden, J. L.; McGrath, K. P.; Wirtz, D.; Tirrell, D. A., Reversible hydrogels from self-assembling artificial proteins. *Science* 1998, *281*, 389-392.

26. McDaniel, J. R.; Callahan, D. J.; Chilkoti, A., Drug delivery to solid tumors by elastin-like polypeptides. *Adv. Drug Deliv. Rev.* 2010, *62*, 1456-1467.

27. Osborne, J. L.; Farmer, R.; Woodhouse, K. A., Self-assembled elastin-like polypeptide particles. *Acta Biomaterialia* 2008, *4*, 49-57.

28. Ruff, Y.; Moyer, T.; Newcomb, C. J.; Demeler, B.; Stupp, S. I., Precision Templating with DNA of a Virus-like Particle with Peptide Nanostructures. *J. Am. Chem. Soc.* 2013, *135*, 6211-6219.

29. Chow, D.; Nunalee, M. L.; Lim, D. W.; Simnick, A. J.; Chilkoti, A., Peptide-based biopolymers in biomedicine and biotechnology. *Materials Science & Engineering R-Reports* 2008, *62*, 125-155.

30. Aida, T.; Meijer, E. W.; Stupp, S. I., Functional Supramolecular Polymers. *Science* 2012, *335*, 813-817.

31. Schor, M.; Bolhuis, P. G., The self-assembly mechanism of fibril-forming silk-based block copolymers. *Phys. Chem. Chem. Phys.* 2011, *13*, 10457-10467.

32. Ni, R.; Abeln, S.; Schor, M.; Cohen Stuart, M. A.; Bolhuis, P. G., Interplay between Folding and Assembly of Fibril-Forming Polypeptides. *Physical Review Letters* 2013, *111*, 058101.

33. Krejchi, M. T.; Cooper, S. J.; Deguchi, Y.; Atkins, E. D. T.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A., Crystal structures of chain-folded antiparallel beta-sheet assemblies from sequence-designed periodic polypeptides. *Macromolecules* 1997, *30*, 5012-5024.

34. Schor, M.; Martens, A. A.; Dewolf, F. A.; Stuart, M. A. C.; Bolhuis, P. G., Prediction of solvent dependent beta-roll formation of a self-assembling silk-like protein domain. *Soft Matter* 2009, *5*, 2658-2665.

35. Smeenk, J. M.; Otten, M. B. J.; Thies, J.; Tirrell, D. A.; Stunnenberg, H. G.; van Hest, J. C. M., Controlled assembly of macromolecular beta-sheet fibrils. *Angewandte Chemie-International Edition* 2005, *44*, 1968-1971.

36. Krejchi, M. T.; Atkins, E. D. T.; Waddon, A. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A., CHEMICAL SEQUENCE CONTROL OF BETA-SHEET ASSEMBLY IN MACROMOLECULAR CRYSTALS OF PERIODIC POLYPEPTIDES. *Science* 1994, *265*, 1427-1432.

37. Beun, L. H.; Beaudoux, X. J.; Kleijn, J. M.; de Wolf, F. A.; Stuart, M. A. C., Self-Assembly of Silk-Collagen-like Triblock Copolymers Resembles a Supramolecular Living Polymerization. *Acs Nano* 2012, *6*, 133-140.

38. Burkoth, T. S.; Benzinger, T. L. S.; Jones, D. N. M.; Hallenga, K.; Meredith, S. C.; Lynn, D. G., C-terminal PEG blocks the irreversible step in beta-amyloid(10-35) fibrillogenesis. *J. Am. Chem. Soc.* 1998, *120*, 7655-7656.

39. Khoe, U.; Yang, Y. L.; Zhang, S. G., Synergistic Effect and Hierarchical Nanostructure Formation in Mixing Two Designer Lipid-Like Peptide Surfactants Ac-A(6)D-OH and Ac-A(6)K-NH2. *Macromolecular Bioscience* 2008, *8*, 1060-1067.

40. Nagai, A.; Nagai, Y.; Qu, H. J.; Zhang, S. G., Dynamic behaviors of lipid-like self-assembling peptide A(6)D and A(6)K nanotubes. *Journal of Nanoscience and Nanotechnology* 2007, *7*, 2246-2252.

41. Yang, S. J.; Zhang, S. G., Self-assembling behavior of designer lipid-like peptides. *Supramolecular Chemistry* 2006, *18*, 389-396.

42. Whitty, A., Cooperativity and biological complexity. *Nature Chemical Biology* 2008, *4*, 435-439.

43. Williamson, J. R., Cooperativity in macromolecular assembly. *Nature Chemical Biology* 2008, *4*, 458-465.

44. Vendruscolo, M.; Dobson, C. M., Protein self-assembly intermediates. *Nature Chemical Biology* 2013, *9*, 216-217.

45. Lim, Y. B.; Lee, E.; Yoon, Y. R.; Lee, M. S.; Lee, M., Filamentous artificial virus from a self-assembled discrete nanoribbon. *Angewandte Chemie-International Edition* 2008, *47*, 4525-4528.

46. Doshi, N.; Mitragotri, S., Needle-shaped polymeric particles induce transient disruption of cell membranes. *Journal of the Royal Society Interface* 2010, *7*, S403-S410.

47. Champion, J. A.; Katare, Y. K.; Mitragotri, S., Particle shape: A new design parameter for micro- and nanoscale drug delivery carriers. *J. Control. Release* 2007, *121*, 3-9.

48. Barnard, A.; Smith, D. K., Self-Assembled Multivalency: Dynamic Ligand Arrays for High-Affinity Binding. *Angewandte Chemie-International Edition* 2012, *51*, 6572-6581.

49. Boato, F.; Thomas, R. M.; Ghasparian, A.; Freund-Renard, A.; Moehle, K.; Robinson, J. A., Synthetic virus-like particles from self-assembling coiled-coil lipopeptides and their use in antigen display to the immune system. *Angewandte Chemie-International Edition* 2007, *46*, 9015-9018.

6.6 Appendix

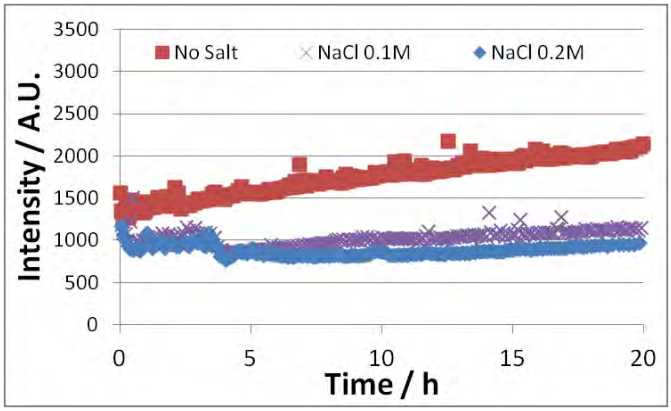


Figure 6.9 Kinetic profile of a solution 22.1 µM of C_4-S_{10}-B^{K12} with different NaCl concentrations.

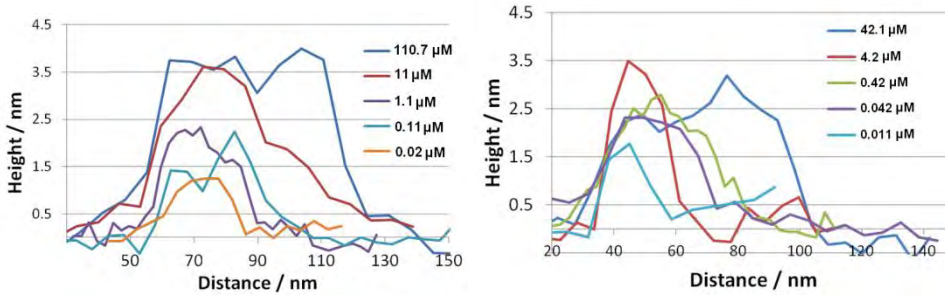


Figure 6.10 Height profiles along the longest axis of self-assembled protein solutions at different concentrations. Left, C_4-S_{10}-B^{K12} height profiles along large axis, right C_4-S_{14}-B^{K12} height profiles. Width C_4-S_{10}-B^{K12}: 13.1 ± 2.1 nm and C_4-S_{14}-B^{K12}: 8.0 ± 1.8 nm.

Table 6.1. Amino Acid Sequences of Recombinant Protein Triblock Copolymers

C_4-S_{10}-B^{K12}
GCGLGAGAPGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQGNPGKNGQPGSPGSQGSPGNQGSP GQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQGNPGKNGQPGSP GSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQ GNPGKNGQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQP GEPGSNGPQGSQGNPGKNGQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGA**GAGAGAGQGAGAGAGQGAGAGAGQGAGA** **GAGQGAGAGAGQGAGAGAGQGAGAGAGQGAGAGAGQGAGAGAGQGA***KKKKKKKKKKKKGA*
C_4-S_{14}-B^{K12}
GCGLGAGAPGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQGNPGKNGQPGSPGSQGSPGNQGSP GQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQGNPGKNGQPGSP GSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQ GNPGKNGQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQP GEPGSNGPQGSQGNPGKNGQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGA**GAGAGAGQGAGAGAGQGAGAGAGQGAGA** **GAGQGAGAGAGQGAGAGAGQGAGAGAGQGAGAGAGQGAGAGAGQGAGAGAGQGAGAGAGQGA***K* *KKKKKKKKKKKGA*

Codes: colloidal stability "C_4" block, **self-assembly "Sn" block** and *cationic "B^{K12}" block.*

Chapter 7

Self-assembled protein-metal ion nanorods[1]

[1] Armando Hernandez-Garcia, Junyou Wang, Wolf Rombouts, Aldrik Velders, Martien Cohen-Stuart and Renko de Vries. Self-assembled protein-metal ion nanorods. Manuscript in preparation. 2013.

7.1 Introduction

Self-assembly routes to build precise supramolecular functional nanostructures that exploit the photonic, magnetic or electronic properties of metal atoms are believed to have many potential applications in medicine and electronics.[1-5] For example, it was recently shown how to create micelles based on cationic block copolymers that can encapsulate metal ions in their micellar core, with potential applications in nanomedicine.[3, 6-8] While promising, these micelles still suffer from a number of problems, with respect to potential applications in nanomedicine: they were not designed to be biocompatible, their nanoarchitecture cannot be precisely controlled and they have limited stability, for example against salt.[7, 9] Hence there is a need for new self-assembly strategies that combine biocompatibility with precise control over co-assembly with metals into supramolecular nanostructures of controlled architecture. Recently, so-called protein-based polymers have emerged as highly tunable biocompatible polymers that can be precisely engineered to self-assemble into a range of nanostructures.[10-13] The protein-based polymers are made up of tandem repeats of functional sequences or motifs inspired by natural proteins and organized in a modular (multi-block) architecture. They can be produced using recombinant DNA technology such that they are perfectly monodisperse. Expression in yeast allows for the production of large amounts of these materials, and designs have been developed that self-assemble into a range of precisely defined nanoarchitectures.[14-17] So far, these novel bio-inspired and biosynthetic polymers have not been used to create biocompatible protein-metal hybrid nanostructures. Here, we use protein-based polymers to incorporate and organize metal atoms into highly ordered supramolecular nanostructures with rod-shaped virus-like shapes. These protein-metal nanorods have superior stability as compared to previous micellar structures with metal ion cores, and can be prepared in aqueous solutions under mild conditions.

7.2 Materials and Methods

7.2.1. Materials. Protein C_4-S_{10}-B^{K12} was biosynthesized by recombinant *Pichia pastoris* yeast strains carrying the synthetic gene encoding for C_4-S_{10}-B^{K12} protein as reported somewhere else. Protein was purified from the medium where was secreted during production phase according to reported in literature.[13] Integrity and correct molecular weight was verified by MALDI-TOF and SDS-PAGE. Lyophilized protein was stored in sealed polypropylene tubes till further use. Metal ion Zn(II) was obtained from $Zn(NO_4)_2{\cdot}4H_2O$ (analytical grade), which was purchased from Aldrich and used without further purification. The bis-ligand compound 1,11-bis(2,6-dicarboxypyridin-4-yloxy)-3,6,9-trioxaundecane (L_2EO_4) was prepared following the reported in literature.[18]

7.2.2. Sample Preparation. $[Zn\text{-}L_2EO_4]^{2-}$ was prepared mixing equimolar concentrations of $Zn(NO_4)_2{\cdot}4H_2O$ and bis-ligand L_2EO_4 for a final concentration of 5 mM (stock solution). The mixture Zn-L_2EO_4 + C_4-S_{10}-B^{K12} was prepared in the following way. A fixed amount of protein solution recently prepared was mixed with $[Zn\text{-}L_2EO_4]^{2-}$, prepared as reported previously for the stock solution, to the desired final +/- ratio. It was considered that each C_4-S_{10}-B^{K12} molecule carries 12 positive charges from lysines. The final concentration of protein in the mixed solution is ~0.7 mg/mL. All stock solutions are prepared in buffer 10mM phosphate, pH 7.4, containing 0.1 mM DTT and using Milli-Q water. The mixture was then filtrated (0.45 μm), sonicated for 10 min to mix well and leave it at room temperature for incubation for the desired time.

7.2.3. Light Scattering. Dynamic light scattering (DLS) was used to characterize the electrostatic complexation process. The DLS measurement was performed with an ALV light scattering apparatus, equipped with a 400 mW argon ion laser operating at a wavelength of 532 nm. Measurements were done at a detection angle of 90° (corresponding to q = 22 247 820 m^{-1}), unless stated otherwise. All measurements were performed at room temperature. The size and size distribution

are presented as the diffusion coefficient obtained from both CUMULANT and CONTIN methods.

The light scattering intensity is expressed as the excess Rayleigh ratio R_θ divided by the polymer concentration. R_θ is obtained as (1) where I_{sample} is the scattering intensity of the complex solution and $I_{solvent}$ is the intensity of the solvent. $I_{toluene}$ is the scattering intensity of toluene (the reference), and $R_{toluene}$ is the known Rayleigh ratio of toluene (2.1×10^{-2} m^{-1}).

$$R_\theta = \frac{I_{sample} - I_{solvent}}{I_{toluene}} \times R_{toluene} \times \frac{n^2_{solvent}}{n^2_{toluene}} \qquad (7.1)$$

The salt stability of the supramolecular was done mixing a portion of the sample solution with a very small portion of concentrated NaCl solution, incubated for few minutes and carrying the DLS measure as before. The last addition of NaCl for a final concentration of ~1.4 M was incubated overnight and the measurement taken next day.

7.2.4. Atomic Force Microscopy (AFM). AFM was carry diluting sample with phosphate buffer (1x) 10-100 times and 3-5 µL of sample of the diluted sample were deposited onto a clean silicon surface. After 2 minutes the surface was rinsed carefully with approximately 1 mL of Milli-Q water in order to remove salts and non-absorbed particles, followed by soaking up of excess water using a tissue and slow drying under a N_2 stream.

7.2.5. Cryogenic Transmission Electronic Microscopy (Cryo-TEM). Samples for Cryo-TEM were prepared by depositing a 5 µL aliquot of the sample solution onto a copper grid (C-Flat 4/1 400 mesh), previously glow-discharged under low pressure. After deposition, all the samples were blotted with an automated system in a controlled environment (Vitrobot MarkIV, FEI) for at least 5 seconds to remove excess liquid from the copper grids to form a thin sample layer. Afterward, the samples were plunged into liquid ethane by the vitrobot and transferred to a cryo-stage (Gatan) precooled to −175 °C with liquid nitrogen. Imaging was carried

out at 200 kV on a JEOL JEM 2100 transmission electron microscope, and 4K x 4K images were recorded under low-dose conditions on a CCD camera (Gatan, US4000). During imaging, the temperature of the sample holder was kept at −175 °C by liquid nitrogen to prevent crystal formation of vitrified water.

7.2.6. Measurement of Length Distributions. ImageJ was used to measure the length of the rod-like nanostructures from AFM images.

7.3 Results and Discussion

Our approach is based on a triblock protein-based polymer with the abbreviated structure C_4-S_{10}-B^{K12} that was originally designed to co-assemble with DNA into rod-shaped virus-like particles. In the present work we investigate its co-assembly with Zn- L_2EO_4 anionic supramolecular polymers, consisting of L_2EO_4 bis-ligands, coordinated by Zn(II) ions. Together with the C_4-S_{10}-B^{K12} triblock protein-based polymers, the Zn coordinated supramolecular polymers assemble in a highly cooperative fashion, into highly organized protein-metal nanorods, as illustrated schematically in Figure 7.1. Each of the three blocks of the C_4-S_{10}-B^{K12} protein-based polymer encodes a specific physical-chemical functionality that can also be recognized in natural viral coat proteins: the "C" block is included for colloidal stabilization of the nanostructures, the "B" for cationic binding (12 positive charges) of the protein-based polymer to the template,[13] and the silk-like "S_{10}" blocks for side-ways interaction between the artificial coat proteins, in order to give rise to cooperative co-assembly with the template (full sequences in supplementary information, table S1). Multimers of the silk-like "S" sequence fold into so-called beta-rolls that stack into stiff, elongated fibers through hydrophobic beta-roll stacking.[12, 19] In previous work we have found that a length of 10 repeats of the silk-like sequence is optimal for cooperative co-assembly (chapter 3). The negatively charged coordination polymer "Zn- L_2EO_4" is formed between Zinc (II) and the coordination ligand "L_2EO_4" [1,11-bis(2,6-dicarboxypyridin-4-yloxy)-3,6,9-trioxaundecane], which chelates a range of metal ions.[20] At sufficiently high, but

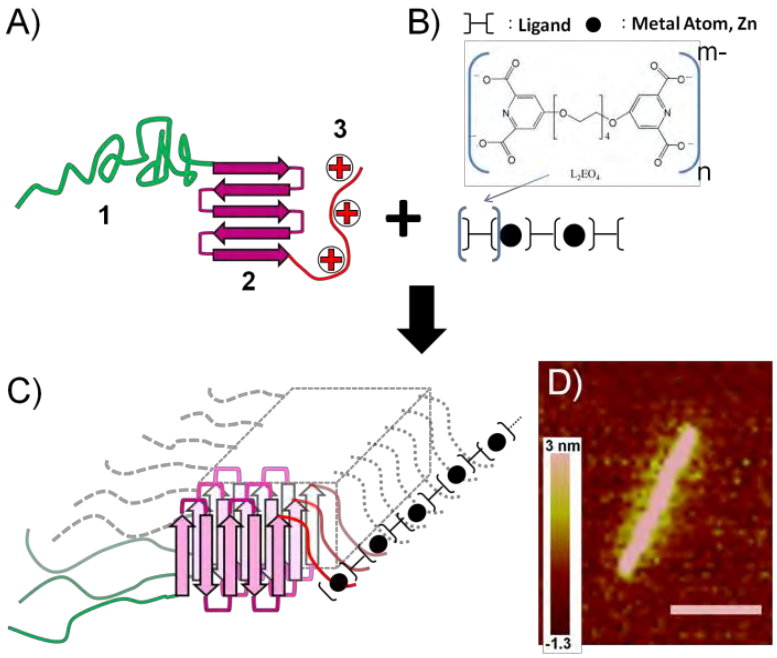


Figure 7.1. A) Cationic protein-based polymer C_4-S_{10}-B^{K12}, 1: Colloidal stability domain "C", 2: self-assembly domain "S_{10}" and 3: cationic domain "B". B) Anionic supramolecular polymer "Zn-L_2EO_4". C) Sketch of metal-protein hybrid nanomaterials formed by electrostatic co-assembly. D) AFM image of highly ordered virus-like C_4-S_{10}-B^{K12} + Zn-L_2EO_4 protein-metal nanorods. Scale bar is 100 nm.

equimolar concentrations, Zn(II) and L_2EO_4 form long, supramolecular coordination polymers.[20, 21] Instead of Zn(II), (the model metal ion chosen here), other transition and lanthanide metals could also have been used.[6, 22] For the experiments described below, we have started at low, but equimolar concentrations of Zn(II) ions and L_2EO_4 bis ligands, where supramolecular polymers do not yet form. When such a sample is mixed with a solution of the C_4-S_{10}-B^{K12} protein-based triblock copolymer at a charge ratio +/- =1 (where we divide the total positive charge of cationic binding blocks by the total negative charge of hypothetical polyanionic coordination polymers), a 6-fold increase in the light scattering intensity was detected, as compared with a control solution composed of only the protein-based polymer (Figure 7.2a).

