Chapter 5 | Real-time assembly of virus capsids onto DNA

Abstract
Viruses are the cause of many diseases and to efficiently cure these diseases the working mechanisms of viruses have to be better understood. Often, they self-assemble around a piece of genomic material inside the host cell. This self-assembly process is poorly understood and can be essential for drug development. In addition, viruses also have been tested and are already being used as drug delivery systems. Here, we exam the suitability of AFS to study the assembly process of a natural virus and of an artificial virus-like particle onto a DNA tether. AFS is stable over long time, it is an intrinsic force clamp and allows to study many molecules in parallel, making it an ideal technique to follow the dynamics of the assembly process over time. Combining our results with data obtained from optical tweezers and atomic force spectroscopy leads to a deeper understanding of the self-assembly process of these viruses.

Based on: Margherita Marchetti, Douwe Kamsma, Renko de Vries, Paul van der Schoot, Wouter H. Roos & Gijs J.L. Wuite; Real-time investigation of the assembly dynamics of an artificial virus-like particle; manuscript in preparation
&
Mariska G.M. van Rosmalen, Douwe Kamsma, Andreas S. Biebricher, Chenlei Li, Adam Zlotnick, Wouter H. Roos and Gijs J.L. Wuite; Real-time assembly of SV40 VP1 capsids onto DNA studied using Optical Tweezers and Acoustic Force Spectroscopy; manuscript in preparation
5.1 Introduction

Viruses are small infectious agents that can infiltrate all classes of organisms to replicate themselves inside the host cell\textsuperscript{61}. They consist of a piece of genomic material, either DNA or RNA, an outer shell of proteins called ‘capsid’ and in some cases also a lipid layer derived from the host cell\textsuperscript{62}. The capsid keeps the virus together and can have many shapes, including helical, icosahedral, prolate and other more complex shapes.

Besides the implications of viruses in the transmission of diseases\textsuperscript{63,64}, they have also been studied extensively over the past decade because of their ability to package, transport and deliver drugs at specific locations in the human body\textsuperscript{65–67}. If manipulated in the right way, viruses seem to be ideal for this task, because they prevent drug degradation, minimize interactions with other tissues and enhance drug uptake at the desired tissues\textsuperscript{68}. Despite major accomplishments in this field\textsuperscript{65,69,70}, many questions remain about the nature and function of viruses. One of the outstanding and complicated questions is how the virus can exactly assemble in a well-defined shape.

There is quite some variation in the way viruses are assembled. In some cases, a full capsid is formed first and at a later stage the genome is pumped into the capsid via a portal protein\textsuperscript{71,72}, while other viruses assemble around the genomic material\textsuperscript{73,74}. In the latter scenario, it has been proposed that capsid proteins have fast protein binding rate to the genome and therefore genomic material is used as a scaffold to form a virus capsid. First, it is believed that the proteins are randomly bound to the genome after which they rearrange in a highly cooperative fashion to form a capsid around the genome. An animation from a coarse-grained molecular dynamics model of this process is made by Perlmutter et al\textsuperscript{73}. This movie beautifully and intuitively captures the formation of a capsid around the genome. So far, experimental data is mainly obtained from structure analyses with Electron Microscopy (EM), electrophoresis\textsuperscript{75} and Atomic Force Microscopy (AFM), capturing only the end product of the reassembly process\textsuperscript{76}. This does not provide any information about the actual assembly process itself.

We aim to obtain more real-time information on this assembly process using different single-molecule techniques such as OT, AFM and AFS. With OT, biomechanical properties of the DNA with bound virus proteins can be probed (\textbf{Figure 5.1a, Supplementary figure 5.1}). Furthermore, if the protein of interest is labeled with a fluorescent dye, the binding of proteins and diffusion on the DNA can be visualised (see \textbf{Figure 5.1d}). However, full capsid
Figure 5.1 | Principles of OT, AFM and AFS
(a) Illustrative image of a dual trap OT combined with confocal fluorescence microscopy. Two focused lasers (red) trap two microspheres, that are chemically attached to a DNA molecule. Proteins in solution bind to the DNA and their fluorescence tag lights up when the scanning laser illuminates them. Fluorescence image is displayed in d1, showing the illuminated proteins and microspheres. Accumulating 1D scans of the DNA (x-axes) over time (y-axes) results in a kymograph (d2). Here the binding and diffusion of 50 nM CS0B protein is displayed. (b) Illustrative image of the AFM setup. The tip of the cantilever is scanned over the surface where VLPs collapsed. By measuring the reflection of a laser off the cantilever the applied force can be probed. A rod shaped VLP of about 200 nm in length, measured with AFM in air and displayed in e. (c) Illustrative image of the AFS setup. DNA tethered microspheres are pushed towards the lower pressure gradient. The VP1 proteins bind and compact the DNA. When higher forces are applied, the DNA is extended again (f).

