LETTERS
PUBLISHED ONLINE: 24 AUGUST 2014 | DOI: 10.1038/NNANO.2014.169
nature nanotechnology

Design and self-assembly of simple coat proteins for artificial viruses

Armando Hernandez-Garcia[1,2†*], Daniela J. Kraft[3,4], Anne F. J. Janssen[1], Paul H. H. Bomans[5], Nico A. J. M. Sommerdijk[5], Dominique M. E. Thies-Weesie[6], Marco E. Favretto[7,2], Roland Brock[7], Frits A. de Wolf[8], Marc W. T. Werten[8], Paul van der Schoot[9,10], Martien Cohen Stuart[1] and Renko de Vries[1,11*]

Viruses are among the simplest biological systems and are highly effective vehicles for the delivery of genetic material into susceptible host cells[1]. Artificial viruses can be used as model systems for providing insights into natural viruses and can be considered a testing ground for developing artificial life. Moreover, they are used in biomedical and biotechnological applications, such as targeted delivery of nucleic acids for gene therapy[1,2] and as scaffolds in material science[3–5]. In a natural setting, survival of viruses requires that a significant fraction of the replicated genomes be completely protected by coat proteins. Complete protection of the genome is ensured by a highly cooperative supramolecular process between the coat proteins and the nucleic acids, which is based on reversible, weak and allosteric interactions only[6–9]. However, incorporating this type of supramolecular cooperativity into artificial viruses remains challenging[10–15]. Here, we report a rational design for a self-assembling minimal viral coat protein based on simple polypeptide domains. Our coat protein features precise control over the cooperativity of its self-assembly with single DNA molecules to finally form rod-shaped virus-like particles. We confirm the validity of our design principles by showing that the kinetics of self-assembly of our virus-like particles follows a previous model developed for tobacco mosaic virus[9]. We show that our virus-like particles protect DNA against enzymatic degradation and transfect cells with considerable efficiency, making them promising delivery vehicles.

Natural viruses, such as the tobacco mosaic virus (TMV), exhibit a high degree of cooperativity when self-assembling on their nucleic acid templates, with an initial binding event of coat proteins to the template rendering subsequent interactions much more favourable[16]. For the case of TMV, the cooperativity arises through allosteric conformational switching of the coat proteins upon binding to nucleic acids. The assembly process starts at a specific region on the RNA template called the origin of assembly[6]. Taking into account allosteric interactions and a specific origin-of-assembly region, we recently put forward a model that accurately describes the assembly kinetics of TMV *in vitro*[9]. Based on these mechanistic insights and inspired by the structure of the TMV coat protein[6–8,17], we 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 B = K_{12} that binds non-sequence-specifically through electrostatic interactions[18]. Precisely tuned cooperativity is provided by a silk-like sequence S_n = $(GAGAGAGQ)_n$ that folds and stacks in solution into stiff filamentous structures[19,20]. The number of silk strands n dictates the level of cooperativity of the binding of the protein and is therefore our key variable. To prevent aggregation of the assembled artificial virus-like particles (VLPs) we include a ~400-amino-acid-long hydrophilic random-coil sequence C with a high fraction of prolines and hydrophilic (mostly uncharged) amino acids. The C block was initially developed as a highly hydrophilic, non-gelling recombinant gelatin[21]. It does not contain motifs that interact with the environment, such as any cell-binding motifs, but could, if desired, be modified to establish interactions with the environment other than just the steric repulsion that we rely on here.

Proteins C–S_n–B (shown schematically in Fig. 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 Supplementary Tables 1 and 2 and Supplementary Fig. 1.