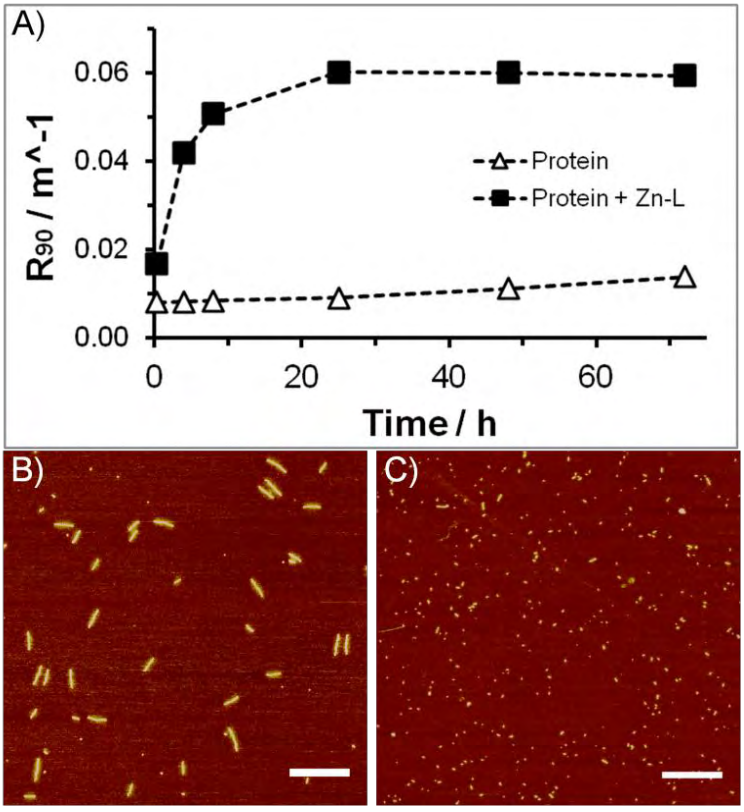


Figure 7.2. A) Intensity of scattered light (Rayleigh ratio R_{90} at a 90 degrees scattering angle, in m^{-1}) versus co-assembly time in h of C_4-S_{10}-B^{K12} + Zn-L_2EO_4 at a charge ratio +/- = 1 (black squares) and protein-only control C_4-S_{10}-B^{K12} (white triangles) at concentration of ~0.7 mg/mL. B) AFM image after 24h of incubation for C_4-S_{10}-B^{K12} + Zn-L_2EO_4 solution and C) protein-only control C_4-S_{10}-B^{K12} after 24h. Scale bars are 500 nm.

Dynamic light scattering corroborates that the increase in scattering is caused by the formation of nanoscale self-assembled objects, with hydrodynamic radii in the range 50-100 nm (Appendix Figure 7.6). Note that scattering from the initial equimolar solution of Zn-L_2EO_4 was so weak that no nanostructures could be detected by DLS, which implies that at these low concentrations, supramolecular polymers do not yet form by themselves.[20, 23] Likewise, using DLS we have also confirmed that at the concentrations used for the co-assembly, the protein polymers C_4-S_{10}-B^{K12} were unable to self-assembly into long stacks as they do at

much higher concentrations (chapter 8). Next, we have performed AFM imaging of the nanostructures formed during co-assembly (C_4-S_{10}-B^{K12} + Zn-L_2EO_4 at +/- = 1) after approximately 24h of incubation at room temperature. We find elongated virus-like nanostructures (Figure 7.1c, 7.2b and Appendix Figure 7.9). On the other hand, the C_4-S_{10}-B^{K12} protein polymer without added coordination polymer only shows very small structures, confirming that by itself, at these low concentrations, it does not yet self-assemble into large supramolecular aggregates (Figure 7.2c and Appendix Figure 7.9).

Our results are evidence for electrostatically driven co-assembly of the triblock protein polymers C_4-S_{10}-B^{K12} with the metal coordinated bis-ligands. Presumably, the initial electrostatic binding of short negative Zn-L_2EO_4 segments to the cationic domains in the C_4-S_{10}-B^{K12} monomers, nucleates the self-assembly of the protein-metal nanorods. Such a nucleation step is required for the stacking and elongation of the self-assembly domain, as was shown before.[24] At the same time, the fiber growth most likely also facilitates the growth of short Zn-L_2EO_4 segments into long negatively charged supramolecular polymers, by increasing the local concentration of the components of the coordination polymers, as has also been suggested to be the case for the micellar systems.[7, 25] Hence we find that polymerization of the supramolecular polymer and the elongation of the protein-metal nanorods go hand-in-hand, and that co-assembly can already take place at very low concentrations, where neither of the building blocks (protein polymer and metal-ligand) self-assembles by itself.

Next, we followed co-assembly as a function of incubation time, using AFM (see Appendix Figure 7.7). After 2h of incubation, only very small structures without a well-defined structure were detected. After 8 hours, some elongated nanostructures were already detected. Consistent with DLS (see Figure 7.2a), the protein-metal rods reached their maximum lengths after approximately 24h. Note that other protein-polymers have also been used for co-assembly with metal coordination polymers, but these did not form the extremely regular protein-metal nano rods, with virus-like shapes, that we find here.[26] The length distribution of the nano rods

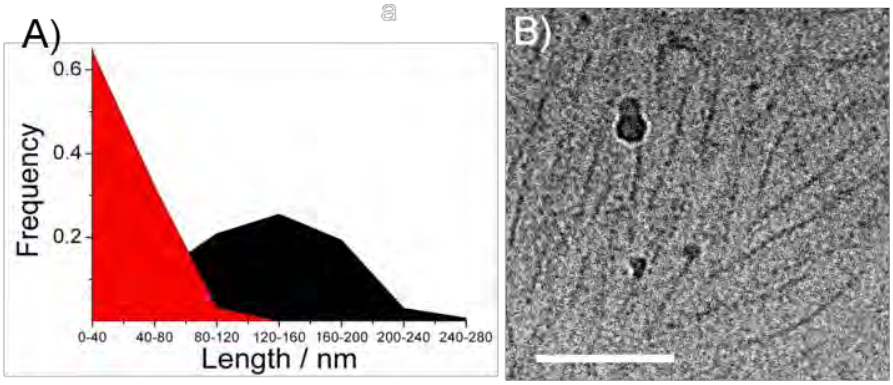


Figure 7.3. A) Plot of the length distributions as determined from AFM for C_4-S_{10}-B^{K12} + Zn-L_2EO_4 at +/- = 1 (black) and protein-only control C_4-S_{10}-B^{K12} (red) after 24h of incubation. B) Cryo-TEM of image of the virus-like protein polymer-metal co-assembled nanostructures. Scale bar is 150 nm.

was determined from AFM images, and is shown in Figure 7.3a. We find that most of the metal-protein virus-like particles have a length of 120-160 nm, with the longest lengths being found are 240-280 nm.

This was further confirmed by cryo-TEM (Figure 7.3b and Appendix Figure 7.8). Interestingly, the maximum scattering intensity is already found at a coordination polymer to protein polymer charge ratio around +/- = 0.5 (Figure 7.4). This means that the co-assembly process between C_4-S_{10}-B^{K12} and Zn-L_2EO_4 does not require the neutralization of all the negative charges of the supramolecular polymer.

With an eye to potential applications, we have finally also tested the stability of the protein-metal nano-rods against changes in solution conditions. First, complexes were incubated with increasing concentrations of NaCl, while measuring the intensity of scattered light. No significant variations in the light scattering intensity were detected even after an incubation for at least 15h at [NaCl] = 1.4 M, which implies that the protein-metal nanorods, once formed, are extremely salt-stable (Figure 7.5). Likewise, protein complexes alone remain stable for at least 4h after sudden dilution by a factor of 500 times with buffer (Chapter 6). We have not tested assembly of protein nano-rods at high salt and at high dilutions, hence the observed stability can be either kinetic or thermodynamic.

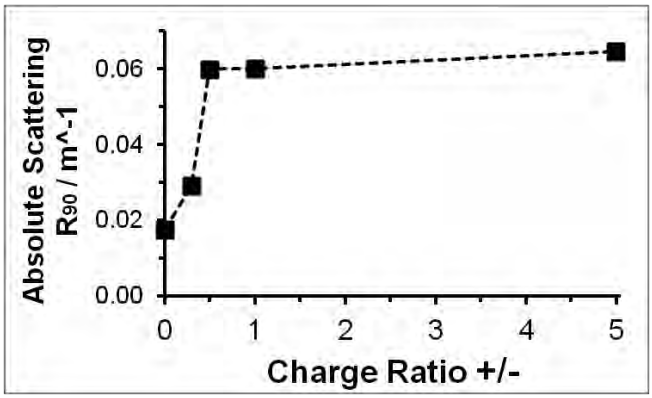


Figure 7.4. Absolute light scattering titration at different cationic protein to negative Zn-L_2EO_4 (+/-) ratio

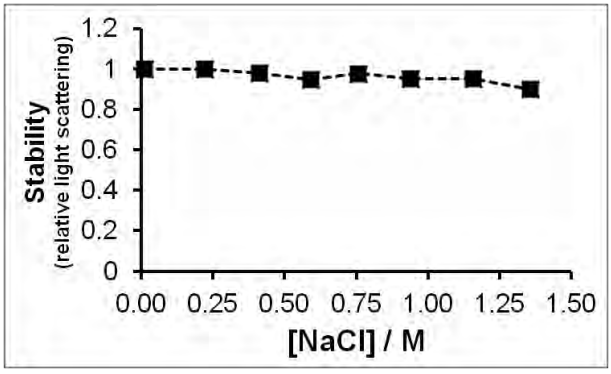


Figure 7.5. Salt stability of supramolecular complexes C_4-S_{10}-B^{K12} + Zn-L_2EO_4 at +/- = 1. Scattering in relation to the initial protein-metal solution without any salt addition.

Kinetic stability of protein-metal rods (once formed) would be consistent with the observation that their assembly kinetics is also very slow (structures equilibrate after more than 20 hours).

Protein and peptide building blocks such as used in our multi-block protein copolymer C_4-S_{10}-B^{K12} allow for extremely good control over the self-assembly of nanostructures,[5] and we have shown that this can be extended to protein-metal composite nanostructures. Our electrostatically direct co-assembly approach can be used with all other metal transition metals that are able to form coordination

polymers with the L_2EO_4 bis ligands, such as Eu^{3+}, Gd^{2+}, Fe^{2+}, Nd^{3+}, La^{3+}, etc.,[3, 6, 8, 20, 23] making our approach highly versatile. The known biocompatibility of our protein polymers (see chapter 3) implies important advantages of our approach over other synthetic approaches, when it comes to medical applications. It addition, it has been shown that has elongated nano particles can have significantly longer circulation times than spherical nanoparticles of similar average size.[27] Our co-assembly approach could be used to create biocompatible luminescent nano-rods to be used as tracers of biosensors when Europium (III)[8, 22] is used to coordinate the bis-ligands. Alternatively, protein nano-rods with Gadolinium(II) cores could be very useful as magnetic nanoprobes,[8] in MRI contrast agent applications. Finally, the recombinant origin of our protein-based polymers allows for straightforward fine-tuning and the possibility to add extra functional domains such as for targeting.

7.4 Conclusion

We have demonstrated a versatile approach for creating self-assembled protein-metal nanorods that are highly regular, very stable and biocompatible. Our results illustrate the usefulness of the multiblock protein-polymer design methodology for designing novel biomacromolecules that assemble and co-assemble into well-defined, functional nanostructures, with many potential applications.

7.5 References

1. Lu, G.; Li, S. Z.; Guo, Z.; Farha, O. K.; Hauser, B. G.; Qi, X. Y.; Wang, Y.; Wang, X.; Han, S. Y.; Liu, X. G.; DuChene, J. S.; Zhang, H.; Zhang, Q. C.; Chen, X. D.; Ma, J.; Loo, S. C. J.; Wei, W. D.; Yang, Y. H.; Hupp, J. T.; Huo, F. W., Imparting functionality to a metal-organic framework material by controlled nanoparticle encapsulation. *Nat. Chem.* 2012, *4*, 310-316.

2. Lopes, W. A.; Jaeger, H. M., Hierarchical self-assembly of metal nanostructures on diblock copolymer scaffolds. *Nature* 2001, *414*, 735-738.

3. Wang, X. S.; McHale, R., Metal-Containing Polymers: Building Blocks for Functional (Nano)Materials. *Macromol. Rapid Commun.* 2010, *31*, 331-350.

4. Carne, A.; Carbonell, C.; Imaz, I.; Maspoch, D., Nanoscale metal-organic materials. *Chem. Soc. Rev.* 2011, *40*, 291-305.

5. Aida, T.; Meijer, E. W.; Stupp, S. I., Functional Supramolecular Polymers. *Science* 2012, *335*, 813-817.

6. Yan, Y.; Lan, Y. R.; de Keizer, A.; Drechsler, M.; Van As, H.; Stuart, M. A. C.; Besseling, N. A. M., Redox responsive molecular assemblies based on metallic coordination polymers. *Soft Matter* 2010, *6*, 3244-3248.

7. Yan, Y.; Besseling, N. A. M.; de Keizer, A.; Marcelis, A. T. M.; Drechsler, M.; Stuart, M. A. C., Hierarchical self-assembly in solutions containing metal ions, ligand, and diblock copolymer. *Angew. Chem.-Int. Edit.* 2007, *46*, 1807-1809.

8. Wang, J.; Velders, A. H.; Gianolio, E.; Aime, S.; Vergeldt, F. J.; Van As, H.; Yan, Y.; Drechsler, M.; de Keizer, A.; Cohen Stuart, M. A.; van der Gucht, J., Controlled mixing of lanthanide(iii) ions in coacervate core micelles. *Chemical Communications* 2013, *49*, 3736-3738.

9. Ding, Y.; Yang, Y.; Yang, L.; Yan, Y.; Huang, J. B.; Stuart, M. A. C., A Case of Adaptive Self-Assembly. *Acs Nano* 2012, *6*, 1004-1010.

10. Rabotyagova, O. S.; Cebe, P.; Kaplan, D. L., Protein-Based Block Copolymers. *Biomacromolecules* 2011, *12*, 269-289.

11. Amiram, M.; Quiroz, F. G.; Callahan, D. J.; Chilkoti, A., A highly parallel method for synthesizing DNA repeats enables the discovery of 'smart' protein polymers. *Nat. Mater.* 2011, *10*, 141-148.

12. Beun, L. H.; Beaudoux, X. J.; Kleijn, J. M.; de Wolf, F. A.; Stuart, M. A. C., Self-Assembly of Silk-Collagen-like Triblock Copolymers Resembles a Supramolecular Living Polymerization. *Acs Nano* 2012, *6*, 133-140.

13. Hernandez-Garcia, A.; Werten, M. W. T.; Stuart, M. C.; de Wolf, F. A.; de Vries, R., Coating of Single DNA Molecules by Genetically Engineered Protein Diblock Copolymers. *Small* 2012, *8*, 3491-3501.

14. Yi, C. Q.; Liu, D. D.; Yang, M. S., Building Nanoscale Architectures by Directed Synthesis and Self-Assembly. *Curr. Nanosci.* 2009, *5*, 75-87.

15. Ulijn, R. V.; Smith, A. M., Designing peptide based nanomaterials. *Chem. Soc. Rev.* 2008, *37*, 664-675.

16. Zhang, S. G., Fabrication of novel biomaterials through molecular self-assembly. *Nat. Biotechnol.* 2003, *21*, 1171-1178.

17. Carlsen, A.; Lecommandoux, S., Self-assembly of polypeptide-based block copolymer amphiphiles. *Curr. Opin. Colloid Interface Sci.* 2009, *14*, 329-339.

18. Vermonden, T.; Branowska, D.; Marcelis, A. T. M.; Sudholter, E. J. R., Synthesis of 4-functionalized terdendate pyridine-based ligands. *Tetrahedron* 2003, *59*, 5039-5045.

19. Schor, M.; Bolhuis, P. G., The self-assembly mechanism of fibril-forming silk-based block copolymers. *Phys. Chem. Chem. Phys.* 2011, *13*, 10457-10467.

20. Vermonden, T.; van der Gucht, J.; de Waard, P.; Marcelis, A. T. M.; Besseling, N. A. M.; Sudholter, E. J. R.; Fleer, G. J.; Stuart, M. A. C., Water-soluble reversible coordination polymers: Chains and rings. *Macromolecules* 2003, *36*, 7035-7044.

21. Schwarz, G.; Bodenthin, Y.; Geue, T.; Koetz, J.; Kurth, D. G., Structure and Properties of Dynamic Rigid Rod-Like Metallo-Supramolecular Polyelectrolytes in Solution. *Macromolecules* 2010, *43*, 494-500.

22. Yang, L.; Ding, Y.; Yang, Y.; Yan, Y.; Huang, J. B.; de Keizer, A.; Stuart, M. A. C., Fluorescence enhancement by microphase separation-induced chain extension of Eu3+ coordination polymers: phenomenon and analysis. *Soft Matter* 2011, *7*, 2720-2724.

23. Vermonden, T.; de Vos, W. M.; Marcelis, A. T. M.; Sudholter, E. J. R., 3-d water-soluble reversible neodymium(III) and lanthanum(III) coordination polymers. *Eur. J. Inorg. Chem.* 2004, 2847-2852.

24. Ni, R.; Abeln, S.; Schor, M.; Cohen Stuart, M. A.; Bolhuis, P. G., Interplay between Folding and Assembly of Fibril-Forming Polypeptides. *Physical Review Letters* 2013, *111*, 058101.

25. Stuart, M. A. C.; Besseling, N. A. M.; Fokkink, R. G., Formation of micelles with complex coacervate cores. *Langmuir* 1998, *14*, 6846-6849.

26. Yan, Y.; Martens, A. A.; Besseling, N. A. M.; de Wolf, F. A.; de Keizer, A.; Drechsler, M.; Stuart, M. A. C., Nanoribbons self-assembled from triblock peptide polymers and coordination polymers. *Angew. Chem.-Int. Edit.* 2008, *47*, 4192-4195.

27. Geng, Y.; Dalhaimer, P.; Cai, S. S.; Tsai, R.; Tewari, M.; Minko, T.; Discher, D. E., Shape effects of filaments versus spherical particles in flow and drug delivery. *Nat. Nanotechnol.* 2007, *2*, 249-255.

7.6 Appendix

Table 7.1. Amino acid sequence of protein-based polymer C_4-S_{10}-B^{K12}

GCGLGAGAPGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQGNPGKNGQPGSPGSQGSPGNQG SPGQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQGNPGKNGQPG SPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSNGPQG SQGNPGKNGQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNG QPGEPGSNGPQGSQGNPGKNGQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGA**GAGAGAGQGAGAGAGQGAGA GAGAGQGAGAGAGQGAGAGAGQGAGAGAGQGAGAGAGQGAGAGAGQGA**KKKKKKKKKKKKGA

Codes: colloidal stability "C_4" block, **self-assembly "Sn" block** and *cationic "B^{K12}" block.*

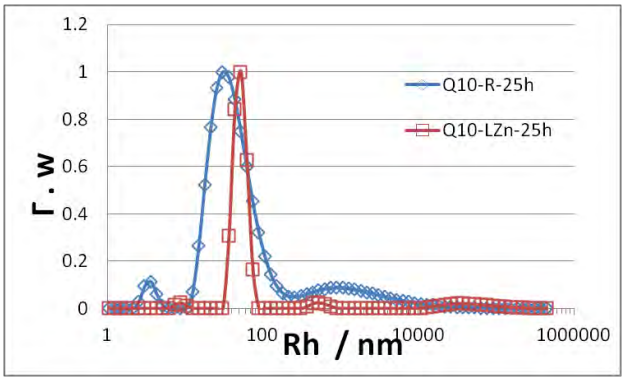


Figure 7.6. Hydrodynamic radius of C_4-S_{10}-B^{K12} incubated with negative coordination polymer Zn-L_2EO_4 at +/- = 1 and C_4-S_{10}-B^{K12} alone control after 25h.

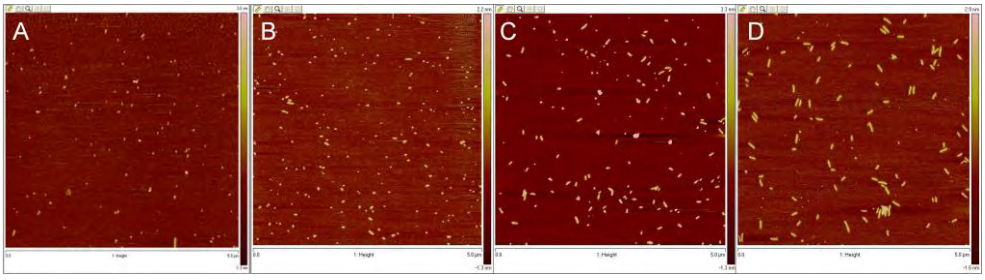


Figure 7.7. Kinetics of supramolecular electrostatic co-assembly. AFM images at different incubation times of solution C_4-S_{10}-B^{K12} + Zn-L_2EO_4. Incubation times are (a) 20min, (b) 1h, (c) 4h and (d) 30h. Images are 5x5μm. Assemblies have same height (~3 to 4 nm).