formation is difficult to observe, because of the low degree of force control at low forces and the limited long-time stability. Furthermore, to measure on a single capsid, short DNA is often needed and it is challenging to measure a fluorescence signal on short DNA held with OT. With AFM, the shape and stiffness of the capsid can be probed (Figure 5.1b, Supplementary figure 5.2), however, since the construct has to be fixed on the surface, no dynamic information can be obtained and surface effects cannot be excluded. To follow the assembly process over time, AFS is the instrument of choice, because it is an intrinsic force clamp, allows accurate application of low forces, and it is stable over long times (see Chapter 2). Moreover, combining observations with OT, AFM and AFS might reveal a complete picture of the virus assembly process.

In this chapter, we will investigate capsid formation onto the DNA of a natural virus and of an artificial virus-like particle. The first virus is a natural virus, called Simian Vacuolating Virus 40 (SV40). SV40 virus is about 45 nm and is has
a T=7 icosahedral structure. It consists of 72 VP1 pentamers forming a stable construct and in vivo it encapsulate 5.2 kbp of circular DNA. The proteins of the SV40 virus are obtained from intact viruses that are disassembled before and it has been shown that this virus can reassemble in vitro77,75,78. Because this virus can self-assemble in vivo and in vitro, SV40 is a potential candidate for drug delivery. We will also study an artificial Virus Like Particle (VLP), built up from an artificial protein, here referred as CS10B79. The VLP we study, was designed by Armando Hernandez-Garcia et al.79 to self-assemble onto a DNA molecule. The protein consists of three main domains: a part that binds non specifically to the DNA via electrostatic interactions (oligolysine block B = K1280), a silk-like sequence (Sn = GAGAGAGQn, n stands for the number of repeats) that in solution stacks cooperatively into a stiff filamentous structure81,82, and a random ~400-amino-acid long tail that prevent aggregation of the protein (see Supplementary figure 5.3). Using AFM, it has been shown that this protein forms rod shaped particles (Sn, with n < 10) in the presence of DNA79.

5.2 Methods and materials

In this chapter, measurements are performed with the AFS system as described in Chapter 3. We measure on VP1 proteins obtained from the SV40 virus provided by the lab of Adam Zlotnick and disassembled in our lab in single VP1 protein. Also, we measure on CS10B proteins provided by the lab of Renko de Vries. Both proteins form capsids in the presence of the DNA. In this chapter, we used ~2.8 μm pKYB1 DNA tethers, as described in section 4.3.2 and ~1 μm pRSET DNA tethers as described below.

5.2.1 Protocols to obtain SV40

Stable SV40 capsids are expressed in Spodoptera frugiperda (Sf9) cells, purified and provided to our lab. A detailed protocol is given by Mukherjee et al.83. On the day of experiments, SV40 are disassembled by the addition of the 20 mM TrisCl pH 8.9, 5 mM EDTA and 2 mM DTT. After 30 minutes, the sample is imaged using AFM to confirm complete dissociation of the capsids into single VP1 proteins. Afterwards, buffer is exchanged to 50 mM MOPS pH 7.2 and 125 mM NaCl by ultracentrifugation, using Amicon Ultracentrifugal filters (Millipore, 100k). All reassembly experiments are performed at a VP1 proteins concentration of 360 nM. The total protein concentration is determined using the absorption coefficient of ε280 = 32890 M/cm per VP1 monomer, which based on the length of amino acids chain84.

Truncated VP1 proteins are obtained by storing the VP1 protein at 4 degrees
Celcius for more than 3 days. Traces of protease present in the buffer cleaves
the C-terminal arm, reducing the length of the protein from 41 to 37 kDa,
determined by electrophoresis. The measurement buffer concentration used
for the truncated VP1 is 10 mM MOPS pH 7.2 and 50 mM NaCl, the rest of the
protocol is the same as the intact VP1 protein.