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 do the coat proteins of natural viruses. Indeed, binding to nucleic acids leads to the cooperative formation of compact, rod-shaped 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

[1]Laboratory of Physical Chemistry and Colloid Science, Wageningen University, Dreijenplein 6, 6703 HB Wageningen, The Netherlands, [2]Dutch Polymer Institute, John F. Kennedylaan 2, 5612 AB Eindhoven, The Netherlands, [3]Soft Matter Physics, Huygens-Kamerlingh Onnes Laboratory, Leiden University, PO Box 9504, 2300 RA Leiden, The Netherlands, [4]Center for Soft Matter Research, Department of Physics, New York University, 4 Washington Place, New York, New York 10003, USA, [5]Laboratory of Materials and Interface Chemistry & Soft Matter CryoTEM Research Unit, Department of Chemical Engineering and Chemistry, and Institute for Complex Molecular Systems, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands, [6]Utrecht University, Van't Hoff Laboratory for Physical and Colloid Chemistry, Debye Institute of Nanomaterials Science, PO Box 80.051, 3508 TB Utrecht, The Netherlands, [7]Department of Biochemistry, Radboud Institute of Molecular Life Sciences, Radboud University Medical Centre, Geert Grooteplein 28, 6525 GA Nijmegen, The Netherlands, [8]Wageningen UR Food & Biobased Research, Bornse Weilanden 9, 6708 WG Wageningen, The Netherlands, [9]Theory of Polymers and Soft Matter, Department of Applied Physics, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands, [10]Institute for Theoretical Physics, Utrecht University, Leuvenlaan 4, 3584 CE Utrecht, The Netherlands, [11]Department of Biomedical Engineering, University Medical Center Groningen, University of Groningen, PO Box 196, 9700 AD Groningen, The Netherlands; [†]Present address: Institute for BioNanotechnology in Medicine, Northwestern University, Chicago, Illinois 60611, USA. *e-mail: armaquim@gmail.com; renko.devries@wur.nl

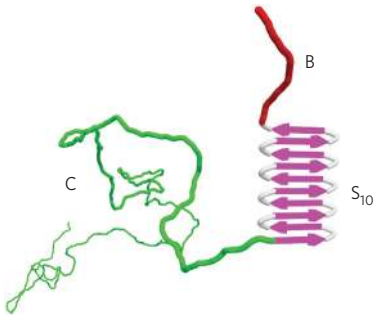

Figure 1 | Design of the minimal viral coat protein C–S_n–B. Shown is the case for $n = 10$, where B = K_{12} is the N-terminal dodecalysine DNA-binding block (red), S_{10} is a tenfold repetition of an octapeptide S = (GAGAGAGQ) that constitutes the self-assembly block (pink), and C is a C-terminal 407-amino-acid hydrophilic random coil block (green).

approaches for artificial viruses that do not feature precise control over cooperativity[2,11–15,22,23].

A small number of repeats n of the central block S_n is insufficient to induce the cooperative formation of compact VLPs. Complexes of linear double-stranded DNA (dsDNA) with proteins C–S_n–B ($n = 2, 4$) have contour lengths that are exactly equal to that of the dsDNA template, and retain much of its original flexibility (Fig. 2b,c). These complexes have the same 'bottle brush' structure that we have previously found[18] for the protein C–S_0–B without the central block (Fig. 2a): a single DNA chain surrounded by a dense array of pendant side chains (hydrodynamic thickness in solution, ~15 nm; Supplementary Fig. 2). Other groups have also demonstrated non-cooperative coating of single nucleic acid molecules by polymers featuring DNA-binding blocks in combination with large hydrophilic shielding blocks[12,22–24].

We do, however, achieve full cooperativity with C–S_n–B proteins 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 Fig. 2d–g (Supplementary Figs 3 and 4). Over a wide range of template lengths, DNA in VLPs is compacted to about one-third of its original length (Supplementary Figs 5 and 6). This compaction factor is controlled by charge stoichiometry: from the solution molar mass of VLPs formed with C–S_{10}–B, as determined using static light scattering, we estimate that 95% of the DNA phosphate charges are neutralized by compensating charges on binding blocks B (Supplementary Fig. 7). To further test the idea that charge stoichiometry determines the compaction factor we also prepared VLPs using single-stranded DNA (ssDNA) templates. These have a linear charge density that is smaller than that of dsDNA, by a factor of approximately 2. As expected, the ssDNA templates form VLPs with a compaction factor of around one-sixth (Supplementary Fig. 8).