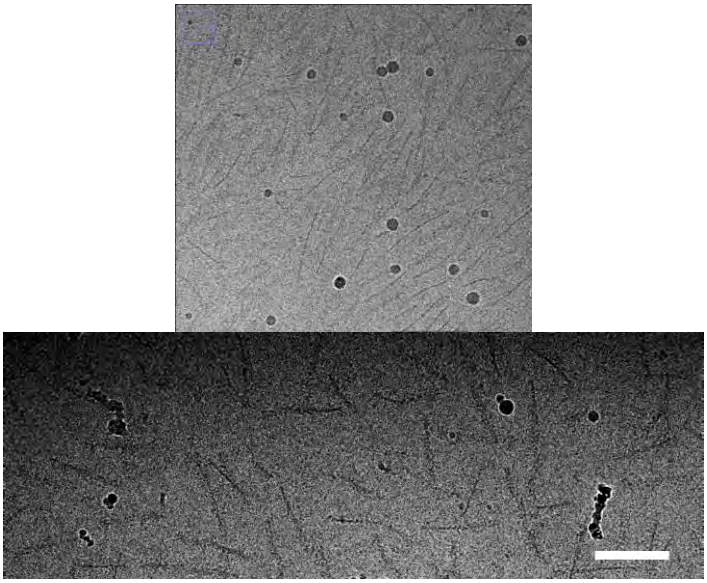


Figure 7.8. Cryo-TEM image of SQ10 complexed with negative coordination polymer L-Zn (N/P = 1) with similar length sizes that observed in AFM. Scale bars are 200 nm.

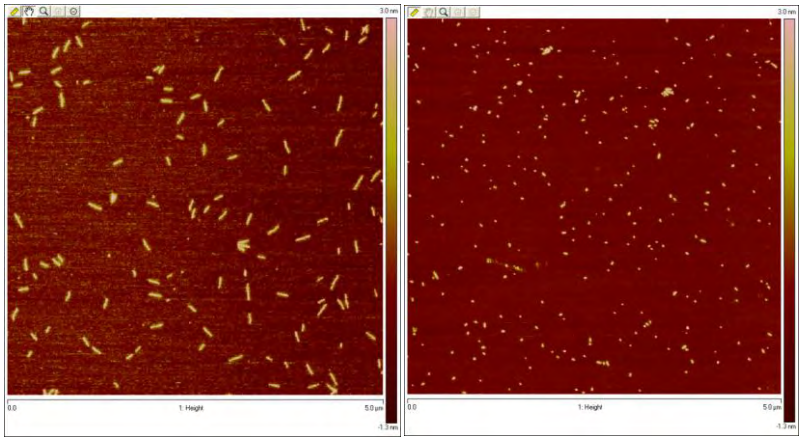


Figure 7.9. Atomic Force Microscopy of supramolecular hybrid metal-protein nanostructures. Left, C_4-S_{10}-B^{K12} polypeptide incubated with negative coordination polymer Zn-L_2EO_4 at +/- = 1 for 30h. Right, C_4-S_{10}-B^{K12} polypeptide alone.

Chapter 8

Templating of artificial viral coat proteins by a range of negatively charged polyelectrolytes[1]

[1] This chapter is based on Armando Hernandez-Garcia, Martien Cohen Stuart and Renko de Vries Templating of artificial viral coat proteins by a range of negatively charged polyelectrolytes, Manuscript in preparation. 2013.

8.1 Introduction

The self-assembly of protein block copolymers has been used to build a range of nanostructured supramolecular assemblies that could be used in nanomedicine and material science.[1, 2] We have recently designed and produced protein polymers with the abbreviated structure "C_4-S_n-B^{K12}" that pack single molecules of negatively charged DNA into regular virus-like supramolecular nanostructures (chapter 3). These are composed of three functional domains: "C_4" is a hydrophilic random coil of about 400 amino acids for colloidal stability,[3] "B^{K12}" a cationic segment made up of 12 lysines[1] and "S_n" is an octapeptide hydrophobic silk-like sequence repeated "n" times that folds into beta-roll secondary structure, and stacks into fibers, as reported for similar sequences.[2] These novel protein polymers, that mimic viral coat proteins in many ways, are highly promising to be developed as nanocarriers for gene delivery. The artificial virus structures can also be used for templating nanomaterials.[1, 4] While we have extensively characterized the co-assembly of these novel protein polymers with DNA, it may be expected that similar co-assembly should be possible with other negatively charged polyelectrolytes, since the oligolysine binding is not at all specific to DNA. Here we therefore investigate the co-assembly of the artificial viral coat proteins with a range of other anionic polyelectrolytes. We use atomic force microscopy (AFM) to characterize at single molecule level the supramolecular complexes formed by the artificial virus capsid proteins C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12} with a range of anionic polyelectrolytes: poly-acrylic acid (PAA) of different lengths, poly-styrene sulfonate (PSS) and Xanthan. This will allows discussing how parameters such as charge density, template contour length and stiffness influence the co-assembly and final architecture of the complexes.

In this way, we can arrive at a better understanding of the co-assembly process[5] of our artificial virus capsid proteins with polyanionic templates, a strategy that has already been shown to be very fruitful for studying the co-assembly of natural viral capsid proteins with synthetic polyelectrolytes.[6-12] In addition, nanorods templated with other anionic polyelectrolytes may be used in similar ways as nanorods

templated with nucleic acids (be it natural, or artificial rod-like virus structures), to create nanostructures with a range of applications in biomedical and material sciences.[13]

8.2 Materials and Methods

8.2.1. Materials. Na-PAA with molar masses (M_W) of 150,000–800,000 g mol^{-1} was purchased from Polysciences, Inc. Reported polydispersities $D_P = M_W/M_N$ range between 1.3 and 2.15. Poly(styrene sulfonic acid sodium salt) or just PSS Mw = 263 000 g/mol with $D_P = M_W/M_N = 1.04$ was bought from Polymer Source Inc. Xanthan gum from *Xanthomonas campestris wa*s bought from Sigma-Aldrich. The cationic protein polymers C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12} were produced biosynthetically by recombinant yeast *Pichia pastoris* and purified as described in chapter 3.

8.2.2. Sample Preparation. Aqueous solutions of polyelectrolytes and C_4-S_{10}-B^{K12} or C_4-S_{10}-B^{K12} (up to 5 g L^{-1}) were prepared by mixing aliquots of the polymers to the final desired bulk charge ratio (+/−) in 10 mM phosphate buffer pH 7.4 and 0.1 mM dithiothreitol (DTT) solution. All the solutions were incubated overnight (17-23h) at room temperature before imaging. The ratio +/− is defined as a total number of positively charged lysine groups of the *B* domain over the total number of (negatively) chargeable groups from polyelectrolytes. Final concentration of PAA and PSS was 4.0 ug/mL when complexed with C_4-S_{10}-B^{K12} (1.56 mg/mL and 0.616 mg/mL, respectively) at +/- = 7.5. When complexed to C_4-S_{14}-B^{K12} (0.985 and 0.387 mg/mL, respectively) at +/- = 7.5, PAA and PSS were 2.4 μg/mL. The concentration of Xanthan was 2 μg/mL when complexed with 1.56 mg/mL C_4-S_{10}-B^{K12}.

8.2.3. Atomic Force Microscopy (AFM). Samples for AFM were prepared by dropping 5 μl of sample on a clean silicon substrate and left for 5 min followed by washing with 1 ml of filtrated Milli-Q water to removed salts and non-absorbed particles. Finally the sample was slowly dried under a nitrogen steam. Multiple images were collected at 1024 samples per line at scan rate of 0.977 Hz using a

Digital Instruments NanoScope V and a silicon nitride cantilevers (Veeco, NY, USA) with a nominal spring constant of 0.32 N m^{-1}. The ScanAsyst™ image mode was used in air at room temperature.

8.2.4. Image Analysis. Images were processed using NanoScope Analysis 1.20 software and also to measure the height. Length of complexes was measured with ImageJ software.

8.3 Results and Discussion

8.3.1. Flexible Anionic Polyelectrolytes as Templates. First we consider co-assembly of the flexible and anionic poly-acrylic acid (PAA) with the cationic, artificial viral coat protein polymers C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12}. AFM images of the final co-assembled structures are shown in Figure 8.1. For both proteins we find a polydisperse mixture of small micelle-like structures (that sometimes appear be slightly elongated) and long and elongated rod-like structures. The small (rod-like) micelles correspond to self-assembled proteins as has been discussed in previous chapters; see the control image in Figure 8.1c. This is expected since we have used a large excess of protein in order to obtain coat the PAA. The polydispersity of the long rods clearly reflects the polydispersity of the underlying PAA template, but we cannot exclude that several PAAs molecules may be involved in individual rods. Structures formed by C_4-S_{10}-B^{K12} appear as somewhat less structured than those formed with C_4-S_{14}-B^{K12}. The cross-section of the rods in most cases appears to be quite homogeneous, although for C_4-S_{10}-B^{K12} some sections of the complexes look thicker than others. Occasionally we find large clusters of several connected rods (data not shown in image). This may be due to the large excess of protein (+/- = 7.5) since a PAA solution incubated with C_4-S_{10}-B^{K12} using less protein (charge ratio +/- = 1) did not show such clusters (Appendix Figure 8.6).

Previously, we have studied the formation of electrostatic complexes between long PAA (MW = $8x10^5$ Da) with a cationic diblock protein polymer C_4-B^{K12}, lacking the silk-like self-assembly block. For that case, single-chain complexes were

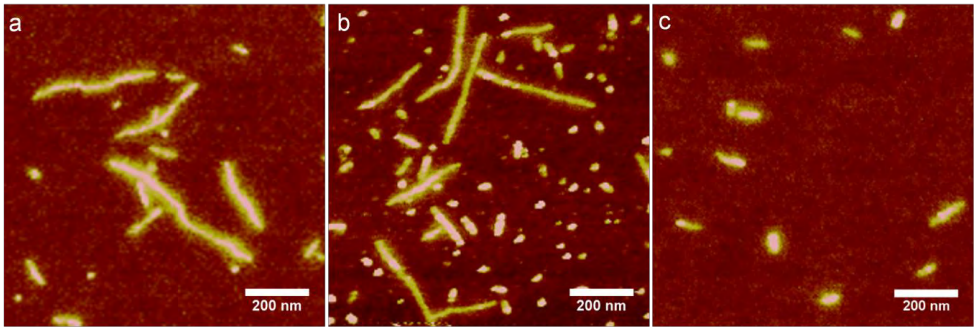


Figure 8.1. Co-assembly of polyacrylic acid (PAA) M_W = 8.0x10^5 Da with cationic protein-based polymers. A) C_4-S_{10}-B^{K12} and b) C_4-S_{14}-B^{K12}. c) Control C_4-S_{10}-B^{K12} without any PAA. Complexes formed at +/-= 7.5 after 24 hrs of incubation at room temperature.

found with a pearl-necklace structure of connected spherical micelles.[14] In that case, the local structure of the micelles is basically dictated by the balance between the long steric "C" block and the short cationic binding block (B, 12 lysines). In contrast, for the artificial viral coat proteins that additionally have the silk-like self-assembly block, the final structures seem to be dictated by the stacking of the silk blocks.

In order to analyze in detail the effect of the polyelectrolyte length on the co-assembled structures, PAA molecules of different lengths were co-assembled with the C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12} proteins. High magnification AFM images of the resulting structures for PAA with MW of 1.5 to 8.0x10^5 Da with SQ10 (a-d) and with SQ14 (e-h) are shown in Figure 8.2. In both cases the length of the complex increases with the length of the PAA as expected for a complexation of single polyelectrolyte chains, hence it seems unlikely that complexes involve multiple PAA chains, as was also concluded for the case of PAA complexes with the C_4-B^{K12} diblock copolymer.[14] In all cases the final length of the rod-like complexes is shorter than the average expected contour length of the PAA, indicating some form of packing of the PAA chains inside the rods, as was also found for DNA (chapter 3). The heights of the PAA-rods was measured from AFM images, and it was found that there is a very well defined average height, that was between 2.7-2.8 nm for C_4-S_{10}-B^{K12}/PAA rods and 2.5-3.0 nm for C_4-S_{14}-B^{K12}/PAA rods. A summary of other

measured values is shown in Table 8.1, while measured heights can be found in Table 8.2 of the Appendix. The height of micelles in a C_4-S_{10}-B^{K12} protein-only control sample was approximately 2.6 nm, closely matching the height of approximately 2.8 nm of self-assembled S_n domains (with beta-roll or beta sheet secondary structure) as determined using small-angle X-ray scattering experiments and computer simulations.[2, 15, 16] The fact that protein-PAA complexes and protein structures have the same height strongly suggests that the structure of the complexes is basically that of a stack of beta-roll or beta-sheet silk-like blocks, with the PAA wrapping the binding blocks B that are located right next to the silk-stack. As for the case of DNA, binding to PAA apparently triggers the self-assembly of the silk-like S_n domain, and the PAA template length controls the length of the final supramolecular complex structure, very similar to what was observed for DNA. In the final structures, the silk-like core, surrounded by the thin layer of complex coacervate, is of course stabilized by the very large bottle-brush corona formed by the C domains. It appears as if the width of C_4-S_{14}-B^{K12} rods is significantly less than that of C_4-S_{10}-B^{K12} rods (xy-axis). This could suggest that the two proteins co-assemble with PAA in somewhat different ways. Also, as discussed above, for the case of long PAA ($8.0x10^5$ Da) co-assembled with C_4-S_{10}-B^{K12}, some segments appeared to be wider than others, and this could possibly point to local backfolding of silk-stack.

Despite the large polydispersity of the PAA samples, we have attempted to estimate a packing factor (contour length of template/contour length of rods), as we have done for the case of DNA. Lengths of complexes of PAA of 2.5 and $8.0x10^5$ Da with C_4-S_{10}-B^{K12} were measured from AFM images, and compared with sizes of micelles for a protein-only control sample. A plot showing the distributions of the measured lengths is shown in Figure 8.3. The distribution of the protein-only control sample corroborates that in this case only short (rod-like) micelles are formed with lengths <100 nm (~85% of the population). For the PAA complexes (formed in the presence of excess protein) these micelles were also present, but these were not taken into account in measuring the length distribution of the complexes (this was done by ignoring all structures with lengths <100 nm).

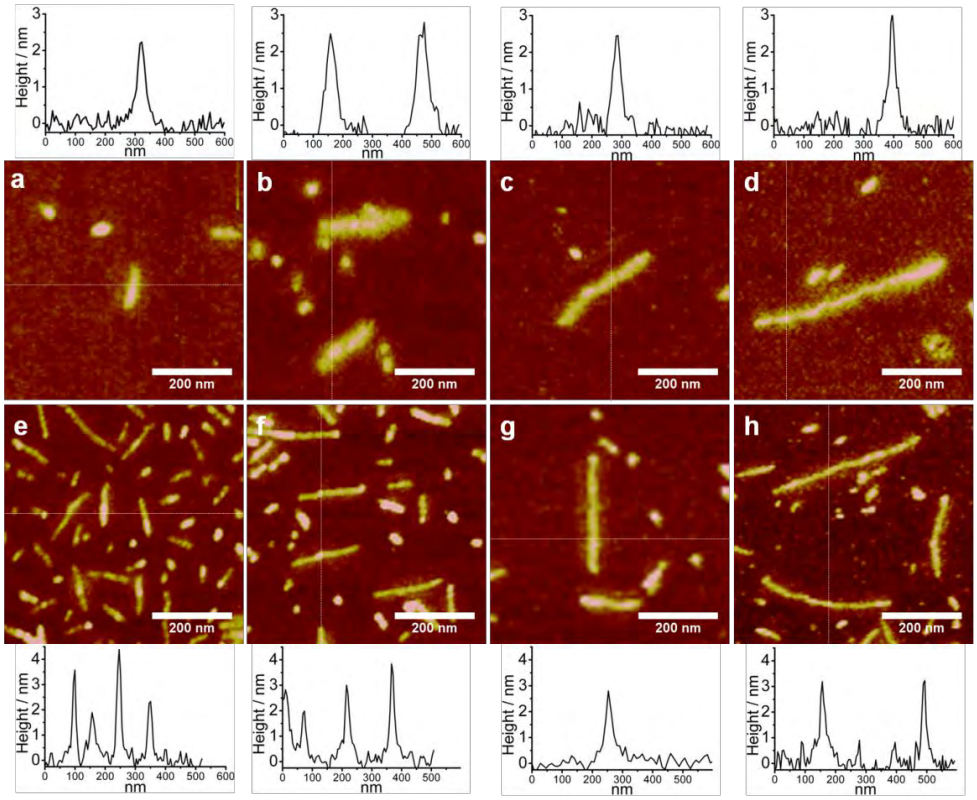


Figure 8.2. Typical supramolecular structures formed between proteins and PAA of different sizes. Top row are complexes with C_4-S_{10}-B^{K12} (a-d) and bottom row are with C_4-S_{14}-B^{K12} (e-h). PAA sizes are: 1.5 (a and e), 2.5 (b and f), 5.0 (c and g) and 8.0x10^5 Da (d and h). Plots next to images are height profiles of section indicated with a dashed line.

In view of the large polydispersity of the PAA samples, we have used a relatively large bin size of 100 nm.

The length distribution of complexes for the PAA $8.0x10^5$ Da sample shows two peaks, one at with average length of 291 ± 80 nm and a second with an average length of 649 ± 144 nm. Both are longer than the single peak observed for distribution of PAA 2.5 complexes (average length of 201 ± 73 nm). Next we estimated packing factors, assuming each rod is formed from a PAA with a mass equal to the average mass of the distribution. For PAA of $8.0x10^5$ Da, this gives

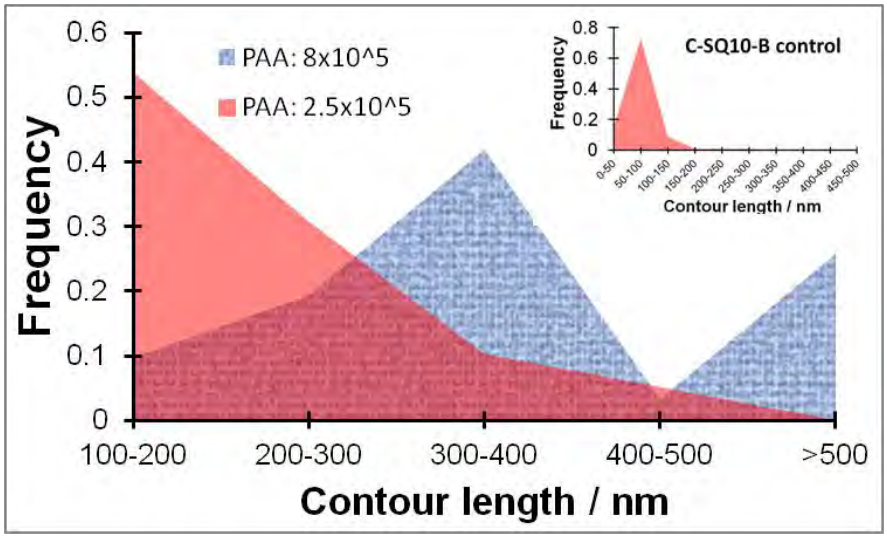


Figure 8.3. Length distributions of complexes C_4-S_{10}-B^{K12} with PAA of 2.5 and 8.0x10^5 Da. The distribution of complexes with short PAA (2.5x10^5 Da) is monomodal, whereas that for te longer PAA (8.0x10^5 Da) appears to be bimodal.

packing factors of 10 and 4.8, respectively, for peak 1 and 2. For PAA of 2.5x10^5 Da, we find a packing factor of 4.8. Possibly, the largest PAA has a higher average packing factor due to the possible back folding that we observed in AFM, as deduced from the observation that some complex sections were thicker than others (Appendix Figure 8.7).The second peak for the PAA 8.0x10^5 Da distribution (649 ± 144 nm) could correspond to non back folded structures, and indeed gives a packing factor of 4.8, equal to that for the shorter PAA. This packing factor corresponds to a charge density of approximately ~12 negative charges (required to neutralize a single binding block B) for 0.6 nm of contour length of the rod. This value is close to values for the height of a single silk-like self-assembly block in the stack: inter-protein distances found for related silk-like domains are 0.48 nm for a fully folded beta-roll and 0.7 to 0.8 nm for a partially unfolded meta-stable state.[2, 16] Like for the case of DNA, it appears that charge neutralization fully determines the binding stoichiometry. Indeed, for that case we found that whereas a double stranded DNA molecule has a packing factor of 3, reducing the linear charge

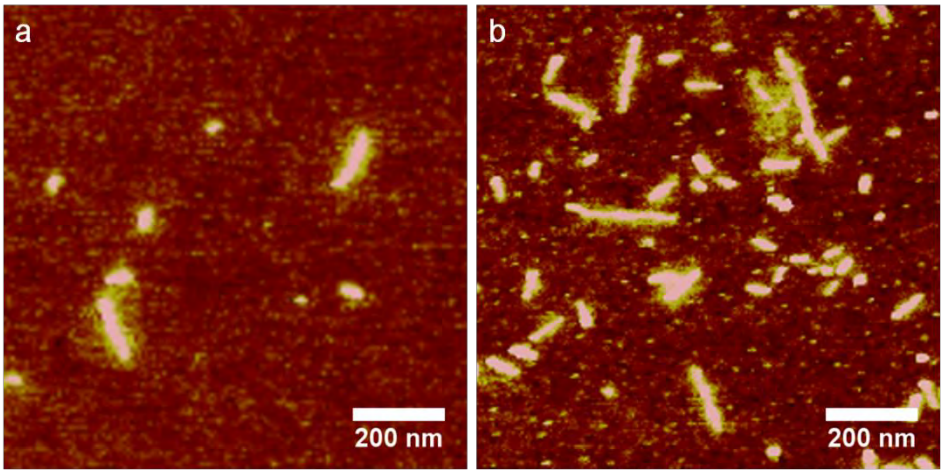


Figure 8.4. Complexation of poly-styrene sulfonate (MW = 2.6x10^5 Da) with C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12}.