### 5.2.2 Protocols to obtain VLPs

VLPs are provided to us as dry protein polymer filaments (CASQ10K12). In our
lab we dissolve 0.5 mg/mL of the filaments in demi water, followed by 10
minutes incubation at 65°C. Next, the sample is diluted to concentration of 10
mM (MCASQ10K12 = 44.7492 kDa) with phosphate buffer (10 mM, pH 7.4). Finally,
0.1 mM dithiothreitol is added.

The artificial protein is labeled with Alexa Fluor™ 555 C2 Maleimide (Thermo
Fisher Scientific) that binds to the only present cysteine located at one
extremity of the random coil block C. The ‘Thiol-Reactive Probes’ protocol
provided by the supplier is followed. A Pierce BCA protein assay (Thermo
Fisher Scientific kit) in combination with Nanodrop was then used for
determining the protein concentration and extrapolate the dye/protein ratio.

### 5.2.3 Preparation of the DNA tethers

pKYB1 DNA tethers are prepared in a flow cell as described in section 4.3.4.
For VLPs, also a smaller (~1 μm) pRSET DNA construct is used, as described
below (section 5.2.4). The pRSET construct is pre-incubated with the
microspheres. ~50 pg of pRSET DNA was mixed and incubated for 10 min with
10 μL of 4.5 μm streptavidin coated polystyrene microspheres. Then, 2 μL of
digoxigenin labeled microspheres (3.05 μm; the protocol can be found in
section 4.3.3) is added. The microspheres are cleaned by adding 1 mL of PBS
(138 mM NaCl, 2.7 mM KCl and 10 mM phosphate (pH 7.4)) supplemented by
EDTA (5 mM), casein (0.02% w/v), pluronics F127 (0.02% w/v) and Na-azide
(0.6 mg/mL), the buffer we will refer to as measuring buffer. Next, the sample
is spun down (2000 g for 2 minutes) and afterward the residue is removed.
This cleaning step, of suspending of the pellet, is repeated three times and to
the final product is 20 μL of measuring buffer is added. This solution is flushed
in an AFS flow cell coated with anti-digoxigenin and passivated with BSA and
pluronic (section 4.3.4, step 3). After 30 minutes of incubation, the free
microspheres are flushed out with measuring buffer and the construct is ready
for the measurement.
5.2.4 pRSET DNA

The smaller pRSET-A plasmid (Thermo Fisher Scientific, V35120) is used to make the pRSET DNA construct (2909bp, ~1 μm). It is fabricated in the same way as pKYB1 DNA, as explained in section 4.3.2. Besides the different plasmid, the only other difference is that the last purification step (step 9, section 4.3.2) is not performed. About 1 μg (7 μL of a concentration of 149ng/μL) of plasmid is used in step 1 (section 4.3.2). pRSET DNA also has KpnI and EcoRI cutting sites. Following the rest of the steps in the same way, a volume of 50 μL with a final concentration of 6 ng/μL is obtained.

5.2.5 Optical tweezers

In this chapter, a single-trap OT is used to measure biomechanical properties of VP1 proteins on DNA and a dual trap OT combined with confocal fluorescence microscopy is used to follow the binding kinetic rates of the VLPs on DNA (see Figure 5.1d).

For the single-trap OT, pKYB1 DNA tethers are prepared as described in section 4.3.4, except for the use of 1.84 μm microspheres instead of 4.5 μm. All measurements are performed on 100 nM truncated VP1 pentamers in 10 mM MOPS pH 7.2 and 50 mM NaCl buffer at room temperature. DNA tethers are stretched along the surface to obtain force-extension curves of the DNA incubated with VP1 proteins. The advantage of this single-trap surface stretching protocol is that multiple constructs can be incubated for a longer time without applying force. Measurements can last up to 130 minutes, in which constructs are put under tension one at the time.

To follow the binding kinetics of the VLPs on the DNA, dual-trap OT combined with confocal fluorescence microscopy is used, as described previously41. The two ends of a 16 μm long lambda–phage DNA are bound to two polystyrene microspheres and using confocal fluorescence microscopy the fluorescently labeled VLPs were imaged and followed along the DNA in real time (see Figure 5.1a&d).