Similar to natural viral capsid proteins, the VLP-forming proteins also tend to form rod-like self-assemblies in the absence of a template (Supplementary Fig. 9), but only above a well-defined critical aggregation concentration (CAC). Below the CAC, protein monomers coexist with much smaller protein micelles. For C–S_{10}–B the CAC as detected by static light scattering is ~80 μM (or 3.6 mg ml^{-1}), whereas for C–S_{14}–B it is ~12 μM (or 0.56 mg ml^{-1}). For all experiments with VLPs, except cryogenic transmission electron microscopy (cryo-TEM) experiments, protein concentrations were much lower than the CAC.

The stiff rod-like appearance of the VLPs strongly suggests that the silk-like mid-blocks fold and stack to form a stiff fibre. Previously, we have shown that a signature of the folding and stacking of similar silk-blocks in other proteins is a decrease in the deep random coil minimum in the circular dichroism (CD) spectra of the proteins at 200 nm (ref. 20). Indeed, during the formation of VLPs with C–S_{10}–B we find a similar decrease of the CD spectra at 200 nm, whereas no changes in the CD spectra are observed for the C–S_0–B protein that does not contain the silk-like midblock (Supplementary Figs 10 and 11).

Like natural viruses, the VLPs protect their nucleic acid cargo against enzymatic attack and are capable of delivering their cargo into cells. To assess the protection, we incubated VLPs with high concentrations of DNAse I and analysed the degradation using agarose-gel electrophoresis. Conditions were chosen such that no intact bare DNA was present after 1 min 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–S_n–B proteins with $n = 10$ and 14 resist attack for periods longer than 60 min (Supplementary Fig. 12). This suggests a link between compaction and protection of the nucleic acids and the cooperativity of the VLP assembly, which we investigate in more detail below.

Transfection of cells by complexes of plasmid DNA with C–S_n–B proteins 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

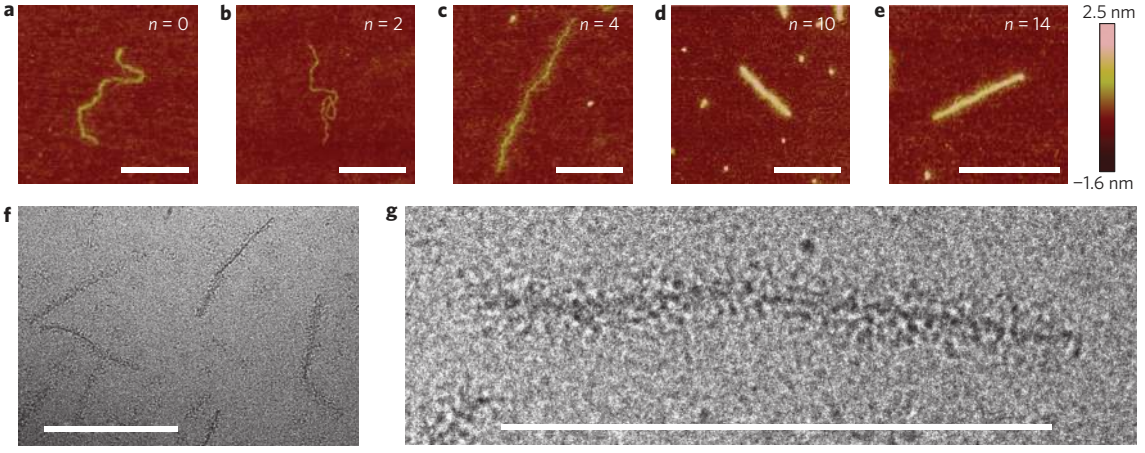

Figure 2 | Self-assembly of VLPs: AFM and cryo-TEM images of complexes of linear dsDNA with C–S_n–B show morphologies that depend on the size n of the self-assembly block. a–c, Coating of DNA templates. Contour lengths of complexes are close to that of the template DNA (850 nm): 831 nm ($n = 0$) (**a**), 842 nm ($n = 2$) (**b**), 817 nm ($n = 4$) (**c**). **d**,**e**, Formation of VLPs. Contour lengths of VLPs are 315 nm ($n = 10$) (**d**) and 304 nm ($n = 14$) (**e**). **f**,**g**, Cryo-TEM images of VLPs for $n = 10$. DNA was incubated with C–S_n–B protein for 24 h at a charge ratio N/P = 7. Scale bars, 300 nm.