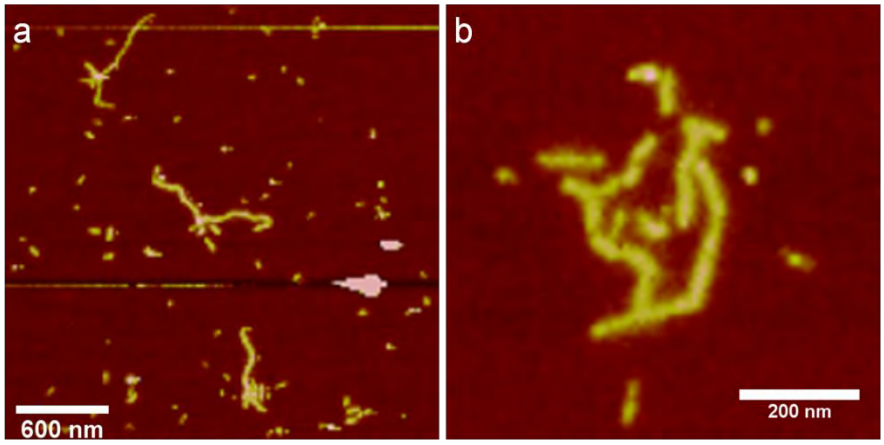


Figure 8.5. Complexation of xanthan complexed with protein C_4-S_{10}-B^{K12}. a) Large and b) zoom in view.

density by a factor of two by going to single stranded DNA, leads to a packing factor of 6.

Next we compare the results found for PAA with those for some other anionic polyelectrolytes. For poly-styrene sulfonate (PSS), we find very similar rod-like

structures (Figure 8.4), with a maximum length found was of ~150 nm. Unfortunately, in this case, the short length of the template and the excess of protein added make it difficult to distinguish protein-only micelles from PSS/protein complexes such that a packing factor cannot be calculated.

8.3.2. Semiflexible Anionic Polyelectrolyte as Template. Finally, we also consider the case of xhantan, a semiflexible anionic polyelectrolyte with a persistence length that is significantly larger than DNA: in a recent study, a value for the persistence length of xhantan was found of 370 nm.[17] Images of structures of xhantan co-assembled with C_4-S_{10}-B^{K12} are shown in Figure 8.5. The complexes again have a rod-like character similar to the cases of PAA and PSS, but for xhantan it was much more common to find branches. These were also found in our study with DNA, more frequent for C_4-S_{14}-B^{K12} than for C_4-S_{10}-B^{K12}, and more frequently for very long DNA than for short DNA. The average height of the rods is similar to that found for PAA and PSS: 2.8 ± 0.8 nm. Most likely, rod-shaped complexes again consist of a single xhantan chain. They have an average length of 1169 ± 142 nm, a value that is quite close to the lengths of chains for a xhantan sample from the same producer, as determined using AFM.[17] We have not determined the actual contour length of the chains for our xhantan samples, hence for the moment we will simply assume that the contour length of our xhantan is indeed similar to that reported before.[17] This then would mean that opposite to the cases of DNA, and the more flexible polyanions PAA and PSS, for xhantan, the packing factor is close to one. This is in fact quite consistent with the much higher reported persistence length of 370 nm,[17] that would make it much more difficult to introduce the bending that is necessary to achieve packing factors > 1.

Hence, we conclude that there is a competition between the forces that drive packing (electrostatic complexation and the stacking of the silk-like blocks) and the elasticity of the template, which opposes packing. Apparently, if the rigidity of the template exceeds a certain value, co-assembly will still occur, but there is no packing of the template. It would be interesting to determine using more

Table 8.1. Summary of physical parameters of naked negative polyelectrolytes and its protein C_4-S_{10}-B^{K12} complexes.

	Anionic Polyelectrolytes				Complexes with protein		
	P_L (nm)	L_C (nm)	Charge density (one – per nm)	Height (nm)	L_C (nm)	Height (nm)	Packing Factor
PAA $2.5x10^5$ Da	~0.7[18]	964	0.24	nd	231 ± 82	2.7 ± 0.7	~4.17
PAA $8.0x10^5$ Da	~0.7[18]	3087	0.24	nd	404 ± 189	2.8 ± 0.7	~7.64
PSS $2.62x10^5$ Da	~18.6[19]	396	0.24	nd	nd	2.5 ± 0.6	nd
ssDNA[a] 7249 nt	~3[20]	2465	0.17	nd	~400	nd	~6
dsDNA[a] 2.5kbp	~50	850	0.17	nd	~300	2.8	~3
Xanthan	370[17]	1170[17]	nd	0.48	1170 ± 142	2.8 ± 0.8	~1
C_4-S_{10}-B^{K12} control	--	--	--	nd	76 ± 35	2.6 ± 0.7	--

P_L: Persistence length, L_C: Contour length, ND: Unknown. [a] Data from Chapter 3

quantitative binding experiments whether this also implies that co-assembly with stiff templates does not lead to full charge neutralization of the binding blocks.

8.4 Conclusion

We have shown that, similar to the case of DNA, the artificial virus capsid protein-based polymers C_4-S_{10}-B^{K12} and C_4-S_{14}-B^{K12} form elongated rod-like complexes with single chains of a range of negatively charged polyelectrolytes (PAA, PSS and xanthan). The structure of the complexes most likely is that of a stack of silk-blocks with pendant binding blocks B that are completely or partially

neutralized by the polyelectrolytes. This core structure is surrounded by a bottle-brush corona formed by the C_4-domain. Flexible polyelectrolytes, such as PAA and PSS were found to have packing factors > 1, whereas the much stiffer xanthan polyelectrolyte was presumably not packed: the length of the Xhantan complexes was equal to the (assumed) length of the Xhantan template. Hence, there is a competition between the forces that drive packing (electrostatic complexation and the stacking of the silk-like blocks) and the elasticity of the template, which opposes packing. Possibly, for stiff templates, co-assembly takes place without the full neutralization of the binding blocks that appears to take place for more flexible templates. As was done previously for natural virus capsid proteins, we here find significant mechanistic insight in the co-assembly of our artificial virus capsid proteins, by considering their interaction with a range of other anionic polyelectrolyte templates.

In addition, the present study highlights a novel way of templating anionic polyelectrolytes, thus providing a new route to prepare nanostructured materials with high aspect ratios, for a range of applications in material sciences.

8.5 References

1. Hernandez-Garcia, A.; Werten, M. W. T.; Stuart, M. C.; de Wolf, F. A.; de Vries, R., Coating of Single DNA Molecules by Genetically Engineered Protein Diblock Copolymers. *Small* 2012, *8*, 3491-3501.

2. Martens, A. A.; Portale, G.; Werten, M. W. T.; de Vries, R. J.; Eggink, G.; Stuart, M. A. C.; de Wolf, F. A., Triblock Protein Copolymers Forming Supramolecular Nanotapes and pH-Responsive Gels. *Macromolecules* 2009, *42*, 1002-1009.

3. Werten, M. W. T.; Wisselink, W. H.; van den Bosch, T. J. J.; de Bruin, E. C.; de Wolf, F. A., Secreted production of a custom-designed, highly hydrophilic gelatin in Pichia pastoris. *Protein Engineering* 2001, *14*, 447-454.

4. Ruff, Y.; Moyer, T.; Newcomb, C. J.; Demeler, B.; Stupp, S. I., Precision Templating with DNA of a Virus-like Particle with Peptide Nanostructures. *Journal of the American Chemical Society* 2013, *135*, 6211-6219.

5. Ni, R.; Abeln, S.; Schor, M.; Cohen Stuart, M. A.; Bolhuis, P. G., Interplay between Folding and Assembly of Fibril-Forming Polypeptides. *Physical Review Letters* 2013, *111*, 058101.

6. Wang, X. Y.; Deng, Y. Q.; Shi, H. Y.; Mei, Z.; Zhao, H.; Xiong, W.; Liu, P.; Zhao, Y.; Qin, C. F.; Tang, R. K., Functional Single-Virus-Polyelectrolyte Hybrids Make Large-Scale Applications of Viral Nanoparticles More Efficient. *Small* 2010, *6*, 351-354.

7. Kivenson, A.; Hagan, M. F., Mechanisms of Capsid Assembly around a Polymer. *Biophys. J.* 2010, *99*, 619-628.

8. Minten, I. J.; Ma, Y. J.; Hempenius, M. A.; Vancso, G. J.; Nolte, R. J. M.; Cornelissen, J., CCMV capsid formation induced by a functional negatively charged polymer. *Org. Biomol. Chem.* 2009, *7*, 4685-4688.

9. Yoo, P. J.; Nam, K. T.; Qi, J. F.; Lee, S. K.; Park, J.; Belcher, A. M.; Hammond, P. T., Spontaneous assembly of viruses on multilayered polymer surfaces. *Nat. Mater.* 2006, *5*, 234-240.

10. Ng, B. C.; Chan, S. T.; Lin, J.; Tolbert, S. H., Using Polymer Conformation to Control Architecture in Semiconducting Polymer/Viral Capsid Assemblies. *Acs Nano 5*, 7730-7738.

11. Kostiainen, M. A.; Hiekkataipale, P.; de la Torre, J. A.; Nolte, R. J. M.; Cornelissen, J. J. L. M., Electrostatic self-assembly of virus-polymer complexes. *Journal of Materials Chemistry 21*, 2112-2117.

12. Cadena-Nava, R. D.; Hu, Y.; Garmann, R. F.; Ng, B.; Zelikin, A. N.; Knobler, C. M.; Gelbart, W. M., Exploiting Fluorescent Polymers To Probe the Self-Assembly of Virus-like Particles. *J. Phys. Chem. B 115*, 2386-2391.

13. van Rijn, P., Polymer Directed Protein Assemblies. *Polymers* 2013, *5*, 576-599.

14. Golinska, M. D.; de Wolf, F.; Stuart, M. A. C.; Hernandez-Garcia, A.; de Vries, R., Pearl-necklace complexes of flexible polyanions with neutral-cationic diblock copolymers. *Soft Matter* 2013, *9*, 6406-6411.

15. Schor, M.; Bolhuis, P. G., The self-assembly mechanism of fibril-forming silk-based block copolymers. *Phys. Chem. Chem. Phys.* 2011, *13*, 10457-10467.

16. Schor, M.; Martens, A. A.; Dewolf, F. A.; Stuart, M. A. C.; Bolhuis, P. G., Prediction of solvent dependent beta-roll formation of a self-assembling silk-like protein domain. *Soft Matter* 2009, *5*, 2658-2665.

17. Camesano, T. A.; Wilkinson, K. J., Single Molecule Study of Xanthan Conformation Using Atomic Force Microscopy. *Biomacromolecules* 2001, *2*, 1184-1191.

18. Manning, G. S., The Persistence Length of DNA Is Reached from the Persistence Length of Its Null Isomer through an Internal Electrostatic Stretching Force. *Biophys. J.* 2006, *91*, 3607-3616.

19. Kassapidou, K.; Jesse, W.; Kuil, M. E.; Lapp, A.; Egelhaaf, S.; van der Maarel, J. R. C., Structure and Charge Distribution in DNA and Poly(styrenesulfonate) Aqueous Solutions. *Macromolecules* 1997, *30*, 2671-2684.

20. Tinland, B.; Pluen, A.; Sturm, J.; Weill, G., Persistence Length of Single-Stranded DNA. *Macromolecules* 1997, *30*, 5763-5765.

8.6 Appendix

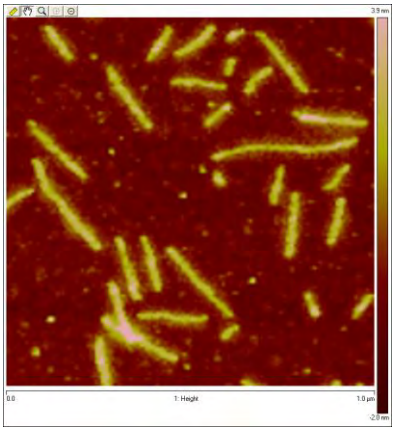


Figure 8.6. Complexation of polyacrylic acid (PAA) $8.0x10^5$ Da with cationic protein-based polymers C_4-S_{10}-B^{K12} at charge ratio +/-= 1 after 24 hrs of incubation.

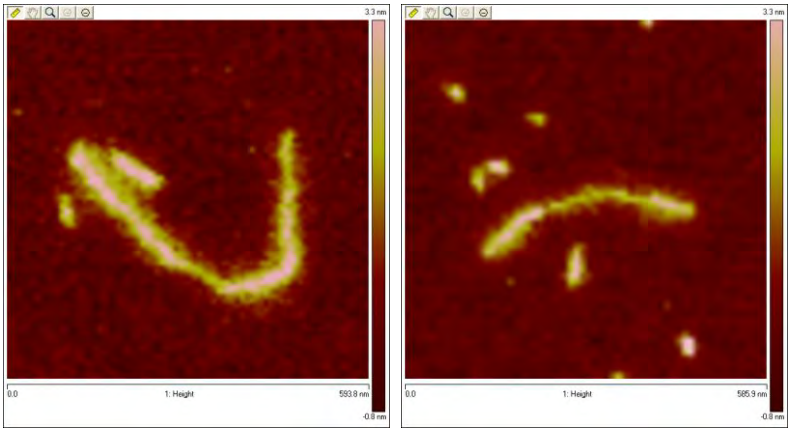


Figure 8.7. Complexation of PAA MW = $8.0x10^5$ Da with cationic protein-based polymers C_4-S_{10}-B^{K12} at charge ratio +/-= 7.5 after 24 hrs of incubation forming partially folded structures seen as wider segments.

Table 8.2. Height profiles (nm) of PAA complexed with protein polymers

	PAA size MW (x10^5 Da)			
	1.5	2.5	5	8
C_4-S_{10}-B^{K12}	2.7 ± 0.6	2.7 ± 0.7	2.7 ± 0.7	2.8 ± 0.7
C_4-S_{14}-B^{K12}	2.5 ± 0.8	3.0 ± 0.7	2.5 ± 0.6	2.6 ± 0.8

Chapter 9

Main concepts, summary and general discussion

9.1 Main concepts

In this thesis I demonstrate that it is possible to use **Protein-based Polymers** (**PbPs**) as synthetic binders of DNA (or other negatively charged polyelectrolytes). The PbPs co-assemble with their DNA templates to form highly organized virus-like particles and supramolecular structures. A range of PbPs have been developed over the last decades that can be used as precision functional polymers, and which integrate the unique properties of both proteins and polymers. Many PbPs are based on nature-inspired simple repetitive amino acid sequences. In this thesis, different kinds of such sequences have been combined into PbPs that mimic complex natural functionalities. Being intermediate between proteins and polymers, we have been able to mimic complex functionalities typically found for folded proteins, while retaining the tunability and ease of control that is more characteristic for (synthetic) polymers. Indeed, using clear design rules, biosynthetic PbPs sequences have been obtained and produced that co-assemble with nucleic acids to form true artificial viruses, which mimic their natural counterparts in many respects.

Our motivation for developing artificial viruses derives among others from the growing interest in exploiting natural self-assembled virus structures to develop nanostructured materials. In addition, natural viruses are being used as scaffolds for delivering DNA in the context of gene therapy, to serve as vaccines (by displaying antigens), to template diverse materials, to produce energy, to catch light, to catalyze reactions, to serve as contrasting agents, etc. Developing artificial viruses would serve not only to advance our capabilities to understand and control the co-assembly of nanostructures, but would also generate useful synthetic biomaterials that are even more suited than natural viruses to be used as building blocks for nanostructured materials. In short, the successful development of artificial viruses may be expected to give rise not only to new insights on templated self-assembly, but will also be very important for a range of applications. Key concepts in this thesis are:

Recombinant production of Protein-based Polymers. In chapter 2, 3 and 5 we show that the recombinant production of Protein-based Polymers that bind DNA is possible with relative high yields. We have shown this using both purely cationic blocks (Chapter 2 and 3) and using a naturally occurring folded (basic) DNA binding domain (Chapter 5). Yields ranged between 0.1-1 g per L of (cell-free) medium, when using 2.5 L fermentors. This demonstrates that DNA-binding PbPs that featuring both silk-like self-assembly domains and basic (cationic) DNA-binding domains are not too toxic to be efficiently produced by the recombinant yeast cells used for fermentation.

Coating/packing of single DNA molecules. This is demonstrated in chapters 2, 3 and 5. We also demonstrate coating of single *negatively charged supramolecular metal-coordinated polymers* in chapter 7, and coating of flexible *synthetic negatively charged polyelectrolytes* in chapter 8. Inserting of coated DNA into nanochannels was shown in chapter 4.

Templating. The coating of single chain molecules by our PbPs implies that they can turn a wide range of linear chain molecules into **templates for creating precisely defined nanostructures** (Chapter 2, 3, 5, 6, 7).

Bottle-brush Stiffening. Stiffening a stretching of single DNA molecules by coating them with polypeptide bottle brushes is illustrated and applied in chapters 2, 4 and 5.

Supramolecular virus-like nanostructures. Elongated supramolecular virus-like nanostructures were obtained for all the PbPs when complexed with DNA as illustrated in chapters 2, 3 and 5. These structures are highly reminiscent of the structure of cylindrical viruses such as TMV. In chapter 7 and 8 we show that virus-like nanostructures can also be obtained using other negatively polyelectrolytes and supramolecular polymers as a template.

Non viral gene delivery. The developed artificial viruses that we have developed are able to do gene delivery with efficiencies superior to polymeric standards and similar to that of lipid standards (chapter 3 and 5).

9.2 Summary

The main part of the thesis is divided into three parts. In part I, "Complexation of DNA into virus-like particles", we describe details of the molecular biomimetic strategy to design and produce PbPs with functionalities that mimic those of natural viruses. Part II, "Applications of protein-DNA complexes", deals with the development of diblock PbP that coat DNA, and with their applications in gene delivery and optical mapping of long DNA. Finally, in part III: "Supramolecular nanostructures beyond DNA" we consider the co-assembly of our PbPs with templates other than DNA, and also consider their self-assembly in the absence of DNA.

Part I "Complexation of DNA into virus-like particles". In **chapter 2** we describe the recombinant production of a diblock PbP with a large hydrophilic sequence for colloidal stability and a short cationic sequence for electrostatic complexation. Due to the extreme asymmetry in the lengths of the two blocks, complexes of these PbPs with (monodisperse) DNA are monodisperse elongated rod-like structures that reassemble rod-like virus structures. When coated with the diblock PbP, the DNA acquires a "bottle-brush" corona that provides colloidal stability against aggregation, for at least several days. The length of the co-assembled structures is dictated by the contour length of the DNA template, such that the coated DNA becomes very suitable for use as a template for creating further nanomaterials. The diblock PbPs can also help stretching DNA molecules, by making them stiffer. Finally, the diblock PbPs are a scaffold structure that can eventually be developed into artificial viruses.