5.3 Results

5.3.1 VP1 proteins interacting with DNA tethers

Before studying the full capsid formation of VP1 on the DNA, we measured on truncated VP1. These truncated proteins miss their carboxyl-terminal arms, responsible for the strong intrapentamer interactions (e.g. disulfide bonds
Figure 5.2 | Measuring the interaction of truncated VP1 on the tethered DNA
(a) A graph of a typical AFS measurement of the virus capsids self-assembly process onto DNA. Here the end-to-end distance and the applied force are plotted over time. The blue area indicates the flushing of truncated VP1 proteins and colored arrows indicate the stretching curves. (b) Stretching curves taken at different times during the measurement, colored correspond to the arrows in a. (c) Average binding rates obtained by fitting the slope of the end-to-end distance over time, for different force clamps. A total of 8 traces were analysed, data is binned in 0.25 pN steps and error bars represent the standard error of the mean.

and calcium bridges\(^{86,87}\), and are therefore not able to form complete capsids. This is also confirmed with AFM measurements performed in our own lab (Supplementary figure 5.2). Therefore we only observe how the VP1 protein binds to the DNA and changes the mechanical properties.

A typical AFS measurement of virus capsid formation on a DNA tether is shown in Figure 5.2a. In the first stage of the experiment, the anchor point of the DNA was determined, the force was calibrated for each tether in the field of view (typically 50 tethers), and a force-extensions curve was obtained (see section 4.3.5). From these calibration steps single tethers could be discriminated (see
section 4.3.6). Then we flushed in ~100 μL of protein buffer (see section 5.2.1), while keeping the tethers under tensions (~5 pN), preventing compaction and non-specific binding to the surface. Next, we stopped the flow and started reducing the acoustic force in a stepwise fashion, until DNA-protein interaction was observed. Also, at the end of each force clamp a force-extension curve was obtained. When the force was clamped below 0.8 pN, we saw that end-to-end length of the DNA slowly decreased, suggesting that the truncated VP1 interacted with the DNA. For the DNA with decreased contour length, we also observed an earlier onset of force during the forward stretching curve, which we did not see in the relaxation curve (see Figure 5.2b). This means that the DNA is being compacted by the VP1 proteins. When applying a higher load on the construct (~20 pN), the compacting of the DNA was eliminated and the original the end-to-end length of the DNA was retrieved (Figure 5.2b and Supplementary figure 5.1a). This suggests that, while applying higher force, some weak interaction between the truncated VP1 and the DNA are broken. Notably, it appears that the proteins stayed attached to the DNA, because the protein-DNA interactions were recovered directly when the force was lowered again. Data measured with the OT tweezers confirmed these results (see Supplementary figure 5.1).

Due to the early rise in tension during DNA stretching, it is not possible to fit the WLC model here. The relaxation curves on the other hand do exhibit WLC behavior and therefore we used these curves to obtain the changes in mechanical response, caused by VP1 protein binding, by fitting a WLC. Under our buffer conditions (10mM MOPS pH7.2 and 50mM NaCl) we obtained a $L_p$ for bare DNA of 51 ± 6 nm (n = 19). For DNA incubated with truncated VP1 pentamers we observe a drastic decrease of the $L_p$ from 48 ± 3 to 8 ± 1 nm (n = 23). To quantify the speed of DNA compaction by these weak protein-DNA interactions, we determined the shortening compared to the expected contour length for a range of applied tensions (see Figure 5.2c). From these measurement we observed that above ~1 pN the truncated VP1 protein was not able to compact the DNA, while at lower tensions the compaction rate increases with a decrease in tension.

Next, we measured on ‘intact’ VP1 pentamers. These VP1 proteins have their carboxyl-terminal arms and can therefore form strong protein-protein interaction. Here the same sequence of steps is applied as above (see Figure 5.2a), however, after flushing in the protein, the tension on the DNA is set to <0.5 pN and force-extension curves are measured after different periods of incubation for individual experiments (between 5 and 180 minutes).
Figure 5.3 | Quantifying the force-extension curves of DNA incubated with ‘intact’ VP1 proteins
(a) 6 representative force-extension curves of DNA with VP1 proteins incubated for different times. Each curve is obtained from independent measurements. (b) The contour length and the persistence length are obtained from fitting WLC to the relaxation curves (a). The total number of stretching curves measured are 123, 97, 25, 16, 29 and 13 for bare DNA, <10, 60, 90, 120 and 180 minutes of incubation, respectively. Error bars represent standard errors of the mean and the blue arrow indicates the drastic drop in persistence length after flushing in the VP1 protein. (c) A representative example of the found rupture events using an in-house step-finding algorithm. The red lines are the found steps and the blue lines are the used events. The program was set to find only steps with a fixed slope. The rupture force is taken at the beginning of the rupture event. (d) Force-extension curves are sorted by their ratio of compaction. The ratio is defined as the contour length of the bare DNA divided by the contour length of the back-stretching curve. We divided the compaction ratio in 3 equally spaced bins, containing 60, 16 and 13 stretching curves and a total of 337, 60 and 14 rupture events were found, respectively. Error bars represent the standard error of the mean.