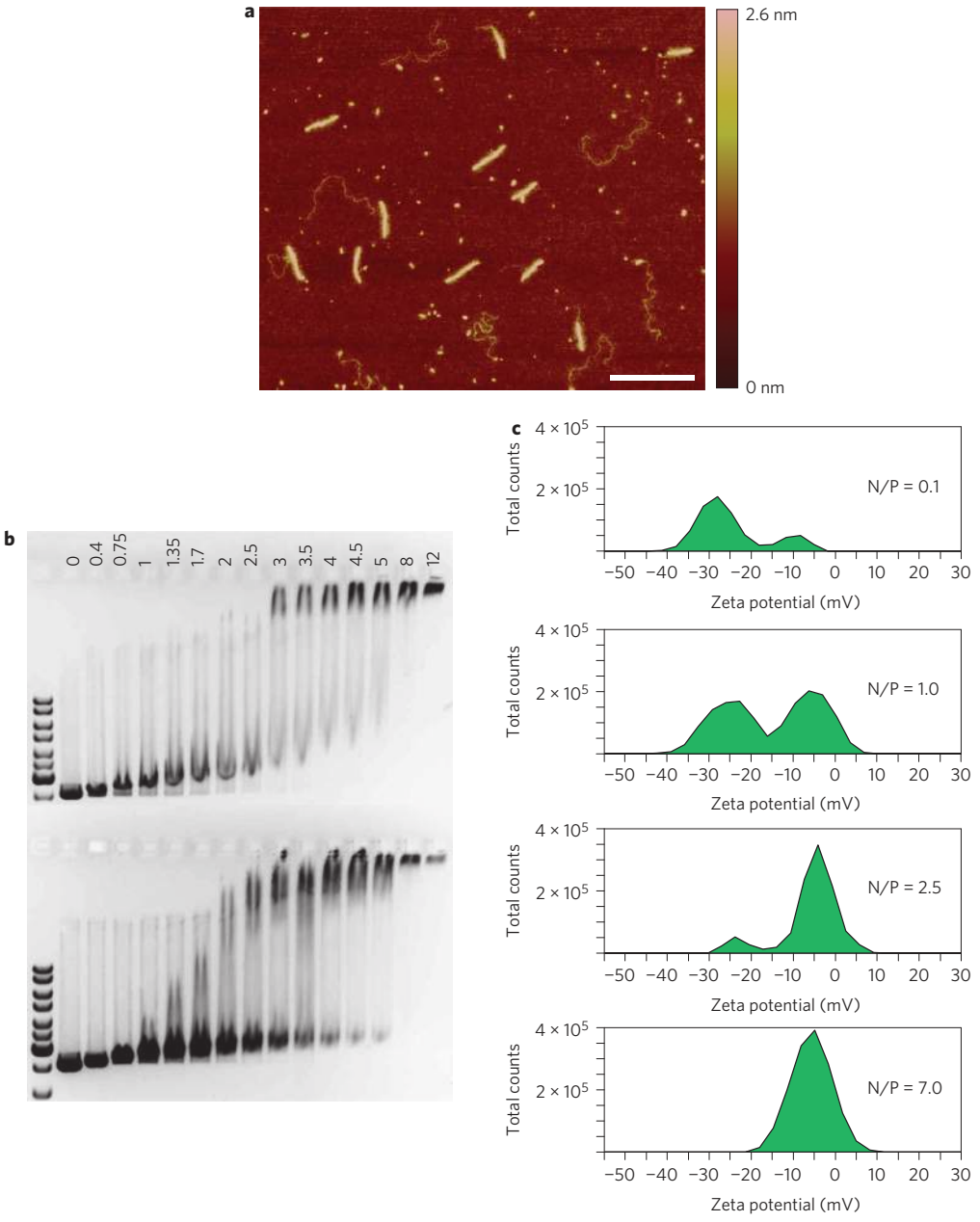

Figure 3 | Cooperativity of the self-assembly of VLPs. a, AFM image of incompletely assembled VLPs of C–S_{10}–B with linear dsDNA (incubation time 1 h, charge ratio N/P = 15). Smaller objects are protein-only assemblies. Scale bar, 500 nm. **b**, Electrophoretic mobility shift assay for C–S_0–B (top) and C–S_{10}–B (bottom) interacting with linear dsDNA. Charge ratios N/P for different lanes are indicated at the top. For C–S_0–B only a single type of complex is visible, but for C–S_{10}–B two types of complexes coexist. For both samples, incubation time is 1.5 h. **c**, Distributions of solution-based electrophoretic mobility (zeta potential) for C–S_{10}–B interacting with linear dsDNA at different N/Ps (incubation times, 30–60 min). The distributions show the coexistence of two types of complexes at intermediate N/Ps.

standards poly-ethyleneimine and Lipofectamine 2000 (Supplementary Fig. 13). Interestingly, we find that proteins without the central block C–S_0–B have the highest transfection efficiency and that it is somewhat lower for VLPs based on C–S_{10}–B. We speculate that the reduced transfection efficiency is related to the more difficult disassembly that is concomitant with the increased protection due to cooperativity and the associated compaction of VLPs. This ties in with what is known about natural viruses, which have to strike a balance between the degree of protection and ease of disassembly[25].

A defining test for cooperativity is 'all-or-nothing'-type behaviour[16,26]. For the *in vitro* co-assembly of viral capsid proteins with nucleic acids[6,9], as well as for other proteins that cooperatively co-assemble with nucleic acids[27], this translates into the coexistence of fully encapsulated and bare nucleic acids. Atomic force microscopy (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 (Fig. 3a). Similarly, gel electrophoretic mobility shift assays (Fig. 3b) and determinations of the distribution of the bulk electrophoretic mobility (Fig. 3c) demonstrate the coexistence of high- and low-mobility complexes. Further evidence for the coexistence of VLPs and non-encapsulated DNA also comes from analytical ultracentrifugation (Supplementary Fig. 14).

Nucleation of VLPs mostly occurs at one end of the DNA, so the ends of the DNA act as an effective origin of assembly. The TMV model that motivated our design should therefore be quantitatively