In **chapter 3** we elaborate on the diblock design of chapter 2 and recombinantly produce four new triblock PbPs by also including a central silk-like self-assembly

domain (each PbP having a different length of the self-assembly domain). The self-assembly domains establish lateral protein-protein interactions with a strength that depends on their length. We find that triblock PbPs with a length of the self-assembly domain beyond some critical value behave as minimal artificial viral coat proteins. These triblock PbPs bind to- and pack single DNA molecules into rod-shaped virus like particles (VLPs) in a highly cooperative way. Inside the VLPs, DNA is protected from enzymatic digestion, and the VLPs efficiently deliver DNA to human cancer cells for expression. The kinetics of encapsulation was found to be highly similar to that of TMV, and can be described by the same kinetic model. Thermodynamic parameters for the kinetics of VLP formation were found to be very similar to those for the *in vitro* formation of TMV. In short, the VLPs behave in many ways as artificial viruses. We conclude that simple rules to translate physical functionalities of natural proteins into minimalistic polypeptide domains allows for the design multiblock PbPs that can faithfully mimic the assembly and some of the functional properties of viruses. Such a biomimetic supramolecular approach holds a great promising for mimicking other natural functional supramolecular structures.

Part II "Applications of protein-DNA complexes". In **Chapter 4** we demonstrated that by coating single DNA molecules and placing them inside nanochannels is possible to resolve fluorescent markers along the DNA chain. This is made possible by the stiffening effect provided by the diblock PbP (chapter 2) that forms a bottle-brush around the DNA. When the diameter of the nanochannels is below the effective persistence length of the coated DNA, it can no longer bend inside the nanochannels and hence becomes stretched. The imaging of the single DNA molecules was achieved using a fluorescence microscope equipped with an ultra-sensitive camera. This chapter provides proof-of-concept of the use of diblock PbPs for enhancing DNA stretching in nanopores, for optical mapping of long DNAs.

In **Chapter 5** we design and produce recombinantly a diblock PbP that incorporates a natural folded DNA binding domain, called "Sso7d", made up of 63

amino acids. The DNA binding domain is coupled to a very long hydrophilic block (~800 amino acids) that provides shielding and colloidal stability. As compared to the initial DNA-binding diblock PbPs of chapter 2, this eliminates the unspecificity of the cationic domains and gives specificity for DNA. Full coating of different types of single DNA molecules was achieved: double stranded linear DNA, double stranded circular DNA and single stranded DNA. We find a number of remarkable differences as compared to the unspecific DNA-binding diblocks: the DNA twist induced by binding of the B^{Sso7d} domain leads to the unwinding of plectonemically supercoiled DNA such that complexes appear as circles. This is in contrast to the case of the unspecific DNA-binding diblocks, that form complexes that appear as thick rods, with tightened plectonemic supercoiling. When binding to single-stranded DNA, the DNA unspecific diblock PbPs cannot completely prevent intramolecular base pairing such that the resulting complexes appear as compact and highly branched in AFM. In contrast, the DNA specific diblock PbPs are able to nearly completely prevent base pairing, such that complexes with circular single stranded DNA indeed appear as circles in AFM. Since the coated single stranded DNA is quite stiff, this opens up the possibility to stretch single long single stranded DNA in nanochannels, thus allowing optical mapping of single stranded DNA. Finally, the diblock PbP with the B^{Sso7d} DNA binding domain is shown to be an effective nanocarrier of DNA into cells, and to be non-toxic and non-hemolytic. Our results indicate that fusions of natural folded DNA-binding domains with PbPs are promising functional biomaterials for DNA-related applications. More generally, the results indicate the many new application possibilities that may arise when combining the unique properties of folded natural proteins with those of PbPs.

Part III "Supramolecular nanostructures beyond DNA". In Chapter 6, we perform a detailed characterization of the self-assembled nanostructures formed by the triblock PbPs of Chapter 3 in the absence of a template. We characterize the dependence of the self-assembly on key variables such as self-assembly time, and protein concentration. Whereas at low concentrations, the triblock PbPs appear to form irregular micelles, we find that beyond a well defined critical concentration,

the triblock PbPs start self-assembling into well defined nanorods, with lengths that depend on the protein concentration and protein design, in particular the length of the silk-like self-assembly block. Especially the latter correlation is highly relevant for improving design rules for self-assembling PbPs. Finally we show that the self-assembled nanorods are highly stable against dilution. The self-assembled protein nanorods are interesting scaffolds for nanomedicines, e.g. for delivery or multivalent display applications.

Chapter 7 describes the hierarchical co-assembly of the cationic triblock PbPs of Chapter 3 with negatively charged supramolecular polymers that are coordinated by metal ions. This results in highly regular, protein-metal nanorods. These novel self-assembled nanostructures can incorporate a large variety of metal ions, and are highly stable even at high salt concentrations, such that they could have a variety of applications, e.g. in nanomedicine or molecular electronics.

Chapter 8 describes the characterization of the electrostatic co-assembly of the triblock PbPs of Chapter 3 with several negatively charged polyelectrolytes of different molecular weight, and different flexibilities. As it turns out, all of them can function as template for the formation of rod-shaped virus-like particles, as was found earlier for DNA, in chapter 3. The polyelectrolytes that were tested are polyacrylic acid and polysulfonate, and xanthan. From extensive AFM imaging experiments, we conclude that the packing factor (contour length of rod-shaped virus like particles divided by the contour length of the template) depends on the intrinsic properties of the polyelectrolyte, in particular the charge density and stiffness. We conclude that the co-assembly of polyelectrolytes with the triblock PbPs is a very straightforward way to produce a variety of templated linear nanostructures, with many potential applications in the synthesis of nanostructured materials.

9.3 General Discussion: Artificial Viruses

9.3.1. Design of Artificial Viruses. The idea to develop an "artificial virus" or "synthetic virus" was first proposed by a number of groups interested in achieving high efficiency, *non-viral* gene therapy. Other people have also proposed artificial rod-like viruses with high aspect ratio can be interesting model systems for liquid crystals.[1, 2] The first molecules designed for gene therapy were based on cationic lipids or polymers. The compounds condense DNA into relatively large particles of limited stability.[3] As groups started finding more and more limitations of such compounds, they started exploring a range of strategies to improve their non-viral vectors.[4] A recurring theme is the literature on non-viral gene therapy has been the desire to imitate natural viruses that deliver DNA with high efficiency into the cell nucleus.[2]

The term "artificial virus" or "synthetic virus" can be found in the scientific literature starting from the middle of the 1990's[1, 2, 4-8] and the use of this term has slowly increased up to the present day.[9-30] It should not be confused with the de novo synthesis of a viral DNA or an RNA genome, as was reported in 2002.[31] The first mention to a "synthetic virus-like gene-transfer system" was in 1994 by the group of Ernst Wagner. They incorporated a short peptide sequence derived from a virus coat protein into positively charged polylysine-DNA complexes. Once inside the cell, these peptides disrupted the endosome and promote the gene transfer.[2] This original idea has motivated many other groups to also start to incorporate into extra functional sequences into non-viral gene transfer agents in order to mimic functionalities of real viruses. Another common element has become the incorporation of sufficient hydrophilicity, for example through addition of polyethylene glycol (PEG) blocks, in order to obtain colloidal stability,[32] which also reduces toxicity of cationic sequences. Furthermore, specific ligands have been added for cell targeting.[5, 7] Many peptide sequences, especially from viruses, with a so-called "proton-sponge effect" (acidification) were found, and utilized for promoting endosome escape. Also, nuclear pore targeting sequences were

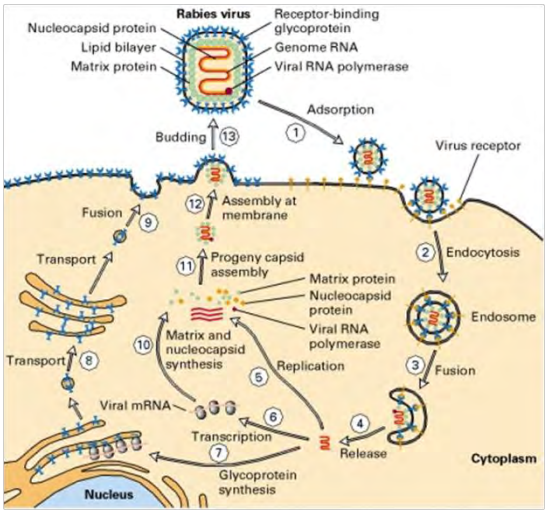


Figure 9.1. Some steps in the lytic replication cycle of an enveloped virus are illustrated for rabies virus. After a virion adsorbs to a specific host membrane protein (step 1), the cell engulfs it in an endosome (step 2). A protein in the endosome membrane pumps protons from the cytosol into the endosome interior. The resulting decrease in endosomal pH induces a conformational change in the viral glycoprotein, leading to fusion of the viral envelope with the endosomal lipid bilayer membrane and release of the nucleocapsid into the cytosol (steps 3 and 4). Replication of the viral RNA genome (step 5) and (step 6). Steps 7 to 13 were omitted. Image taken from: Section 6.3, Viruses: Structure, Function, and Uses Molecular Cell Biology. 4th edition. Lodish H, Berk A, Zipursky SL, et al. New York: W. H. Freeman; 2000. Copyright © 2000, W. H. Freeman and Company.

incorporated in the vectors in order to reach the nucleus.[2, 6] All these improvements, which were obtained through innovative new chemistries, have much improved the initial results obtained with the simpler polymer-based vectors.

A first generation of "artificial viruses" mainly aimed to imitate in non-viral vector designs (polymer or lipid- based) the first steps of the lytic cycle of viral replication: cell targeting, uptake, endosome-disruption and nucleus targeting (Figure 9.1). Those steps indeed determine the high delivery efficiency that characterizes natural viruses.[4, 8-10] However, in this work, no attention whatsoever was given to the structural features of the delivery vehicles at the nanoscale, and properties that arise from their structural features that also determine at least part of the delivery part of the lytic cycle. Such features could include, for example, a highly ordered nanostructure, specific molecular arrangements at the nanoscale,

etc. At least in part due to the limited control over the structure of the non-viral gene delivery vehicles, none of the polymer-based systems have actually reached clinical trials (most did not go beyond animal studies), only a few lipid-based ones have.[33] By looking at the large number of clinical trials that use viral vectors or even naked plasmid DNA instead of non-viral vectors, we can conclude that non-viral vectors have been very unsuccessful so far.

At the same time, our understanding of natural viruses is increasing rapidly. In the last decades, in particular our understanding of the structure of natural viruses has reached new heights. For several viruses, detailed understanding is now available on their mechanism of infection and replication. Following the paradigm "structure determines function", many scientists have advocated that it is necessary to unravel the structural details of viral coat proteins and viruses. Indeed, nowadays sub-nanometric details of many virus structures are well known. After a century of intense research on virus structures, in 1982 the Chemistry Nobel prize was awarded to Aaron Klug for the "structural elucidation of biologically important nucleic acid-protein complexes".[34] This includes the structure and self-assembly mechanism of several viruses, the best known one being the tobacco mosaic virus (TMV)[35-38] (Figure 9.2). His studies "illuminated" important principles about the spontaneous formation of viral particles from their components. Unraveling structural features of simple viruses such as TMV, and mechanistic insights into their formation, have contributed significantly to our ability to understand viruses, which is a prerequisite for eventually designing an artificial, man-made version of a virus. Indeed, designing a delivery system that precisely imitates fundamental structural features of viruses could lead non-viral delivery systems with efficiency close to or even superior to that of natural viruses. Such artificial functional systems may be expected to be (or can be designed to be) much more easy to tune and control than the natural systems, which also provides additional options to remedy problems such as safety issues that arise when natural viruses are used for gene transfer.

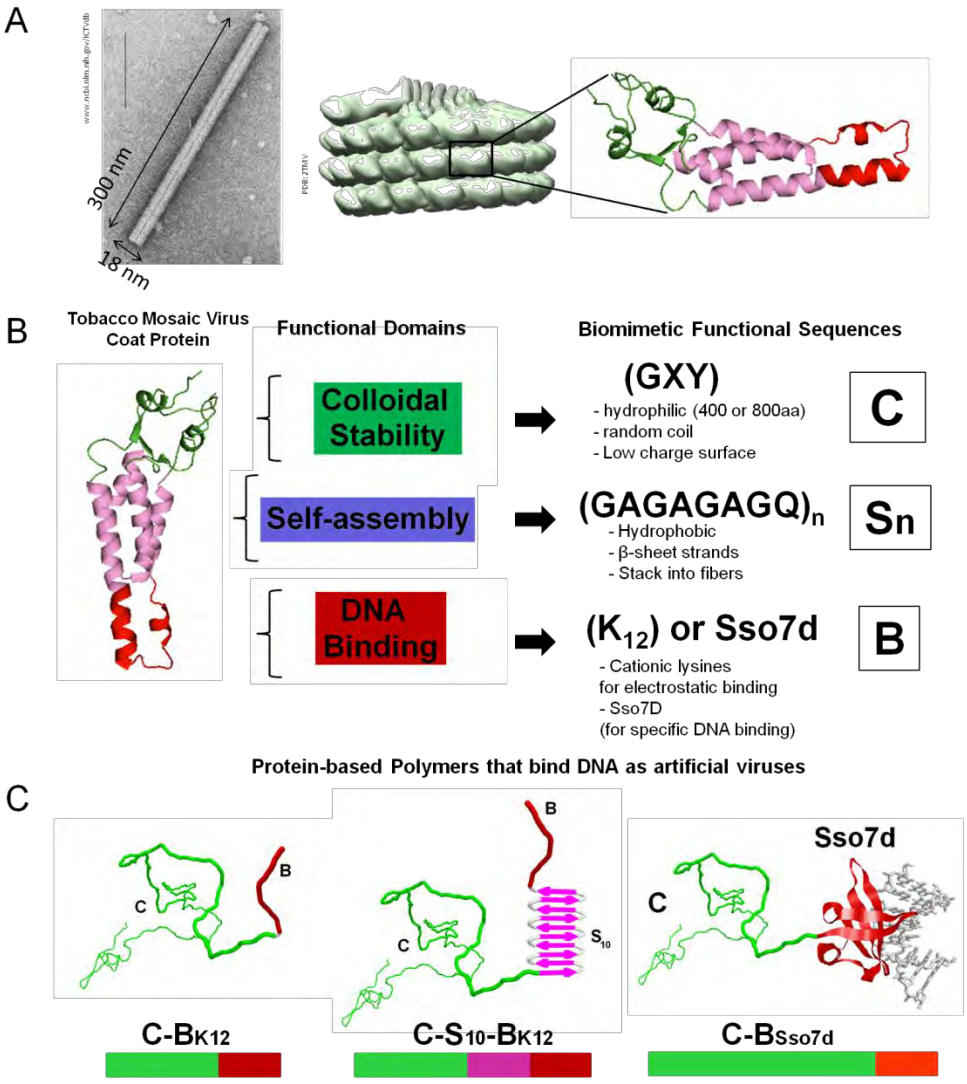


Figure 9.2. A, Self-assembled structure of tobacco mosaic virus capsid protein (TMV). B, Biomimetic approach to build an artificial virus with controlled nanostructure and nanoarchitecture. C, examples of biosynthesized protein-based polymers for binding DNA in this thesis that form nanostructures with DNA that reassemble artificial viruses in many respects.

Some of the key features found in natural viruses that are still missing in current non-viral gene delivery systems are: cooperativity, precise control of nanoarchitecture and the self-assembly of the nanostructure, allosteric regulation, molecular recognition of the cargo, responsiveness and control of virus size dictated by cargo size (templating). It may seem to be a daunting task to design building blocks for mimicking natural viruses, given the fact that nature has had millions of years to evolve current virus structures. An important question that can serve as a start, however, is: how to translate specific viral features and viral functionalities into separate building blocks that we can design independently. Could this be achieved in a "simple" and attainable way?

From the previous discussion it is clear that there are two complementary approaches that both aim at developing an artificial virus: one that focuses on mimicking the "functional" properties of a virus (cell targeting, endosome escape, etc.) and another one that instead aims at mimicking the structural- and physical features of viruses at the nanoscale (cooperativity, nanostructure, molecular recognition, allosteric responsivity but also size, shape and nanoarchitecture). I strongly believe that only a convergence between these approaches can lead to the development of a real nanosystem that can be considering an artificial virus and therefore, can be expected to push forward the development of efficient non-viral delivery systems.

Besides of the interest in gene therapy, several other fields, especially in material sciences, have been inspired by viruses. In materials science, viruses have been used as robust and well-defined structures at the nanoscale, that can be building blocks for further processing.[39] Indeed, natural viruses are now being used as scaffolds and building blocks in material sciences, and are expected to revolutionize methods of synthesizing and manufacturing advanced nanostructured materials.[40, 41] Virus-inspired self-assembly is also important in materials science, for example, peptides have been designed that self-assemble around carbon nanotubes, via molecular recognition.[42] Several other examples exist of designed molecules (mainly protein- or peptide based) that form virus-like particles,[21, 25, 43-46] for which the authors do

not even give a particular application, but instead simply assume that the ability to self-assemble structures with sizes and shapes of viruses can be important in many applications.

The aim of this thesis is to develop a strategy that in principle allows for the convergence of the two approaches mentioned above, even though (due to time restrictions) most of my efforts have been focused on the second approach. The approach involves designing proper polypeptide building blocks, that (when combined into multiblock polypeptides) can co-assemble with DNA into virus-like particles. The particles that we aimed to develop should be highly ordered, and feature precise control of their nanostructure and nanoarchitecture, as well as displaying as much functional virus properties as possible.

Our building blocks are based on simple polypeptides or protein-based polymers (PbPs). PbPs were chosen because we can achieve high control in the interactions that they establish with each other, and with other molecules, in this case DNA. (see Introduction section 1.1.2). PbPs can include blocks with well defined secondary structure, and at the same time behave to some extent as polymers, making them more tunable and designable. The ability to combine multiple functional blocks to create multifunctional PbPs has made them a very important platform to create complex, functional nanostructures. In this thesis we demonstrate that PbPs are highly promising for use as materials to make artificial virus particles. PbPs can encode properties such as:

Control of cargo size for delivery vehicles, by packing a fixed number of DNA molecules (e.g. a single molecule) into a protein nanostructure –chapter 2

Cooperativity and precise control of the nanostructure and nanoarchitecture – chapter 2 and 3

Precise molecular recognition chapter 5 (Figure 9.2c).

Other important features of viral self-assembly, such as allosteric regulation are not addressed in this thesis, but could be added in further work.

It is very important to mention that due to the recombinant origin of the PbPs we have full control of their amino acid sequences. This is important, considering that in some cases, for natural folded proteins, single amino acid variations in the sequence can lead to disastrous complete loss of functionality.

Control of the cargo size for artificial viruses (e.g. make an artificial virus that encapsulates 1 DNA molecule) has been one of the first aims in developing an artificial viruses. Most polymer vectors form large particles carrying multiple DNA molecules, with sizes that are poorly controlled. DeRouchey et. al. 2006 first reported the monomolecular coating of DNA.[47] However, they their polymeric vectors at very low concentrations of DNA, in order to avoid inter-molecular bridging of the DNA. In chapter 2, we demonstrate that the length of the binding block matters and that it should be reduced to very short lengths, in order to avoid any kind of bridging while still allowing for the coating of single DNA molecules. This prevents bridging and hence aggregation at high concentration regimes, such as is observed when using longer binding blocks.[48] Our control of the binding of PbPs to one single DNA molecule leads gives us control of the cargo size, a property exhibited by most natural viruses, and a prerequisite in order to obtain a true artificial virus (Chapter 3). This property of our PbPs, in principle also allows for the use of DNA as a template for developing other well defined linear nanostructures. The capacity of of our PbPs to bind to and encapsulate a single DNA molecule is illustrated in Figure 9.3 (from chapter 3).