The longer the incubation time, the stronger was effect on the mechanical properties of the DNA due to the binding of VP1 proteins (Figure 5.3a). We observed a direct and drastic effect on the persistence length of the DNA compared to the bare DNA, while the contour length was initially the same,
Figure 5.4 | Measuring the compaction of 1 μm DNA tethers by VLPs
(a) A graph of a typical AFS measurement of VLP assembly onto DNA. Here the end-to-end length of the DNA and the applied force are plotted over time. The coloured arrows indicate the force-extension curves taken and the red striped square is focussing on an area used to obtain compaction steps. (b) Stretching curves taken at different times during the experiment, colors correspond to the arrows in a. (c,d) The contour (c) and persistence (d) lengths are obtained from fitting WLC fit to the relaxation curves, respectively. A total of 20, 20, 15 and 9 curves were fitted for 50 nM and 15, 15, 12 and 8 for 1000 nM, for curves obtained after 0, 10, 45 and 85 minutes of incubation, respectively (error bars are standard error of the mean). (e) A zoom in of the graph in a (red striped box), showing the compaction steps found with our in-house step-finding algorithm, based on previously described theory. The program was set to only find horizontal steps.
Figure 5.5 | Compaction and decompaction events found by our step-finding algorithm
We measured on 3 (a) and 1 μm (b,c) DNA tethers to quantify the step size of compaction and decompaction events, obtained with our step-finding algorithm (see Figure 5.4e), using a concentration of 1000 (a,b) and 50 nM (c). The histograms are fitted with a multi-Gaussian function. Here the distances from peaks-to-peak are equally spaced and used as one fit parameter.

Fit values obtained are 21 ± 2, 21.3 ± 0.4 and 22.6 ± 0.3 (fit value ± s.e.m.) nm for the peak-to-peak distance for histogram a, b and c, respectively.

but started to decrease after ~100 minutes (see Figure 5.3b). This suggests that the binding of the proteins happened already before the first stretching curve, while stronger protein-protein bindings took more time.

In addition, we observed rupture events in the forward-stretching curve. Because these ruptures were not detected with the truncated VP1 protein, these ruptures correspond to the stronger protein-protein bindings. We use a step-finding algorithm to determine rupture lengths and forces (see Figure 5.3c). From the contour length, obtained by the WLC fit of the relaxation curve, we determined the amount of DNA compacted by the virus capsid. The forces at which these ruptures occurred increased with the amount that had been...
compacted (Figure 5.3d). The sizes of the steps of these rupture events initially have a wide distribution (250 nm standard deviation) but after the amount of compacted DNA has increased, only small ruptures events are observed (60 ± 22 nm). This observation could be explained by the idea that, when a full capsid is formed, there is increasingly less free DNA available that can contribute to these small rupture events.

5.3.2 CS10B proteins interacting with DNA tethers

With the AFS we perform similar kinds of experiments on the artificial virus protein CS10B as explained above the VP1 virus capsid proteins. After flushing the CS10B protein into the flow cell, we observed compaction of the DNA, while clamping the force at ~1.5 pN (see Figure 5.4a). Force-extension curves were taken sequentially to quantify the bio-mechanical changes caused by the CS10B protein (see Figure 5.4b). In the relaxation curves we initially observed an abrupt and drastic change in the persistence length after 10 minutes (from ~43 ± 1 to ~15 ± 3 nm), while after longer incubation times the persistence still decreases steadily, but only little (see Figure 5.4c). The contour length, on the other hand, decreased more gradually with time (~5 nm per minute, see Figure 5.4d). We conclude from these results that CS10B first binds to the DNA and then starts compacting it.

When applying a force clamp we observed shortening of the DNA in a stepwise fashion (Figure 5.4e). We quantified, the length of the steps using an step-finding algorithm. We obtained step-size distributions for 2.8 and 1 μm long DNA and at 1000 and 50 nM CS10B. The distributions of step sizes were fitted using a multi-Gaussian function with equally spaced peaks, resulting in the same single step size for each condition used here (see Figure 5.5). Remarkably, we observed the same compaction step-size under all conditions. In addition, decompaction steps were observed most noticeably, for 1 μm DNA in combination with 50 nM CS10B. Preliminary data also show that decompaction could be enforced by applying force clamps with increased tension. These observations suggest that the DNA is compacted in a very regular way. Perhaps, the DNA is forming loops that give rise to these specific steps.