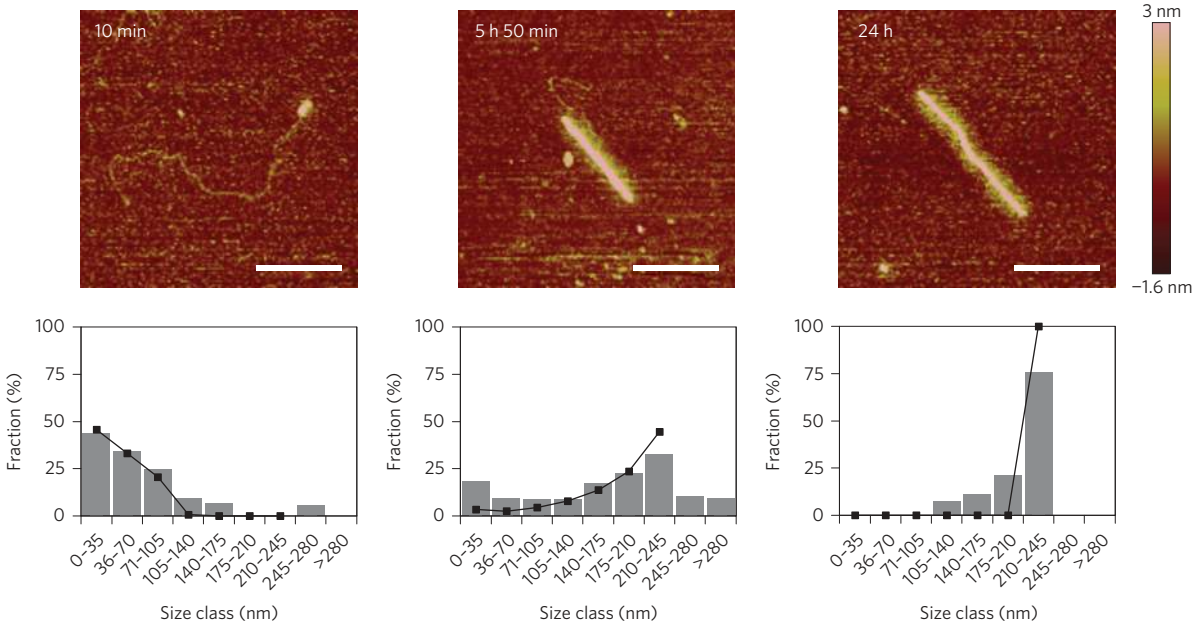

Figure 4 | Formation process of VLPs. Top: representative AFM images of VLPs of linear dsDNA with C–S_{10}–B protein for different incubation times (indicated in the figure; charge ratio N/P = 3). Scale bars, 200 nm. Bottom: size distributions of condensed sections as determined from AFM images (bars) and fit to kinetic model of TMV self-assembly[9] (symbols connected by lines).

applicable to the kinetics of our VLP assembly. We quantified the time evolution of the distribution of lengths of condensed sections of VLPs from AFM images (Fig. 4a–c). Characteristic timescales for VLP formation deduced from AFM match those found from bulk biophysical approaches (circular dichroism, Supplementary Fig. 11; static light scattering, Supplementary Fig. 15). The time evolution of the length distributions from AFM was compared with predictions of the aforementioned kinetic model[9]. The model assumes that a first subunit, which binds to an origin of assembly on a linear array of binding sites, pays a free energy penalty h due to conformational switching, while binding of subsequent subunits involves a favourable binding free energy, g. In addition, bound protein subunits gain a free energy ε per bond due to favourable protein–protein interactions. A comparison with our experimental data leads to estimated parameter values of $h - \varepsilon \approx 5.3 k_B T$ for the effective penalty for nucleation and $g + \varepsilon \approx -18 k_B T$ for the effective binding free energy in units of thermal energy $k_B T$ (Supplementary Fig. 16). These values are very similar in magnitude to those found for TMV[9], again emphasizing the striking similarity between the assembly of our artificial VLPs and natural viruses such as TMV.

In summary, 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[28] cooperatively form rod-shaped VLPs, which protect the encapsulated DNA and show significant transfection activity. The capsids that our proteins make are scaffold materials that may act as monodisperse templates in material science[5,24,29,30] and could be optimized