After achieving precise control in the cargo amount (one single DNA molecule) a next feature that we wanted to incorporate is more control over the nanostructure and nanoarchitecture of the complexes. This requires control over polymer interactions that is hard to achieve using “simple” polymeric systems such as those

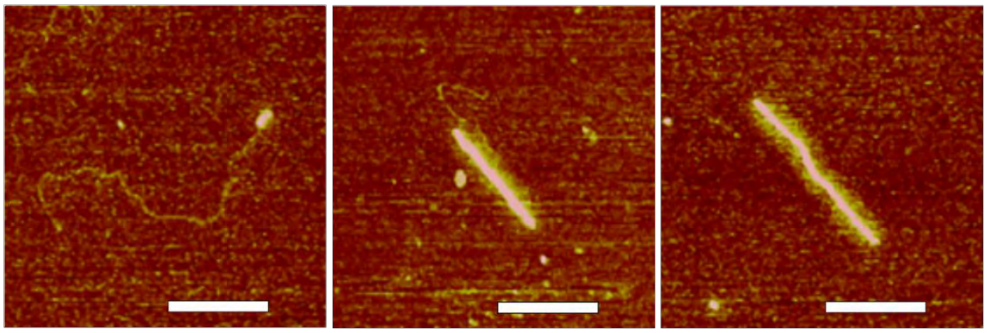


Figure 9.3. Binding and packing of a single DNA molecule by protein-based polymer C_4-S_{10}-B^{K12}. AFM images from ch 3, scale bar is 200nm. (Chapter 3).

based on polyethyleneimine, poly(2-dimethylaminoethyl methacrylate) (PDMAEMA). It is also very difficult to obtain highly regular structures with controlled size when using lipids. Designing from scratch a folded globular protein that achieves similar functionalities as those of natural viral coat proteins is still impossible at the persent state of art of the de-novo protein design.[42] For example, designing protein nanocapsules that self-assemble upon binding to a specific cargo is a goal that has not yet been achieved. This would be an important next step for the future and initial steps in this direction have indeed been recently made (the design of empty nanocapsules.[49] With respect to design of complex multifunctional self-assembling macromolecules, the PbP-approach is uniquely versatile and reliable. Indeed, in chapter 3 we demonstrate that we can combine a silk-like self-assembly domain that folds into precise secondary nanostructures called beta-rolls, with a DNA binding moiety to obtain a molecule that stacks into fiber-like nanostructures when binding to DNA. Several other rather simple self-assembling peptide sequences have been reported in literature, which could also have been used for our purpose. The advantage of using multiple repeats of such simple (short) sequences is that we can control the strength of the interactions easily by tuning the length of the self-assembly block (Chapter 3 and 6, Figure 9.4). As the PbPs already establish lateral interactions, it is sufficient to use weak cationic binding blocks (e.g. short stretches of basic amino acids) to bind to the DNA. This

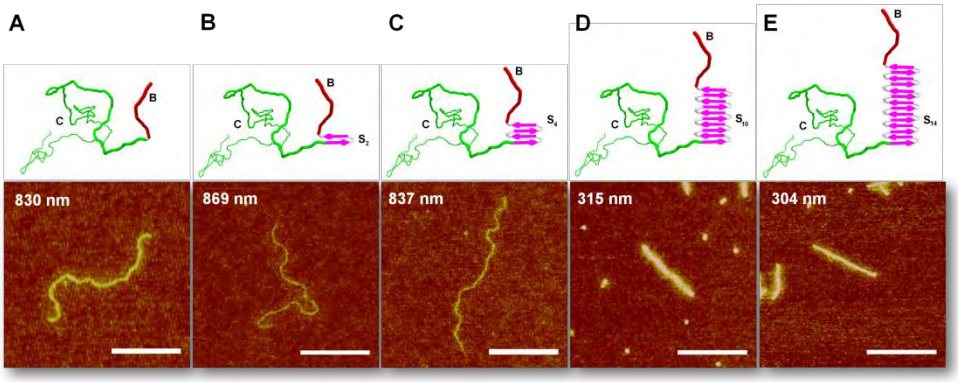


Figure 9.4. Modulation of the cooperativity strength by controlling the length of the self-assembly domain "S_n". (Chapter 3).

approach allows for control of the nanoarchitecture, without using complex and elaborated protein folds. On the other hand, a polypeptide is required to attain the regular nanostructures we find that are ultimately based on secondary structure elements, whereas with traditional polymers this would be, since these typically lack the fine control over secondary structures that is available for polypeptides. Protein science has evolved to the point where there are clearly established consensus sequences that fold into regular secondary structures, which implies that there is a wealth of structural elements that can be directly incorporated into PbPs. Indeed, in chapter 3 we show that by precisely tuning the strength of the silk-like self-assembly block, we can find an optimum for which fibrilization only occurs on the DNA (see Figure 9.4) and not in the bulk.

The incorporation of a self-assembly domain brings about a cooperative self-assembly mechanism, a feature that so far has not been incorporated explicitly in current designs of non-viral vectors. Cooperativity is an essential feature found in all living systems and viruses.[50] It was called "the second secret of life " by the famous scientist Jaques Monod, second only to the structure of DNA.[51] Indeed, cooperativity is a key feature that allows for the reliable organization of small building blocks into highly organized, large structures. Hierarchical complexity mediated by cooperativity has been argued to be responsible for emergent

properties in living systems. In viruses, cooperativity also plays a key role in the co-assembly of viral particles around the target polynucleotide (genome of the virus), and can prevent viral particle formation the absence of the proper template. It assures full protection due to the all-or-nothing encapsulation of the genome. In viruses, cooperativity is often mediated by protein-protein interactions that are allosterically regulated. The allosteric or conformational control has been pinpointed as the heart of the cooperativity. In chapter three we describe similar features for our artificial virus particles. The self-assembly of the silk-like domain is triggered with a nucleation step, where several polypeptides find each other through hydrophobic interactions. Next, they fold in the proper secondary structure and stack into fibers. By binding the PbPs, the template dramatically increases the local concentration of PbPs, and hence promotes the formation of the fiber-like structure that packs the DNA. Although cooperativity typically hard to tune, we have shown how this can be achieved by varying the length of silk-like self-assembly domains, and how such control can be exploited in a synthetic virus. Explicit control of cooperativity is new for non-viral vectors. Lipid-based vectors can also be said to exhibit some kind of cooperativity, in view of the existence of critical concentrations such as the critical micelle - and critical aggregation concentration.

Viruses have evolved to recognize and specifically pack their own genomes. In many cases, they have an origin-of-assembly sequence where the first building blocks start to pack the on the nucleic acid. The ability to trigger self-assembly using its own genome is an advanced molecular recognition that is way beyond the capabilities of the simple cationic binding blocks we use here. As a first step toward the incorporation of more specific molecular recognition in PbPs, in Chapter 4, we joined together a domain that binds specifically to DNA (and not to any negative polyelectrolyte), to a protective PbP sequence. Hence, we had shown that it is also possible to engineer much more specific domains into artificial viral coat protein designs. If a domain that binds sequence specifically would be combined with a self-assembly and a shielding domain that provides colloidal stability, we would

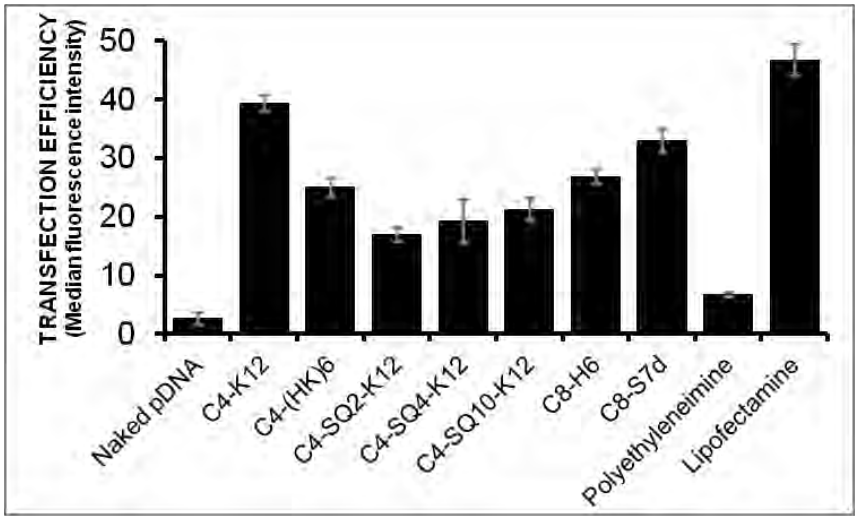


Figure 9.5. Gene delivery efficiency of some produced protein-based polymers in this thesis compared with commercial standards. (Chapter 3).

potentially have a PbP that could preferentially encapsulate specific polynucleotide sequences.

Our artificial virus designs, either with self-assembly domain or without, have been shown to effectively deliver DNA into cells with efficiencies better than a polymer standard and with similar efficiencies as a lipid standard (chapter 3 and 5). Here, not only the structural features are important, but many other parameters also come into play such as biocompatibility and toxicity. We found that our PbPs were highly biocompatible in the sense that they had very low toxicity and hemolytic activity (chapter 5). Their relatively high transfection efficiency combined with the good safety profile, make them promising scaffolds for future non-viral vectors (Figure 9.5), that could additionally include extra functional sequences such as endosome escape- or targeting sequences.

The main message of the above arguments is that the design of efficient non-viral vectors unavoidably has to go in to the direction of control of structural details at the nanoscale, in a similar ways as done by viruses. It is not enough to just

incorporate unstructured sequences for which some particular functionality has been reported when that sequence is in some particular context. A further organizational level is required that can only be achieved using more sophisticated molecular designs, such as protein-based polymers that (as we have shown) are quite amenable to engineering via tuning of relative blocks lengths, and in addition can be produced at a large scale. As long as we do not connect the desired biological functionalities with control over the self-assembly of precisely defined nano-architectures, we will not go beyond the delivery efficiencies of current non-viral gene delivery agents, that cannot really be called artificial viruses yet. Instead, we should take examples from the structures and mechanism found in simplest viruses in nature, in order to expand the boundaries of our molecular design capabilities. Finally, last but not least, learning how to build an artificial virus may also have implications for the study of the origin of life, since precise self-assembly of organic molecules into organized systems, is supposedly how life could originated.[52]

9.3.2. Other Applications. Besides the already mentioned application in gene delivery, our designed PbPs may have further applications in material sciences and in genomics. By coating monodisperse single DNA molecules with the polypeptides (chapter 2 and 4) is possible to deposit on them metal particles that can serve to manufacture metal nanowires.[53] Triblock PbPs have already a cysteine that is solvent exposed (Chapter 3). Through that cysteine is possible to deposit gold and other metals by simple chemical reducing methods.[54, 55] Another interesting application is stretching single DNA molecules inside nanochannels to resolve fluorescently labeled genetic markers or mutations (Figure 9.6). This approach exploits the formation of a bottle brush around the coated DNA to make is stiffer and reduce its bending conformations. Chapters 2, 4 and 5 demonstrate these principles. This idea can be extended to different types of single DNA molecules (single stranded, double stranded, linear, circular and supercoiled) (Figure 9.7). The flexible design of the PbPs, makes them broadly applicable. The enhanced stretching due to a bottle-brush coating is quite promising for

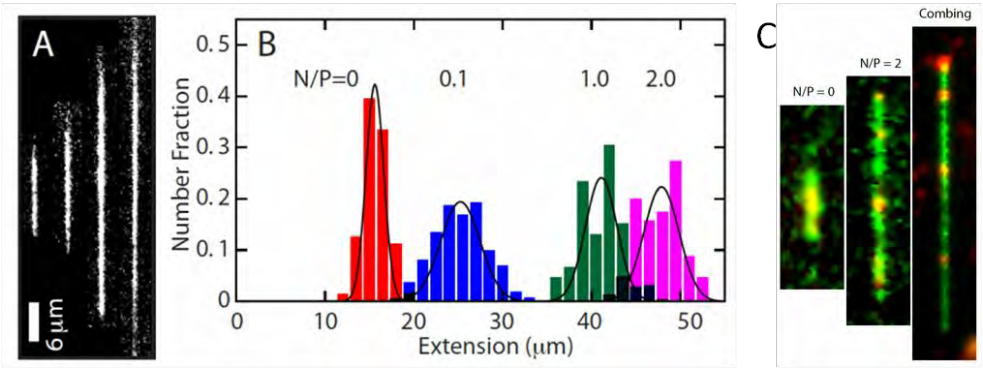


Figure 9.6. (A) Stretching a single DNA molecule inside a nanochannel by coating it with protein-based polymer C_4-B^{K12}. (B) Effect in the extension of the DNA molecule when coated. (C) Resolving fluorescently labeled genetic markers and compared with standard "DNA combing" method. (Chapter 4).

developing technologies for efficient optical mapping and sequencing of long DNA molecules. In chapter 7, we show that triblock PbPs can also form supramolecular rod-like nanostructures of different sizes by themselves, which could have a whole range of other applications.

Due to their high aspect ratio, artificial virus-like particles could be used as a route to produce biomolecular lyotropic liquid crystals with highly tunable properties.[1] In chapter 8, we show the production of hybrid metal ion-proteins nanostructures via co-assembly of PbPs with metal-coordinated supramolecular polymers. In this case some principles used for developing the artificial viruses were used, such as electrostatic co-assembly and cooperativity but we also exploit the capability of the metal ions to coordinate supramolecular polymers. In this way triblock PbPs can be used to develop functional materials based on metals with applicability in material science and nanomedicine (Figure 9.8).

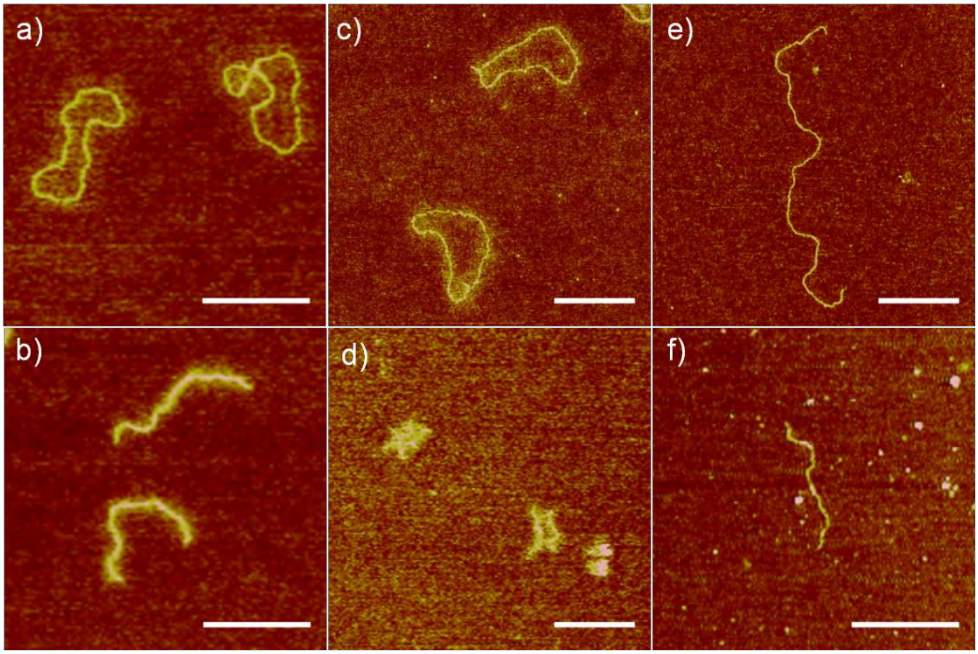


Figure 9.7. Stretching single DNA molecules of different conformations with protein-based polymers. Supercoiled pDNA with (a) C_8-B^{Sso7d} (b) C_4-B^{K12}. Single stranded circular DNA from M13 virus with (c) C_8-B^{Sso7d} (d) C_4-B^{K12}. C_4-$B^{K12}$with double stranded linear DNA (e) 8.0 kbp and (f) 2.5 kbp. (Chapter 5).

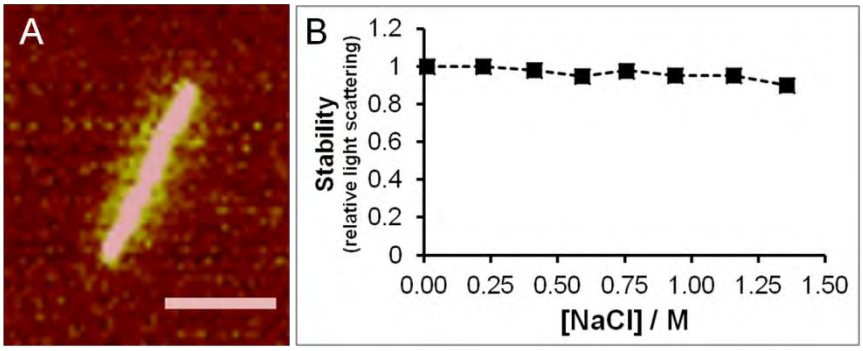


Figure 9.8. (a) Virus-like nanostructures between polymer-based polymer C_4-S_{10}-B^{K12} and a supramolecular metal (Zn^{2+}) polymer. (b) High stability hybrid protein-metal functional materials can be obtained for applications in material science and nanomedicine. (Chapter 7).

9.4 Perspectives

9.4.1. Protein-Based Polymers that Bind DNA. In order to further develop PbPs that bind to DNA we first need to understand some detailed issues related to their co-assembly that are somewhat unclear. This includes, in particular, an understanding of the conformational changes that the self-assembly domain and DNA undergo upon PbP binding, stacking and during elongation of virus-like particles. It is not clear if the DNA conformation inside the complexes has some regular pattern or not. Further questions are: does the self-assembly block also interacts with the hydrophobic core of the DNA? There seem to be some indications for this from fluorescence data (not reported in this thesis). Other questions: do the cationic blocks bind to the minor or major groove or both? Is the hydrophobic self-assembly domain forming beta-rolls or beta-sheets? Are the self-assembly blocks folded when free in solution? Finally, it would also be good to have high quality binding isotherms for PbPs with DNA, which are currently, are not yet available.

Beyond these detailed questions, there is also still much room for improved understanding and control of cooperativity. The latter is extremely important since it can be used to impart a robust responsivity to self-assembling systems. In this thesis we have focused on linear self-assembled structures, and it would obviously also be of great interest to try to design other types of nanostructures using our PbP approach, such as spheres, tubes, triangles etc.

A question of biological relevance is how the complexes between the PbPs and the DNA lead to transfection. How can they escape into cytoplasm and reach the nucleus? Preliminary results indicate that the complexes enter cells through endosomes and they are quickly releases into the cytoplasm (Figure 9.9). Regarding the delivery, multicomplexation of several therapeutic cargos each with a different

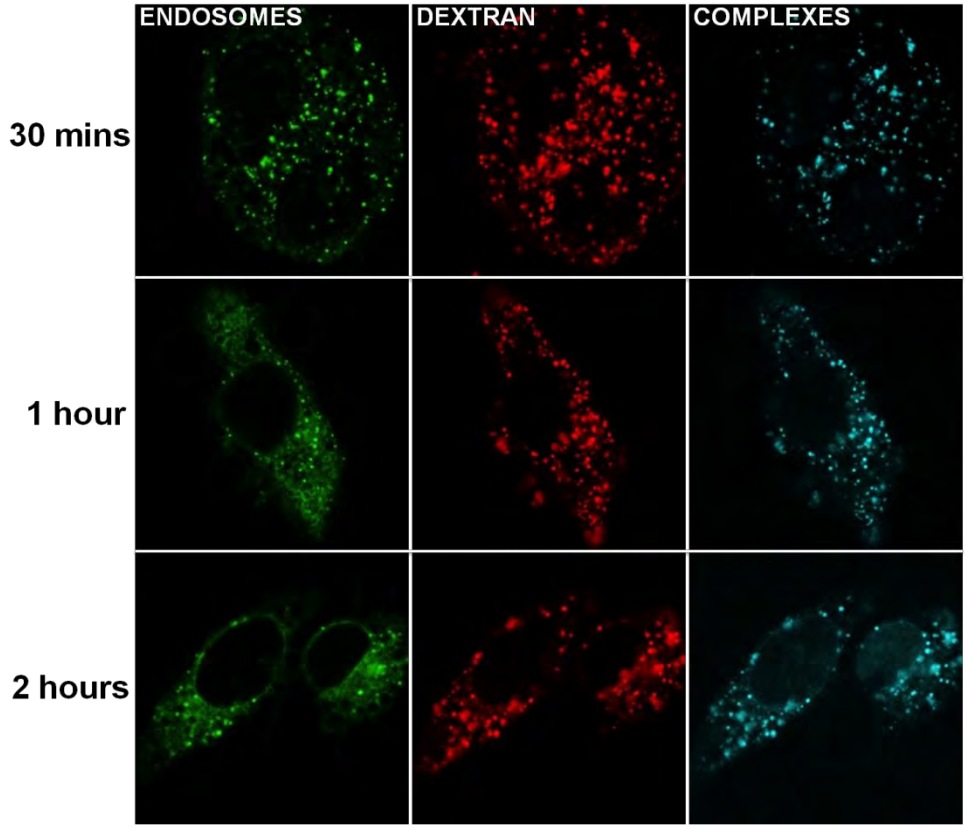


Figure 9.9. Preliminary results showing the uptake by HeLa cells of complexes between pDNA and protein C_4-B^{K12}. In the first column early endosomes are shown through fluorescent imaging of Rab5, in the second column is imaged Dextran which is an indicator of macropinocytosis and in the third column is imaged DNA intercalated with YOYO-1 complexed with the protein. There is larger correlation of the DNA image with the dextran than with Rab5, which suggest that complexes may be entering the cells by macropinocytosis. Also it can be appreciated that the complexes already entered cells after 30 minutes. After 1h pDNA has been released into the cytoplasm and after 2h it is found inside the nucleus.

chemical nature could be explored. We can also add other amino acid sequences to our PbP designs to improve their performance, either for targeting or for reaching the nucleus. In order to work on such improvements however, we first need to know the limiting steps for the current PbPs.