5.4 Discussion

In this chapter we showed that AFS is well capable of following the dynamics of the self-assembly of viruses. Because it acts as a force clamp and it is stable over long time, we were able to follow the dynamics of the process. We
measured on both a natural virus (SV40) and a virus like particle consisting of CS50B proteins. The combined observations made by OT, AFM and AFS lead to a deeper understanding of assembly process of these viruses, as we will further clarify below.

5.4.1 VP1 protein form full SV40 capsids

We started measuring on truncated VP1 proteins. They lack their carboxyl-terminal arms and are therefore unable to bind to other VP1 proteins, allowing us to only monitor DNA-protein interaction. After incubating them with DNA tethers, we observed a drastic drop in the persistence length of the DNA. It has been proposed by Kler et al\textsuperscript{89} that the stiffness of the genomic material highly influences the assembly product. It is indeed logical that it is easier to package a long polymer chain (in the order of micrometres) in round small capsid (45 nm) if the polymer chain is more flexible. Additionally, the drastic decrease in persistence length is also observed for the ‘intact’ VP1 proteins (Figure 5.3b). This is measured at the first force-extension curve obtained (after 10 minutes), suggesting that the DNA is saturated quickly with VP1 proteins. These observations support the assumption that the DNA acts as a scaffold for the formation of the virus capsid.

Next, we observed compacting of the DNA for both the truncated as well as the ‘intact’ VP1 proteins (Figure 5.2 and Figure 5.3b). This suggests that VP1 compacts the DNA also without the need for protein-protein interactions. We think that VP1 has multiple binding sites for DNA: initially, only one binding side is connected, while over time multiple connections are formed, thereby compacting the DNA. In OT experiments, we observed steps of 40 nm during the forward stretching (see Supplementary figure 5.1d). We expect that one step is related to one VP1 protein unbinding from the DNA, however still connected with only one binding site. This one protein binding site to the DNA does not induce tension on the DNA, but when then tension on the DNA is lowered the other binding sites of the protein can reform fast reform rapidly, because of the higher local concentration. This hypothesis of multiple binding sides is also supported by the fact that when measuring a second force-extension curve almost the same breaking events are observed (see Supplementary figure 5.1a). This would mean that the proteins are bound to the DNA, forming more connections over time; when more tension is applied these connections are all broken, except for one, and they quickly reform after the release of tension.

The ‘intact’ VP1 proteins are able to form protein-protein interactions. Initially, we observed the same drastic decrease in persistence length (Figure 5.3b) and
compaction of the DNA as with the truncated virus. We observed after ~100 minutes of incubation a decreased contour length, obtained from the relaxation curve (see Figure 5.3b). This was not observed for the truncated VP1 proteins, after the force-extension curve they always went back to the contour length of bare DNA, meaning that they could not form stable structures. AFM data was in agreement with these observations: no capsids were observed for the truncated VP1 proteins and for the ‘intact’ VP1 proteins capsids were only observed after 120 minutes (see Supplementary figure 5.2).

As concluded before, the decrease of contour length was related to the formation of more stable structures, which could be partially or fully formed capsids. Remarkably, this decrease of contour length happens after around 100 minutes of incubation, about the same time needed to observe capsids with AFM. Furthermore, we reasoned that the more the contour length of the DNA tether is decreased the more DNA is compacted and finally fully encapsulated in a full SV40 capsid. In vivo a capsid contains 5.2 kbp of DNA, that is related to a contour length ~1.7 μm of DNA. It has to be noted that several research groups have reported different encapsulated genome sizes in vitro90–92. pKYB1 DNA used here has a length of 2.8 μm, therefore we expected that only one full capsid can be formed on our tether, which would result in a decreased contour length to ~1.1 μm. The contour length of most DNA molecules, incubated for more than 120 minutes, have a peak between 0.5 and 1 μm (see Supplementary figure 5.4). It is very likely that those molecules have formed one complete capsid, while the other DNA molecules, with still longer contour length, have probably formed intermediate structures.