for the delivery of nucleic acids[2]. In analogy with natural viruses, the cooperative self-assembly mechanism of our VLPs in principle allows for sequence-directed control of nucleation. For example, some degree of sequence specificity could be incorporated into binding block B, which would allow for a rudimentary form of phenotype–genotype linkage.

Methods

Production of proteins. 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. Dialysed and freeze-dried proteins were stored in sealed tubes. Purity and integrity were corroborated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and matrix-assisted laser desorption/ionization mass spectrometry.

DNA–protein complex formation. Complexes were prepared by mixing stock solutions of linear dsDNA 2,500 bp (NoLimits Thermo Scientific) and a stock solution of protein in buffer (10 mM phosphate, pH 7.4, containing dithiothreitol (DTT), 5 to 0.1 mM) at the desired protein-to-DNA 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 mixed by vortexing for a few seconds or pipetting a few times, and incubated at room temperature. Sonication was never used. Fresh protein stocks were prepared immediately before every experiment by dissolving a weighed portion of freeze-dried protein in Milli-Q water followed by vortexing for 1–3 min.

AFM. Typically, 3–5 µl of sample ([DNA] = 1 µg ml^{-1} in 10 mM phosphate buffer, pH 7.4 with 0.1 mM DTT, except for C–S_n–B with $n = 0$, for which DTT was omitted, protein/DNA charge ratios N/P as indicated) were deposited onto a clean silicon surface. After 2–3 min 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. Samples were analysed using a Digital Instruments NanoScope V equipped with a silicon nitride probe (Bruker) with a spring constant of 0.4–0.35 N m^{-1} in the ScanAsyst imaging mode. Images were recorded between 0.488 and 0.977 Hz and at 384–1,024 samples/line. Image processing and height and diameter measurements were performed with NanoScope Analysis 1.20 software. A first-order flattening was used for all images. Contour length measurements were performed using ImageJ software.