Interesting and fruitful work can also be done together with computational scientists. They can provide basic information on building blocks such as thermodynamic parameters and folding pathways that can be used for screening sequences before their actual use in PbPs. Unless a fully integrated approach is taken however, involving closely collaborating multidisciplinary teams, this approach may not work so well, for example due to divergent scientific interests in different disciplines, phasing problems in joint projects, and a lack of mutual understanding due to different conceptual frameworks.

9.4.2. Biomimetic Approach Beyond Artificial Viruses. A fascinating possible application of our PbPs that I would like to mention last is the possibility to carry out biomolecular design of other synthetic biosystems. For example, the generation of clean energy in a sustainable fashion is of great importance in our times. It could ultimately be possible to mimic complex nanosystems such as the ones involved in photosynthesis to produce energy using design approaches such as our PbP approach. So far, only artificial photosynthetic systems based on polymers have been developed.[56-58] Using a biomimetic molecular design based on PbPs this might be feasible, because of the many advantages of PbPs over other approaches with respect to the design of self-assembling nanostructures that we have already advanced. There is even a connection to (artificial) viruses, since some labs have already demonstrated the use of TMV virus to harvest light.[59]

9.5 References

1. Percec, V.; Heck, J.; Johansson, G.; Tomazos, D.; Ungar, G., Towards tobacco mosaic virus-like self-assembled supramolecular architectures. *Macromol. Symp.* 1994, *77*, 237-265.

2. Plank, C.; Oberhauser, B.; Mechtler, K.; Koch, C.; Wagner, E., The influence of endosome-disruptive peptides on gene-transfer using synthetic virus-like gene-transfer systems. *J. Biol. Chem.* 1994, *269*, 12918-12924.

3. De Smedt, S. C.; Demeester, J.; Hennink, W. E., Cationic polymer based gene delivery systems. *Pharm. Res.* 2000, *17*, 113-126.

4. Lehn, P.; Fabrega, S.; Oudrhiri, N.; Navarro, J., Gene delivery systems: Bridging the gap between recombinant viruses and artificial vectors. *Adv. Drug Deliv. Rev.* 1998, *30*, 5-11.

5. Remy, J. S.; Kichler, A.; Mordvinov, V.; Schuber, F.; Behr, J. P., Targeted gene-transfer into hepatoma-cells with lipopolyamine-condensed dna particles presenting galactose ligands - a stage toward artificial viruses. *Proceedings of the National Academy of Sciences of the United States of America* 1995, *92*, 1744-1748.

6. Budker, V.; Gurevich, V.; Hagstrom, J. E.; Bortzov, F.; Wolff, J. A., pH-sensitive, cationic liposomes: A new synthetic virus-like vector. *Nat. Biotechnol.* 1996, *14*, 760-764.

7. Erbacher, P.; Remy, J. S.; Behr, J. P., Gene transfer with synthetic virus-like particles via the integrin-mediated endocytosis pathway. *Gene Ther.* 1999, *6*, 138-145.

8. MacLachlan, I.; Cullis, P.; Graham, R. W., Progress towards a synthetic virus for systemic gene therapy. *Curr. Opin. Mol. Ther.* 1999, *1*, 252-259.

9. Behr, J. P., Towards artificial viruses. *Journal of Controlled Release* 2001, *72*, 225-227.

10. Belguise-Valladier, P.; Behr, J. P., Nonviral gene delivery: Towards artificial viruses. *Cytotechnology* 2001, *35*, 197-201.

11. Nahde, T.; Muller, K.; Fahr, A.; Muller, R.; Brusselbach, S., Combined transductional and transcriptional targeting of melanoma cells by artificial virus-like particles. *J. Gene. Med.* 2001, *3*, 353-361.

12. Zuber, G.; Dauty, E.; Nothisen, M.; Belguise, P.; Behr, J. P., Towards synthetic viruses. *Adv. Drug Deliv. Rev.* 2001, *52*, 245-253.

13. Zhang, Y.; Zhu, C. N.; Pardridge, W. M., Antisense gene therapy of brain cancer with an artificial virus gene delivery system. *Molecular Therapy* 2002, *6*, 67-72.

14. Aoyama, Y.; Kanamori, T.; Nakai, T.; Sasaki, T.; Horiuchi, S.; Sando, S.; Niidome, T., Artificial viruses and their application to gene delivery. Size-controlled gene coating with glycocluster nanoparticles. *Journal of the American Chemical Society* 2003, *125*, 3455-3457.

15. Kakudo, T.; Chaki, S.; Futaki, S.; Nakase, I.; Akaji, K.; Kawakami, T.; Maruyama, K.; Kamiya, H.; Harashima, H., Transferrin-modified liposomes equipped with a pH-sensitive fusogenic peptide: An artificial viral-like delivery system. *Biochemistry* 2004, *43*, 5618-5628.

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

17. Walker, G. F.; Fella, C.; Pelisek, J.; Fahrmeir, J.; Boeckle, S.; Ogris, M.; Wagner, E., Toward synthetic viruses: Endosomal pH-triggered deshielding of targeted polyplexes greatly enhances gene transfer in vitro and in vivo. *Molecular Therapy* 2005, *11*, 418-425.

18. Boeckle, S.; Wagner, E., Optimizing targeted gene delivery: Chemical modification of viral vectors and synthesis of artificial virus vector systems. *Aaps Journal* 2006, *8*, E731-E742.

19. Mastrobattista, E.; van der Aa, M.; Hennink, W. E.; Crommelin, D. J. A., Artificial viruses: a nanotechnological approach to gene delivery. *Nature Reviews Drug Discovery* 2006, *5*, 115-121.

20. Xu, P. S.; Li, S. Y.; Li, Q.; Zhan, Y. H.; Ren, J.; Radosz, M.; Shen, Y. Q., Polymeric artificial virus for gene delivery. *Abstr. Pap. Am. Chem. Soc.* 2006, *231*, 2.

21. Boato, F.; Thomas, R. M.; Ghasparian, A.; Freund-Renard, A.; Moehle, K.; Robinson, J. A., Synthetic virus-like particles from self-assembling coiled-coil lipopeptides and their use in antigen display to the immune system. *Angewandte Chemie-International Edition* 2007, *46*, 9015-9018.

22. Tresset, G.; Cheong, W. C. D.; Boulaire, J.; Tan, S. Y. L.; Lam, Y. M., Exploiting the like-charge attraction between phospholipids and DNA to assemble biologically relevant artificial viruses. *Biophys. J.* 2007, 237A-237A.

23. Tresset, G.; Cheong, W. C. D.; Tan, Y. L. S.; Boulaire, J.; Lam, Y. M., Phospholipid-based artificial viruses assembled by multivalent cations. *Biophys. J.* 2007, *93*, 637-644.

24. Douglas, K. L., Toward development of artificial viruses for gene therapy: A comparative evaluation of viral and non-viral transfection. *Biotechnol. Prog.* 2008, *24*, 871-883.

25. Lim, Y.-b.; Lee, E.; Yoon, Y.-R.; Lee, M. S.; Lee, M., Filamentous artificial virus from a self-assembled discrete nanoribbon. *Angewandte Chemie-International Edition* 2008, *47*, 4525-4528.

26. Sasaki, K.; Kogure, K.; Chaki, S.; Nakamura, Y.; Moriguchi, R.; Hamada, H.; Danev, R.; Nagayama, K.; Futaki, S.; Harashima, H., An artificial virus-like nano carrier system: enhanced endosomal escape of nanoparticles via synergistic action of pH-sensitive fusogenic peptide derivatives. *Anal. Bioanal. Chem.* 2008, *391*, 2717-2727.

27. McCarthy, H. O.; Wang, Y. H.; Mangipudi, S. S.; Hatefi, A., Advances with the use of bio-inspired vectors towards creation of artificial viruses. *Expert Opin. Drug Deliv.* 2010, *7*, 497-512.

28. Lamarre, B.; Ryadnov, M. G., Self-Assembling Viral Mimetics: One Long Journey with Short Steps. *Macromol. Biosci.* 2011, *11*, 503-513.

29. Rodik, R. V.; Klymchenko, A. S.; Jain, N.; Miroshnichenko, S. I.; Richert, L.; Kalchenko, V. I.; Mely, Y., Virus-Sized DNA Nanoparticles for Gene Delivery Based on Micelles of Cationic Calixarenes. *Chem.-Eur. J.* 2011, *17*, 5526-5538.

30. Miyata, K.; Nishiyama, N.; Kataoka, K., Rational design of smart supramolecular assemblies for gene delivery: chemical challenges in the creation of artificial viruses. *Chem. Soc. Rev.* 2012, *41*, 2562-2574.

31. Cello, J.; Paul, A. V.; Wimmer, E., Chemical Synthesis of Poliovirus cDNA: Generation of Infectious Virus in the Absence of Natural Template. *Science* 2002, *297*, 1016-1018.

32. Ogris, M.; Brunner, S.; Schuller, S.; Kircheis, R.; Wagner, E., PEGylated DNA/transferrin-PEI complexes: reduced interaction with blood components, extended circulation in blood and potential for systemic gene delivery. *Gene Ther.* 1999, *6*, 595-605.

33. Yanagie, H., Clinical trials using non-viral gene delivery systems. *Non-Viral Gene Therapy: Gene Design and Delivery* 2005, 261-290.

34. Nobel-Foundation, The Nobel Prize in Chemistry 1982-Press Release "Depicting the building blocks of life". 1982.

35. Butler, P. J. G.; Klug, A., Assembly of particle of tobacco mosaic virus from rna and disks of protein. *Nature-New Biology* 1971, *229*, 47-&.

36. Durham, A. C. H.; Finch, J. T.; Klug, A., States of aggregation of tobacco mosaic virus protein. *Nature-New Biology* 1971, *229*, 37-&.

37. Bloomer, A. C.; Champness, J. N.; Bricogne, G.; Staden, R.; Klug, A., Protein disk of tobacco mosaic-virus at 2.8-A resolution showing interactions within and between subunits. *Nature* 1978, *276*, 362-368.

38. Klug, A., The tobacco mosaic virus particle: structure and assembly. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 1999, *354*, 531-535.

39. Liu, Z.; Qiao, J.; Niu, Z.; Wang, Q., Natural supramolecular building blocks: from virus coat proteins to viral nanoparticles. *Chem. Soc. Rev.* 2012, *41*, 6178-6194.

40. Comellas-Aragones, M.; Engelkamp, H.; Claessen, V. I.; Sommerdijk, N.; Rowan, A. E.; Christianen, P. C. M.; Maan, J. C.; Verduin, B. J. M.; Cornelissen, J.; Nolte, R. J. M., A virus-based single-enzyme nanoreactor. *Nature Nanotechnology* 2007, *2*, 635-639.

41. Douglas, T.; Young, M., Host-guest encapsulation of materials by assembled virus protein cages. *Nature* 1998, *393*, 152-155.

42. Grigoryan, G.; Kim, Y. H.; Acharya, R.; Axelrod, K.; Jain, R. M.; Willis, L.; Drndic, M.; Kikkawa, J. M.; DeGrado, W. F., Computational Design of Virus-Like Protein Assemblies on Carbon Nanotube Surfaces. *Science* 2011, *332*, 1071-1076.

43. Domingo-Espin, J.; Vazquez, E.; Ganz, J.; Conchillo, O.; Garcia-Fruitos, E.; Cedano, J.; Unzueta, U.; Petegnief, V.; Gonzalez-Montalban, N.; Planas, A. M.; Daura, X.; Peluffo, H.; Ferrer-Miralles, N.; Villaverde, A., Nanoparticulate architecture of protein-based artificial viruses is supported by protein-DNA interactions. *Nanomedicine* 2011, *6*, 1047-1061.

44. Hartmann, L.; Haefele, S.; Peschka-Suess, R.; Antonietti, M.; Boerner, H. G., Tailor-made poly(amidoamine)s for controlled complexation and condensation of DNA. *Chem.-Eur. J.* 2008, *14*, 2025-2033.

45. Ruff, Y.; Moyer, T.; Newcomb, C. J.; Demeler, B.; Stupp, S. I., Precision Templating with DNA of a Virus-like Particle with Peptide Nanostructures. *Journal of the American Chemical Society* 2013, *135*, 8.

46. Valery, C.; Paternostre, M.; Robert, B.; Gulik-Krzywicki, T.; Narayanan, T.; Dedieu, J. C.; Keller, G.; Torres, M. L.; Cherif-Cheikh, R.; Calvo, P.; Artzner, F., Biomimetic organization: Octapeptide self-assembly into nanotubes of viral capsid-like dimension. *Proceedings of the National Academy of Sciences of the United States of America* 2003, *100*, 10258-10262.

47. DeRouchey, J.; Walker, G. F.; Wagner, E.; Radler, J. O., Decorated rods: A "bottom-up" self-assembly of monomolecular DNA complexes. *Journal of Physical Chemistry B* 2006, *110*, 4548-4554.

48. Osada, K.; Oshima, H.; Kobayashi, D.; Doi, M.; Enoki, M.; Yamasaki, Y.; Kataoka, K., Quantized Folding of Plasmid DNA Condensed with Block Catiomer into Characteristic Rod Structures Promoting Transgene Efficacy. *J. Am. Chem. Soc.* 2010, *132*, 12343-12348.

49. King, N. P.; Sheffler, W.; Sawaya, M. R.; Vollmar, B. S.; Sumida, J. P.; Andre, I.; Gonen, T.; Yeates, T. O.; Baker, D., Computational Design of Self-Assembling Protein Nanomaterials with Atomic Level Accuracy. *Science* 2012, *336*, 1171-1174.

50. Perutz, M. F., Mechanisms of cooperativity and allosteric regulation in proteins. *Q. Rev. Biophys.* 1989, *22*, 139-236.

51. Fenton, A. W., Allostery: an illustrated definition for the 'second secret of life'. *Trends Biochem.Sci.* 2008, *33*, 420-425.

52. Koonin, E.; Senkevich, T.; Dolja, V., The ancient Virus World and evolution of cells. *Biology Direct* 2006, *1*, 29.

53. Dujardin, E.; Peet, C.; Stubbs, G.; Culver, J. N.; Mann, S., Organization of metallic nanoparticles using tobacco mosaic virus templates. *Nano Letters* 2003, *3*, 413-417.

54. Scheibel, T.; Parthasarathy, R.; Sawicki, G.; Lin, X. M.; Jaeger, H.; Lindquist, S. L., Conducting nanowires built by controlled self-assembly of amyloid fibers and selective metal deposition. *Proceedings of the National Academy of Sciences of the United States of America* 2003, *100*, 4527-4532.

55. Xiong, S. L.; Xi, B. J.; Wang, C. M.; Zou, G. F.; Fei, L. F.; Wang, W. Z.; Qian, Y. T., Shape-controlled synthesis of 3D and 1D structures of CdS in a binary solution with L-cysteine's assistance. *Chem.-Eur. J.* 2007, *13*, 3076-3081.

56. van Hal, P. A.; Christiaans, M. P. T.; Wienk, M. M.; Kroon, J. M.; Janssen, R. A. J., Photoinduced electron transfer from conjugated polymers to TiO2. *Journal of Physical Chemistry B* 1999, *103*, 4352-4359.

57. Wang, X.; Maeda, K.; Chen, X.; Takanabe, K.; Domen, K.; Hou, Y.; Fu, X.; Antonietti, M., Polymer Semiconductors for Artificial Photosynthesis: Hydrogen Evolution by Mesoporous Graphitic Carbon Nitride with Visible Light. *Journal of the American Chemical Society* 2009, *131*, 1680-+.

58. Alstrum-Acevedo, J. H.; Brennaman, M. K.; Meyer, T. J., Chemical approaches to artificial photosynthesis. 2. *Inorganic Chemistry* 2005, *44*, 6802-6827.

59. Miller, R. A.; Presley, A. D.; Francis, M. B., Self-assembling light-harvesting systems from synthetically modified tobacco mosaic virus coat proteins. *Journal of the American Chemical Society* 2007, *129*, 3104-3109.

Samenvatting

Het grootste deel van dit proefschrift is verdeeld in drie delen. In deel I "Complexering van DNA in virusachtige deeltjes", beschrijven we de bijzonderheden van de moleculaire biomimetische strategie om PbPs (Protein Block coPolymer, oftewel eiwitblokpolymeer) te ontwerpen en produceren. Deze PbPs hebben functionaliteiten die de natuurlijke virussen nabootsen. Deel II," Toepassingen van eiwit-DNA complexen" gaat over de ontwikkeling van diblok PbPs die aan DNA binden, en hun toepassingen in gentherapie en optische analyse van lange DNA ketens. In deel III: "Supramoleculaire structuren anders dan DNA" bekijken we de co-assemblage van onze PbPs met andere componenten dan DNA, en ook hun zelfassemblage in de afwezigheid van DNA.

Deel I "Complexatie van DNA in virusachtige deeltjes"

In hoofdstuk 2 beschrijven we de recombinante productie van een diblok PbP met een grote hydrofiele sequentie voor colloïdale stabiliteit en een korte kationische sequentie voor elektrostatische complexering. Vanwege de extreme asymmetrie in de lengtes van de twee blokken, zijn complexen van deze PbPs met (monodispers) DNA langwerpige staafvormige structuren van dezelfde lengte, die op staafvormige virus lijken. Wanneer DNA is bekleed met de diblok PbPs, krijgt het een "bottle-brush" corona dat de colloïdale stabiliteit geeft en dus samenklontering voorkomt voor minstens enkele dagen. De lengte van de co-geassembleerde structuren wordt bepaald door de contourlengte van de DNA keten, zodat de gecoate DNA zeer geschikt wordt voor gebruik als een sjabloon voor het maken van verdere nanomaterialen. De diblok PBPs helpen ook om DNA-moleculen uit te rekken doordat ze de keten stijver maken. Het diblok PBPs kan, tenslotte, als mal fungeren die uiteindelijk kan worden ontwikkeld tot kunstmatige virussen.

In hoofdstuk 3 gaan we dieper in op het diblokontwerp van hoofdstuk 2 en produceren we vier nieuwe triblok PbPs met een zijde-achtige zelfassemblerend middenblok (elk PbP met een andere lengte van het zelf-assemblerende blok). De zelf-assemblerende domeinen interacteren zijdelingse met een kracht die afhangt van de lengte. We vinden dat triblok PbPs met een lengte van het zelfassemblerende blok groter dan een kritische waarde, zich gedragen als kunstmatige virale manteleiwitten. Deze triblok PbPs binden coöperatief aan een enkele DNA keten en verpakken deze in staafvormige virusachtige deeltjes (VLPs). In de VLPs wordt DNA beschermd tegen enzymatische digestie, en VLPs leveren efficiënt DNA aan menselijke kankercellen voor expressie. De kinetiek van inkapseling bleek zeer vergelijkbaar met die van het *Tobacco Mosaic Virus* te zijn en kan worden beschreven met hetzelfde kinetische model. Thermodynamische parameters voor de kinetiek van VLP-vorming bleken vergelijkbaar met de voor de in vitro vorming van TMV. Kortom, de VLPs gedragen zich in veel opzichten als kunstmatige virussen. We concluderen dat we met eenvoudige vuistregels de fysieke functionaliteiten van natuurlijke eiwitten kunnen vertalen

in het ontwerp van minimalistische eiwtiblokken die de assemblage en een aantal van de functionele eigenschappen van virussen nabootsen. Zo'n biomimetische supramoleculaire aanpak is zeer veelbelovend voor het nabootsen van andere natuurlijke functionele supramoleculaire structuren.

Deel II "Toepassingen van eiwit-DNA-complexen"

In hoofdstuk 4 hebben we aangetoond dat door het bekleden van een enkele DNA-ketens en deze in nanokanalen te doen, het mogelijk is om fluorescente markers langs de DNA-keten op te lossen. Dit wordt mogelijk gemaakt door het verstijvende effect van de diblok PbPs (hoofdstuk 2) dat een “bottlebrush” rond het DNA vormt. Wanneer de diameter van de nanokanalen lager is dan de effectieve persistentielengte van het beklede DNA, kan deze niet meer buigen in de nanokanalen en wordt dus uitgerekt. Het in beeld brengen van enkele DNA-ketens is gedaan met een fluorescentiemicroscoop met een ultragevoelige camera. In dit hoofdstuk is dus het “proof-of-concept” beschreven voor het gebruik van diblok PBPs voor het verbeteren van het uitrekken van DNA-ketens in nanokanalen, om zo DNA-ketens optisch in kaart te brengen.