If a full capsid would be formed during our experiments, part of the DNA would still stick out of the capsid and be free to interact with proteins in solution. When less DNA is available, it is expected that only small breaking events would be visible during the force-extension curves of the DNA, as observed with the truncated DNA. These smaller steps, as measured on truncated DNA (40 nm with a standard deviation of 15 nm, see Supplementary figure 5.1d), are very similar to the steps measured on the tethers with decreased contour length (60 nm with a standard deviation of 20 nm, see Figure 5.3d). However, the force of the breaking events is stronger and increases over time, suggesting that they are breaking the stronger protein-protein interactions, however not enough DNA is available to form a full capsid.

We have studied the assembly process of the SV40 virus with AFS and demonstrated that this technique can be useful to extract additional information about the assembly process. Compared to other single-molecule
techniques (OT and AFM), AFS has the advantage of long-time stability, is an intrinsic force clamp and has high data throughput, allowing us to follow the assembly process in real time for many molecules in parallel. Since AFS, OT and AFM are complementary to each other, the combined observations results in a deeper understanding of the reassembly process of SV40 virus in vitro. This knowledge can be used for therapeutic purposes, since this virus is potentially an useful candidate for targeted drug delivery.

5.4.2 CS10B protein forming VLPs onto DNA tethers
CS10B protein is an artificially designed protein, which can form stiff rod shape particles in the presence of DNA, as shown by AFM images, compacting it to one third of its original length. The amount of nucleation points increases with DNA length, with typically one nucleation point every 1 μm of DNA. Furthermore, from dual-trap OT experiments combined with confocal fluorescence microscopy performed in our lab, we know that the protein binds as monomer or oligomers to the DNA (see Supplementary figure 5.5b), it can diffuse over the DNA if the cluster is small (<5 proteins) and under these experimental conditions no unbinding of the protein is observed.

From our AFS measurement, we could confirm that the CS10B protein binds rapidly to the DNA, because of the drastic change in persistence length after only ~10 min of incubation (Figure 5.4d). Furthermore, under low tension the protein starts to compact the DNA in a specific stepwise fashion of ~22 nm (DNA end-to-end length). This corresponds to 30 nm of contour length, calculated using the WLC model (equation 6.16), plugging in the force applied (~1 pN) and the measured persistence length of 13 nm. We measured these step sizes with DNA molecules of two different lengths. For 2.8 μm long DNA, the step-sizes distribution was smeared out (see Figure 5.5a), probably because VLP nucleates at multiple locations along the DNA. The shorter DNA construct (~1 μm) seems to form one nucleation site per tether, resulting in well-defined steps. For shorter DNA and lower concentrations of proteins, also decompaction was observed. Notably, the decompaction steps had the same size, suggesting that VLP compacts and decompacts in the same manner.

To better understand how the protein forms the rod like shape on the DNA, we calculated how much DNA is compacted at each step. Given that the tension of the DNA is ~1 pN, the most frequent steps correspond to 30 nm in contour length. From AFM data, we know that the compaction factor is ~3, meaning that the 1 μm DNA compacts in a 334 nm rod. Therefore, the total length decreases with 667 nm and the total amount of compaction steps are 667/30 nm = ~22 steps. Our DNA tether has 2909 bp, indicating that every
step \( \frac{2909}{22} = \sim 130 \) base pairs are compacted.

All these findings combined lead us to propose the following working mechanism for the assembly mechanism of the CS\(_{10}\)B protein into a VLP (see Supplementary figure 5.3). In solution, aggregation is prevented by the \( \sim 400 \)-amino-acid coil, however we observe that there are still aggregates for our measurement conditions (Supplementary figure 5.5). In the presence of DNA, proteins bind to the DNA via electrostatic interactions (oligolysine block B). DNA functions here as a scaffold to increase the local concentration of proteins. The DNA persistence length decreases upon proteins binding, making it easier to bend and therefore to compact. Single proteins or aggregates smaller than 5 proteins can diffuse over the DNA and when in close proximity to each other, they can merge. Larger and stably bound oligomers might represent the nuclei from which the particle growth can take place. From the nucleation site, the DNA is compacted every time new proteins stack on it. This happens in a steps of 30 nm of contour length. The exact mechanism of how the DNA is compacted and the final particle structure still remains an open question. Our data show a regularity in the compaction process that might be related to the formation of loops. The observed decrease in persistence length is the key change needed to overcome the energy barrier for bending the DNA in such a way. It could be that DNA loops are formed around the protein or any other kind of regular structure by the DNA hold in place by the CS\(_{10}\)B protein. One way to further elucidate the process is to measure on torsional constrained DNA, to check for the presence of DNA loops. In addition, structural analyses or molecular dynamic simulations could tell us more about the last piece of the puzzle.