Cryo-TEM. Samples ([DNA] = 30 µg ml^{-1}) were prepared using a vitrification robot (FEI Vitrobot Mark III) using grids R2/2 (Quantifoil Micro Tools), previously surface-plasma-treated (Cressington 208 carbon coater). Cryo-TEM characterizations were performed on a CryoTitan (FEI) equipped with a field-emission gun operating at 300 kV and a post-column Gatan energy filter. Images were recorded using a 2,000 × 2,000 Gatan charge-coupled device (CCD) camera. Final images were prepared using ImageJ software.

Electrophoretic mobility shift assay. DNA–protein samples ([DNA] = 15 ng µl^{-1}) incubated for 1.5 h were electrophoresed in a 1% agarose gel (95 min/60 V) using 1 × TAE buffer (pH 8). Bands were visualized with ethidium bromide.

Solution-based electrophoretic mobility. Zeta potential measurements of DNA + C–S_{10}–B protein samples ([DNA] = 5 ng µl^{-1}, 10 mM phosphate, pH 7.4, 5 mM DTT) were performed using a Zetasizer NanoZS (Malvern Instruments) with a 4 mW He–Ne ion laser at a wavelength of 633 nm. The angle of detection was 17°. Samples were incubated for 30, 30, 35 and 59 min for N/P = 0.1, 1, 2.5 and 7, respectively, and measured over 17 min (ten runs each measurement) using disposable folded capillary cells with stoppers.

Further details of all methods are provided in the Supplementary Information.

Received 13 November 2013; accepted 17 July 2014; published online 24 August 2014

References

1. Naldini, L. *et al. In vivo* gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* **272,** 263–267 (1996).
2. Mastrobattista, E., van der Aa, M. A. E. M., Hennink, W. E. & Crommelin, D. J. A. Artificial viruses: a nanotechnological approach to gene delivery. *Nature Rev. Drug Discov.* **5,** 115–121 (2006).
3. Comellas-Aragones, M. *et al.* A virus-based single-enzyme nanoreactor. *Nature Nanotech.* **2,** 635–639 (2007).
4. Grigoryan, G. *et al.* Computational design of virus-like protein assemblies on carbon nanotube surfaces. *Science* **332,** 1071–1076 (2011).
5. Douglas, T. & Young, M. Host–guest encapsulation of materials by assembled virus protein cages. *Nature* **393,** 152–155 (1998).
6. Klug, A. The tobacco mosaic virus particle: structure and assembly. *Phil. Trans. R. Soc. Lond. B* **354,** 531–535 (1999).
7. Caspar, D. L. & Namba, K. Switching in the self-assembly of tobacco mosaic virus. *Adv. Biophys.* **26,** 157–185 (1990).
8. Butler, P. J. G. Self–assembly of tobacco mosaic virus: the role of an intermediate aggregate in generating both specificity and speed. *Phil. Trans. R. Soc. Lond. B* **354,** 537–550 (1999).
9. Kraft, D. J., Kegel, W. K. & van der Schoot, P. A kinetic zipper model and the assembly of tobacco mosaic virus. *Biophys. J.* **102,** 2845–2855 (2012).
10. Percec, V., Heck, J., Johansson, G., Tomazos, D. & Ungar, G. Towards tobacco mosaic virus-like self-assembled supramolecular architectures. *Macromol. Symp.* **77,** 237–265 (1994).
11. 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. USA* **92,** 1744–1748 (1995).
12. Aoyama, Y. *et al.* Artificial viruses and their application to gene delivery. Size-controlled gene coating with glycocluster nanoparticles. *J. Am. Chem. Soc.* **125,** 3455–3457 (2003).
13. Wagner, E. Strategies to improve DNA polyplexes for *in vivo* gene transfer: will 'artificial viruses' be the answer? *Pharm. Res.* **21,** 8–14 (2004).
14. 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.* **47,** 4525–4528 (2008).
15. 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.* **41,** 2562–2574 (2012).
16. Whitty, A. Cooperativity and biological complexity. *Nature Chem. Biol.* **4,** 435–439 (2008).
17. 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 Å resolution by X-ray fiber diffraction. *J. Mol. Biol.* **208,** 307–325 (1989).
18. Hernandez-Garcia, A., Werten, M. W., Stuart, M. C., de Wolf, F. A. & de Vries R. Coating of single DNA molecules by genetically engineered protein diblock copolymers. *Small* **8,** 3491–3501 (2012).
19. Krejchi, M. T. *et al.* Chemical sequence control of beta-sheet assembly in macromolecular crystals of periodic polypeptides. *Science* **265,** 1427–1432 (1994).
20. 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* **6,** 133–140 (2012).
21. 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.* **14,** 447–454 (2001).
22. 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* **110,** 4548–4554 (2006).
23. Osada, K. *et al.* Quantized folding of plasmid DNA condensed with block catiomer into characteristic rod structures promoting transgene efficacy. *J. Am. Chem. Soc.* **132,** 12343–12348 (2010).
24. 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.* **135,** 6211–6219 (2013).
25. Lecuyer, K. A., Behlen, L. S. & Uhlenbeck, O. C. Mutants of the bacteriophage MS2 coat protein that alter its cooperative binding to RNA. *Biochemistry* **34,** 10600–10606 (1995).
26. Williamson, J. R. Cooperativity in macromolecular assembly. *Nature Chem. Biol.* **4,** 458–465 (2008).
27. Morrone, S. R. *et al.* Cooperative assembly of IFI16 filaments on dsDNA provides insights into host defense strategy. *Proc. Natl Acad. Sci. USA* **111,** E62 (2014).
28. Rabotyagova, O. S., Cebe, P. & Kaplan, D. L. Protein-based block copolymers. *Biomacromolecules* **12,** 269–289 (2011).
29. Zhang, S. G. Fabrication of novel biomaterials through molecular self-assembly. *Nature Biotechnol.* **21,** 1171–1178 (2003).
30. King, N. P. *et al.* A general computational method is used to design protein building blocks that self-assemble into target architectures. *Science* **336,** 1171–1174 (2012).