In hoofdstuk 5 ontwerpen en produceren we een recombinant diblok PbP dat een DNA-bindingsdomein, genaamd " Sso7d ", bestaande uit 63 aminozuren, bevat. Het DNA-bindende domein is gekoppeld aan een lange hydrofiel blok (~ 800 aminozuren) die afscherming en colloïdale stabiliteit biedt. In tegenstelling tot het oorspronkelijke DNA-bindende diblok PbP in hoofdstuk 2 geeft dit blok de specificiteit voor DNA-interactie. Volledige bedekking van verschillende afzonderlijke DNA-ketens is bereikt: dubbelstrengs lineair DNA, dubbelstrengs circulair DNA en enkelstrengs DNA. We vinden een aantal opvallende verschillen in vergelijking met de niet-specifieke DNA-bindende diblok PbPs : DNA draait door binding van de BSso7d en dit leidt tot leidt tot het ontrollen van supercoiled DNA, zodat er cirkelvormige complexen ontstaan. De niet-specifieke PbPs leiden daarentegen tot complexen in de vorm van dikke staven. Bij binding aan enkelstrengs DNA, kan het niet-specifieke PbP intramoleculaire baseparing niet volledig zodat de resulterende complexen verschijnen als compacte en zeer vertakte structuren. In tegenstelling, het DNA specifieke diblok PbP kan bijna volledig baseparing voorkomen, zodat complexen met cirkelvormige enkelstrengs DNA inderdaad cirkels worden. Aangezien de beklede enkelstrengs DNA vrij stijf is, opent dit de mogelijkheid om lange enkelstrengs DNA-ketens in nanokanalen te rekken, waardoor optische analyse mogelijk wordt. Tenslotte tonen we aan dat het diblok PbP met BSso7d een effectieve nanocarrier van DNA in cellen vorm, en niet toxisch noch hemolytisch is. Onze resultaten tonen aan dat het combineren van natuurlijk gevouwen DNA-bindende domeinen met PbPs veelbelovend functionele biomaterialen opleveren. In het algemeen kunnen we stellen dat de resultaten veel nieuwe toepassingsmogelijkheden laten zien die kunnen ontstaan bij het combineren van de unieke eigenschappen van gevouwen natuurlijke eiwitten met die van PbPs.

Deel III "Supramoleculaire structuren voorbij DNA".

In hoofdstuk 6 geven we een gedetailleerde karakterisering van de zelfgeassembleerde nanostructuren van de triblok PbPs van hoofdstuk 3 in de afwezigheid van DNA. We karakteriseren de afhankelijkheid van de zelfassemblage op tijd en eiwitconcentratie. Bij lage concentraties blijken de triblok PbPs onregelmatige micellen te vormen, en we vinden een goed gedefinieerde kritische concentratie. Bij hogere concentraties assembleren de triblok PbPs in goed gedefinieerde nanostaafjes, met een lengte die afhankelijk is van de concentratie en de lengte van de zijde-achtige middenblok. Vooral de laatste correlatie is zeer relevant voor het verbeteren van de ontwerpregels voor zelfassemblerende PbPs. Tot slot laten we zien dat de zelf-geassembleerde nanostaafjes zeer stabiel zijn tegen verdunning. De zelf-geassembleerde nano-eiwitstaafjes bieden een interessant platform voor nanomedicijnen.

Hoofdstuk 7 beschrijft de hiërarchische co-assemblage van de kationische triblok PbPs van hoofdstuk 3 met negatief geladen supramoleculaire polymeren die worden bijeengehouden door metaalionen. Dit resulteert in zeer regelmatige eiwit-metaal nanostaafjes. Deze nieuwe zelf-geassembleerde nanostructuren kunnen gemaakt worden met een grote verscheidenheid van metaalionen en zijn zeer stabiel, zelfs bij hoge zoutconcentraties, zodat zij toepasbaar zijn in bijvoorbeeld nanomedicijnen of moleculaire elektronica.

Hoofdstuk 8 beschrijft de karakterisering van de elektrostatische co-assemblage van het triblok PbP van hoofdstuk 3 met meerdere negatief geladen polyelektrolyten van verschillend molecuulmassa's en met verschillende stijfheid. Al deze polyelectrolyten kunnen worden gebruikt voor de vorming van staafvormige virusachtige deeltjes, zoals eerder gevonden met DNA (hoofdstuk 3). De polyelektrolyten die werden getest zijn polyacrylzuur, polysulfonaat en xanthaan. Uit uitgebreide AFM experimenten concluderen we dat de pakkingsgraad (contourlengte van het staafvormige virusachtige deeltje gedeeld door de absolute lengte van het polyelectrolyt) afhankelijk is van de intrinsieke eigenschappen van het polyelektrolyt, met name van de ladingsdichtheid en stijfheid. We concluderen dat de co-assemblage van polyelektrolyten met triblok PBP een zeer eenvoudige manier is om een verscheidenheid van lineaire nanostructuren te produceren die veel potentiële toepassingen heeft in de synthese van nanogestructureerde materialen.

Acknowledgements

This research forms part of the research programme of the Dutch Polymer Institute (DPI), project #698. I thanks its funding and all the support given to finish successfully my PhD project. I thank to my dear supervisor Renko for trusting me since the initial interview in Lund, Sweden. You believed in me for the realization of such an ambitious PhD project. We had no idea how far we could take it. I am so glad to have had you as supervisor. Also, I thank you for having opened your house and introduced to me your family the first day I came to Netherlands. I am sure it was not easy to open the door of your home to a stranger and foreigner. That night my adventure as PhD student started. I also thank the freedom that you gave me to carry the project, the very fruitful and enrichment discussions and meetings that we had; especially in the last year, when I became so overwhelmed with so many things. I hope we can keep collaborating in the future. I thank to Martien, my promotor, you have been the paternal figure in the whole FYSKO lab, always leading the navy into quite and rich new oceans. Your open, friendly, fresh, wise and flexible character has always impressed me. I enjoyed the multiple discussions we had and your highly remarkable scientific comments and suggestions that you have always made. Thanks for trusting in me and for all the support given all the way along my PhD. I hope we keep in touch for the time to come.

To the Eindhoven collaborators, Paul and Saber. Paul your humor sense always brought a relaxing moment, especially during the tedious DPI meetings that have a reserved community of scientists. Also, thanks for the scientific coaching, and for providing the virus self-assembly model that has been a powerful contribution to our high impact manuscript, also for contacting Daniela to carry out such job. Your firm commitment with the development of science has always impressed me and it is strongly inspiring.

To Frits and Marc, from FBR, thanks for providing a professional environment that it was more than enough to develop such a fantastic molecules that I call protein polymers. Thanks to introduce me to the *Pichia pastoris* expression system.

To the FYSKO people generation 2009, Suomi, Yunus, Thao, Dmitry, Huanhuan, Monica, Liyakat and Emilia + Yuan Li. I have no words to explain how nice it was to share the time with such an open and friendly gang as you guys. You were such a good company during 4 years and especially during the Southeast Asia PhD trip that I will never forget. Evan, thanks to answer very quick and accurately my e-mails and questions. Yuan Li, thanks for all the jokes made on purpose or not, for the personal invitation to get the chinese visa and for the nice moments together during the trip to Thailand and Cambodia. Dmitry, thanks for all the intricate, deep and passionate discussions we had about almost everything, science, religion, moral, sex, god, prejudices, etc, etc, and for the nice company during the Cambodia and Thailand trip. I still think passions dominate more our world

that the reason. Suomi and Surrender, thank to share your Indian wisdom, strong character and friendly approach to life. I learnt many things from that, also for the dinners, the drinks and all the good moments. Yunus, thanks for being more than a colleague. You have shared always your point of view sincerely and honestly, as it is expected from a cool guy. Thao, you were the little cute friend that I had since we shared the office in FBR, when all our life was cloning and dealing with restriction enzymes and agarose gels.

Thanks to the newer generation of FYSKO members that have enriched my life. Maria, the sincerest and most friendly Greek I have ever met, you have always something nice to talk and to say. Christian, you have an extraordinary ability to empathize with the people. Thanks for sharing and being a nice friend in and outside the lab. I know that deeply inside you are more Latin than German. Jacob, "the gentleman", thanks for the crazy jokes and remarks you brought to make a nice atmosphere to the lab. "Gosias", you brought the cute and sweet polish style to the lab that nobody could rival with it. Rui, the man of the thousand impolite jokes, with them you made me to forget the formality and closeness of many people around. Sabine, you always had a very good and optimistic vibe. To the collaborators in NUS, Prof. Johan and Ce, thanks for the nice work and help during my staying in Singapore.

To the PhD trip committee organizers (Sabine, Christian and Yunus); thanks to our restless work we have organized the most AMAZING trip that FYSKO will ever see. California rocks! It was a difficult task, full of difficulties and misunderstandings but at the end the sincerity and good intentions remains. We all learnt the lesson about cultural differences (I hope). Josie, you were always there for any silly or complicate question and for offering kindly support words when I needed them. Mara, thank to get on time all the chemicals, keep doing such an important job. To my calm and hard worker officemate, Junyou, I have learnt plenty of things about chinese people. You were the only one at FYSKO that has indeed shared most of the time next to me. We shared the office 0016 during 4 years and we could materialize a real scientific collaboration in the group without involving our bosses. I hope our work that has been slowly cooked (during years) finally gets published somewhere. To all the BSc. students that I supervised, Anne, Niek, Ilse and Marcel; I always hoped to give you more time, although I always did my best. All of you have done such amazing thesis. I am so proud of your work. To all the FYSKO members (staff and students), you have always taught me always something new in the good and in the bad moments.

To you, random Dutch person, that have inspired many thoughts about how different can be the world and the cultures, but on the other hand, that how many similarities we share, although you don't want to see them.

To all my amazing Mexican friends I met all the way along during my PhD in Wageningen, Vicente, Emilio, Ale, Ilane, Laura, Gibran, Daniela, Diana, Sugued, Arturo (in Wien), Max, Oswaldo, Yadira, Gabriela, Jennifer, Carolina and Edward and Majorlein

(Mexican adopted). You have been my little Mexico abroad. We have kept the best Mexican food only for us ☺.

Vicente, thanks for all the good moments, it has been good to have such a comrade and friend all this time. We could share and talk almost about anything. I cannot forget the breathtaking trip that we did from Saint Petersburg to Beijing during the 2012 summer. Also thanks to materialize the amazing short movie about Brain Drain.To the crazy "Flash Mob" team (you can see it in YouTube). It was so funny to have a pillow fight in Forum building during lunch.

To my dears Anita and Svetlana, we were such a close family that shared so many things, especially during the Winter of 2009 to Summer of 2010. I won't forget you, and of course also to the crazy Czechs, with all you I had more parties, meetings and dinners than ever. To Pelin, the life indeed is a kite. To the "jam sessioners", Yunus and Kris, there will be a time that our "jam music" will become a musical jewel. To my dear friend Naiara, I have no words to tell you how grateful has been to meet you. During these four years you were always there to have a cup of coffee, a beer, to cook a dinner, or to have a walk. To Jazmin, the little Peruvian girl that nothing could stop her. You have been a central piece of this puzzle that I call life.

To my dear friends in México, Arturo, Fabiola, Mónica, Yololxochitl, Victoria, Victor (si, a pesar de todo), Gustavo, Rosa, Francisco y Esther, Lucia, Lorena y demás. To keep in contact has been always difficult, but the real friendship that we have created cannot be destroyed easily in spite of the longest distance and the largest waiting time.

To my dear Corina, you have been such a shinny sun in a stormy day, and also a fantastic partner all the time. You are an oasis in a dessert that I expect to keep it near to me. I hope you still bring such happiness to my life and I hope I can do the same with you. We are a fantastic team!

To my beloved family (Mama, Papa, Yadira, Erick, Ariel). You have been such a great company and supporters. You have always been in the most difficult and in the happiest moments of my life. Mama y Papa, I have built such an incredible story thanks to your extremely hard and dedicated work to put forward my education and in spite of all the hurdles and limitations you had. That is the purest act of love you have done.

To the people that I met during all my trips around the world, young guests in hostels, passengers in trains, buses or airplanes, walkers in open fields. In a way you are the most valuable treasure I have had during all these trips. The curiosity to meet you is one of the reasons that keep me traveling.

A mi país que aun creo en él y sobre todo en su gente, la humilde, la honesta, la trabajadora, la que te abre su corazón, su mente, sus ideas, su casa y que no se rinde ante nada. Siempre seré Mexicano y con orgullo. Finalmente, espero que este gran esfuerzo personal de crecimiento y aprendizaje que he tenido durante 6 años sirva de algo para hacer que en México haya al menos un pequeño cambio hacia adelante.

List of publications

This thesis:

1. Hernandez-Garcia, A.; Werten, M. W. T.; Stuart, M. C.; de Wolf, F. A.; de Vries, R., Coating of Single DNA Molecules by Genetically Engineered Protein Diblock Copolymers. *Small* **2012**, *8*, 3491-3501.

2. Zhang, C.; Hernandez-Garcia, A.; Jiang, K.; Gong, Z.; Guttula, D.; Ng, S. Y.; Malar, P. P.; Kan, J. A. v.; Dai, L.; Doyle, P. S.; Vries, R. d.; Maarel, J. R. C. v. d., Amplified stretch of bottlebrush-coated DNA in nanofluidic channels. *Nucleic Acids Research* **2013**.

3. Hernandez-Garcia, A.; Kraft, D. J.; Janssen, A. F. J.; Bomans, P. H. H.; Sommerdijk, N. A. J. M.; Favretto, M.; Brock, R.; Wolf, F. A. d.; Werten, M. W.; Schoot, P. v. d.; Stuart, M. C.; Vries, R. d., Supramolecular design of a minimal coat-protein for an artificial virus. *Submitted* **2013**.

4. Hernandez-Garcia, A.; Favretto, M.; Brock, R.; Wolf, F. A. d.; Vries., R. d., Protein polymer incorporating a DNA binding domain: coating, stiffening and delivery of single DNA molecules. *Submitted* **2013**.

5. Hernandez-Garcia, A.; Wang, J.; Rombouts, W.; Velders, A.; Cohen-Stuart, M.; Vries, R. d., Self-assembled protein-metal ion nanorods. *Manuscript* **2013**.

6. Hernandez-Garcia, A.; Wolf, F. A. d.; Stuart, M. C.; Vries, R. d., Supramolecular nanorods from triblock protein polymers. *Manuscript* **2013**.

7. Hernandez-Garcia, A.; Stuart, M. C.; Vries, R. d., Templating of artificial viral coat proteins by a range of negatively charged polyelectrolytes. *Manuscript* **2013**.

Other work not included in this thesis:

8. Lindman, S.; Hernandez-Garcia, A.; Szczepankiewicz, O.; Frohm, B.; Linse, S., In vivo protein stabilization based on fragment complementation and a split GFP system. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 19826-19831.

9. Golinska, M. D.; de Wolf, F.; Stuart, M. A. C.; Hernandez-Garcia, A.; de Vries, R., Pearl-necklace complexes of flexible polyanions with neutral-cationic diblock copolymers. *Soft Matter* **2013**, *9*, 6406-6411.

About the author

Armando Hernandez Garcia was born on 7th of October of 1981 in Mexico City, Mexico. He obtained his high school diploma from Colegio de Ciencias y Humanidades (Vallejo) in 2000 and then started his BSc. studies in Food Chemistry at National Autonomous University of Mexico (UNAM). He carried out his BSc. thesis with Prof. Amanda Galvez and defended it in June, 2007. He was working managing clinical data for clinical trials during a year for a CRO just before starting his MSc. in Biotechnology at Kemicentrum at Lund University (LU), Lund, Sweden. He carried out for 3 months an advanced course at Immunotechnology department of LU doing Phage Display of carbohydrate binding proteins. In January 2009 he started his MSc. thesis at Biophysical Chemistry Department with Prof. Sara Linse where he managed to develop an *in vivo* split GFP biosensor to scan for stable variants of Protein GB1. He graduated in September, 2009 and immediately he started his PhD studies at Laboratory of Physical Chemistry and Colloid Science at Wageningen University, the Netherlands under supervision of Dr. Renko de Vries and Prof. Martien Cohen Stuart with a project founded by the Dutch Polymer Institute and a personal grant for postgraduate studies from CONACyT, Mexico. He has applied to the prestigious PEW Latin America Fellowship in Biomedical Sciences to carry out Postdoctoral studies at Laboratory of Samuel I. Stupp in Northwestern University in Chicago, USA. This thesis is the result of his PhD research. He uses to think that if you do not try out something new you never know how amazing and powerful it could be.

Frontispiece in SMALL, vol. 8, issue 22, November, 2012.

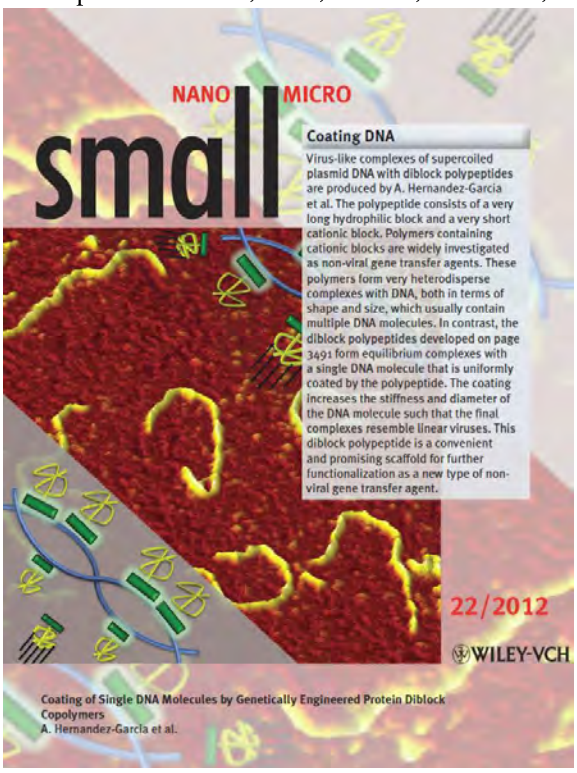


Photo by Armando Hernandez-Garcia

Overview of completed training activities

General Courses

- VLAG - PhD week, 2010.
- Scientific Writing, WUR, Wageningen, 2011.
- Communication in Interdisciplinary Research, WUR, Wageningen, 2011.
- Information Literacy PhD including EndNote Introduction, WUR, 2011.
- Advance Course Guide to Scientific Artwork, WUR, Wageningen, 2011.

Discipline specific

Scientific Courses

- Han-sur-Lesse Wither School, Belgium, January, 2010.
- International Workshop in Physical Chemistry of Biointerfaces II, CIC Biomagune, San Sebastian, Basque Country, Spain, 2012.
- Training in Biophysics of Complex Fluids, NUS, Singapore, 2013.

Meeting Oral Presentation

- European Polymer Congress, Granada, Spain, 2011.
- Chemistry as an Innovative Science, NWO. Maarsen, Netherlands, 2011.
- Dutch Polymer Days, NOW, Lunteren, the Netherlands, 2013.
- Molecular and Cellular Biophysics, Veldhoven, the Netherlands, 2013.
- Becarios CONACyT en Europa, Strasbourg, France, 2013.

Meeting Posters

- Gordon Research Conference Bioinspired Materials, NC, USA, 2012.
- 8th International Conference and Exhibition RNAi. Oxford, UK, 2013.
- Self-assembly of Biomolecules, RCS, Bristol, UK, 2013.
- New Frontiers Symposium in Synthetic Life, Nijmegen, Netherlands, 2013.

Meeting Attendance

- 11th European Symposium on Controlled Drug Delivery, Netherlands, 2010.
- Black Forest Focus on Soft Matter 5, Saig, Germany, 2011.
- Dutch Polymer Days, Lunteren, Netherlands, 2012.
- DPI annual meeting 2010-2012.

Other Activities

- Group Meetings and Colloquia, 2009-2013.
- PhD trip, Southeast Asia, February, 2011.
- PhD trip to California, February, 2013 (Organization and attendance).
- Advanced Soft Matter, WUR, Wageningen, 2010.
- Dutch Polymer Institute, Corporate Cluster Review Meeting, 2010-2013.

This research forms part of the research programme of the Dutch Polymer Institute (DPI), project #698.

Dutch Polymer Institute (DPI), P.O. Box 902, 5600 AX Eindhoven, The Netherlands

Idea original of cover by Armando Hernandez Garcia.
Cover design realized by Marius-Alexandru Surugiu.

Printed by Wohrmann Print Service, Zutphen.