**Acknowledgements**

We thank M. van Rosmalen, M. Marchetti for all the help, work, scientific discussions and all the OT and AFM data shown respectively on the SV40 virus and VLP. We thank A.S. Biebricher and S. Acar for their help and support on sample preparation and OT instrumentation, D. Vorselen for providing the home-built Matlab based step-finding algorithm. Further, we thank the laboratory of prof. A. Zlotnick (Molecular & Cellular Biochemistry, Indiana University, Bloomington) for kindly providing us with the SV40 viruses and dr. Renko de Vries (Physical chemistry and soft matter science, Wageningen University) for kindly providing us with the VLP.
5.5 Supplementary figures

Supplementary figure 5.1 | DNA binding properties of truncated VP1 pentamers measured with OT

Data is obtained on DNA with 100 nM truncated VP1 proteins in 10 mM MOPS pH7.2 and 50 mM NaCl at room temperature. (a) Stretching curves of bare DNA (black) and DNA incubated with truncated VP1 proteins for 9 min (green). WLC Odijk fit of the bare DNA and back-stretching curve of DNA+VP1 is indicated with a dotted line. (b) Scatter plot of all rupture forces obtained from 20 DNA+VP1 molecules after different incubation times. (c) Using the WLC Odijk model we were able to obtain the contour length corresponding to each force point of our stretching curves in time. Then we applied a step-finding algorithm to determine the step wise lengthening of the contour length (inset). This correlates to the ruptures in our stretching curves. (d) Histogram of the lengths of the different steps fitted with a Gaussian function (red line), this resulted in an average rupture step size of 40 ± 20 nm (Gaussian fit ± s.e.m.).
Supplementary figure 5.2 | AFM data of reassembly products of VP1 pentamers incubated with 16 μm lambda DNA

Topographical 2D and 3D image together with a height profile along the green arrow. (a) ‘Intact’ VP1 pentamers results after 120 min of incubation in the formation of multiple capsids on a lambda DNA molecules in a ‘beads-on-a-string’ like formation. (b) When using truncated VP1 pentamers, large unorganized clusters of DNA-VP1 proteins are obtained with a maximum height of 30 nm. Detailed analyses of this study can be found in the thesis of Mariska van Rosmalen.

Supplementary figure 5.3 | Artistic impression of the VLP

(a) Design of the protein that form the VLP. From left to right, the virus is build up from an oligolysine block b= K12(b, ), a ten times repeat of a silk-like sequence Sn = GAGAGAGQ10(S, red) and is a random 407-amino-acid long tail (C, green). (b) Assembly process of the VLP described in 5 steps. 1 DNA tethers are under tension in solution and the protein starts to bind. 2 The protein diffuses over the DNA and encounters other proteins. 3 The proteins compacting the DNA starting from an aggregation site. 4 The compete DNA is compacted, reducing the length with a factor of ~3.5 the amino acid tail is visualized here to give the complete picture.
**Supplementary figure 5.4 | Histogram of the contour length of the DNA incubated with VP1**
Contour length determined from the back-stretching curves for DNA incubated for more than 120 minutes. A total of 32 independent DNA molecules are displayed here, bin size is 0.5 μm.

**Supplementary figure 5.5 | Fluorescence data obtained with dual trap optical tweezers**
16 μm lambda DNA is attached between two microspheres as described previously[19], while incubated with 2 different concentrations (50 and 200 nM, light and dark green, respectively) of the CSαβ protein. (a) Two kymographs show the real-time binding and diffusion of the CSαβ protein onto the DNA. The tension is kept constant (5 pN). (b) Histograms of the number of peptides binding at the same time to the DNA at a scan line time of 30 ms. First peak is fitted with a Gaussian function, resulting in 3.0 ± 0.4 and 6.0 ± 0.3 binding events, for 50 and 200 nM, respectively. (c) Cumulative distribution function of two representative diffusion paths analysed with an in-house MatLab Kymotrackker software, used previously[93]. Exponential fit values obtained are 4.1 ± 0.2 and 2.3 ± 0.1 minutes (fit values ± s.e.m.), for 50 and 200 nM, respectively. (d) Diffusion coefficient obtained from the CDF, calculated for different cluster sizes. Red zone indicates zero diffusion.