Acknowledgements

The authors thank A. Westphal and I. van Hees for help with the fluorescence correlation spectroscopy experiments and R. Fokkink for help with static light scattering. The authors also thank M.W.P. van de Put for help in performing part of the cryo-TEM experiments and W. Kegel for suggesting the electrophoretic light scattering experiment. A.H-G. is financially supported by the Dutch Polymer Institute (DPI), project #698 SynProt, and by the Consejo Nacional de Ciencia y Tecnología (CONACyT), México. M.E.F. is supported by the Dutch Polymer Institute (DPI, Technology area HTE, project #730). M.C.S. is supported by the European Research Council (advanced grant no. 267254). D.J.K. acknowledges financial support through a Rubicon fellowship (grant no. 680-50-1019) from the Netherlands Organization for Scientific Research (NWO).

Author contributions

F.d.W., M.C.S., P.v.d.S. and R.d.V. conceived the initial idea. R.d.V. and A.H-G. conceived and designed the molecules and experiments. A.H-G. produced the proteins, except for C–S_2–B, which was produced by A.F.J.J. M.W. contributed to the production of proteins and to writing the technical sections on protein production. Except for the transfection experiments, analytical ultracentrifugation and cryo-TEM imaging, A.H-G. performed all experiments. A.H-G. and R.d.V. interpreted and analysed all experimental data except for the transfection experiments. P.v.d.S. and D.J.K. performed the theoretical analysis of the VLP assembly kinetics. R.B. designed the transfection experiments. M.F. performed and analysed the transfection experiments. D.M.E.T-W. designed, performed and analysed the analytical ultracentrifugation experiments. N.A.J.M.S. designed the cryo-TEM imaging experiments. P.H.H.B. performed the cryo-TEM imaging. A.H-G., D.J.K., P.v.d.S., M.C.S. and R.d.V. wrote the paper. All authors commented on the manuscript.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.H.G. and R.d.V.

Competing financial interests

The authors declare no competing financial interests.