Modeling project: DNA packaging in a virus

Abstract

In order to infect a host cell, bacteriophage viruses must eject their genome, which is then replicated, followed by the synthesis of the proteins encoded in it. Once the concentration of the proteins is high enough, the construction of the viral capsids, which are rigid protein boxes, is initiated. The next step is the packaging of the viral DNA inside the capsids, which can then be injected into other host cells. The packaging of DNA in these viruses is thus an essential step in their life cycle. In this article, we describe a simple physical model explaining this packaging process for diversely shaped capsids.

1 Introduction

The dynamics of the packing of double-stranded DNA in bacteriophages and other viruses using the same mechanism (ATP-dependent packaging of the DNA into a preformed capsid) has been investigated both experimentally and theoretically. There are two main types of theoretical models: thermodynamic quasi-static models which considers DNA as a polymer (see [Pur+05]) and molecular dynamics models which simulate it as a series of beads connected by springs (see [PH08]). These two approaches lead to different predictions and there is still a debate about the nature of the dynamics and the forces involved in the mechanism.[Ber+14] The physical model we present is of the former type. It assumes that DNA takes a single conformation inside the capsid and organizes itself as a spool. However, the first experimental measurement of the spatial organization of DNA inside the capsid during the packing process has only been observed in 2008 by Comoli and al. [Com+08]. They showed that contrarily to what was believed before (for instance in [Pur+05; PKP03]), DNA tends to organize itself in the spool structure very late in the packaging process. Indeed, until 78% of the genome is packed, the DNA inside the capsid stays in a disordered phase. Thus, the dynamics of DNA packing inside viruses is still an active subject of research.

Name of the variable	Symbol	Value	Source
Diameter of the capsid	R_{out}	$21 \times 54\ nm$	Phage $\phi 29$ [Tao+98]
Diameter of the DNA	d_{dna}	$1\ nm$	[Pur+05]
Length of the DNA strand	L_{tot}	$5.5.10^3\ nm$	Phage $\phi 29$ [Tao+98]
Persistence length	ξ_p	$50\ nm$	[STS01]
Gyration radius	R_G	$300\ nm$	phage $\phi 29$ [Pur+05]
Motor force	F_{mot}	$10\ pN$	[Pur+05]
Typical pressure	p_{poly}	$219\ atm.$	will be explained later
Pressure of non-release	p_{nr}	$20\ atm.$	will be explained in the text

Table 1: Some order of magnitudes about DNA in a viral capsid

The article is structured as follows: the next section introduces the physical model for the packaging of DNA inside a viral capsid. The following derives the expression for the force resisting the packaging for three different geometries and includes a qualitative analysis of data obtained from the literature. The final section concludes with final remarks about the model's predictions and how it could be improved.

2 Simple physical model of DNA packaging inside a viral capsid

The goal of this section is to present how the worm-like chain model (WLC) of DNA can be built from simple physical arguments and used to construct mathematical expressions that are useful to understand the mechanism of DNA packing inside a viral capsid. Our goal in the end is to compute the force resisting the packaging motor. We make the assumption that only the DNA strands participate to the packing and the capsid is assimilated to a

infinitely rigid vessel of radius R_{out}[1]. Our model contains only a polymer, which is charged because DNA is strongly negative. As DNA is solvated in a solution, the nature of the interactions between the DNA strands depend on the surrounding solvent ions.[PKP03] In particular, we consider purely repulsive conditions.

2.1 Building the Hamiltonian of the worm-like chain model

The only element we want to keep in our model is the stiffness of the polymer, which is represented by a binding energy, and the repulsion between two points of the polymer. We do not include tension in our model because the extremity of the polymer is free and any tension will relax very fast (this is an assumption). We define the curvilinear coordinate along the DNA strand s, the "trajectory"/ "parametrised curve" representing the strand $r(s)$, and κ which is a kernel that represents the energy of interaction between two bits of unit length $s, s + ds$ at two distinct points along the strand. We also have k_B the Boltzman constant, T the temperature [2], and ξ_p the persistence length of the polymer [3].

Then, keeping only the lowest order terms in the Hamiltonian, we end up with the following form:

$$G = \int_0^L ds \left[\frac{k_B T \xi_p}{2} \left(\partial_s^2 \vec{r}(s) \right)^2 + \int_0^L ds' \kappa \left(\vec{r}(s), \vec{r}(s') \right) \right]$$

We see that the term in $(\vec{r}(s))^2$ will correspond to bending. The other terms can allow to implement any kind of two point interaction of the DNA strand.

2.2 Building the two terms G_{bend} and G_{int}

The goal of this subsection is to write the two terms we have identified in the Hamiltonian in a more explicit form.

2.2.1 Expression of the bending force

To put the bending force in a nice expression, we start from the empirical observation of DNA packed inside the capsid. The DNA is packed as a spool. So we will assume that it is always so. This leads to the assumption that the radius of bending is almost constant: $\left(\partial_s^2 \vec{r} \right)^2 = \frac{1}{R^2}$. We transform the integral of the curvilinear coordinate into an integral over the radial coordinate of the piece of strand in the spool:

$$\int_0^L ds = \int_{R_{int}}^{R_{out}} \rho(R) dR \int_0^{2\pi} R d\theta$$

Where R_{int} is the internal radius of the spool, R_{out} is the outer radius of the spool (which is the same as the radius of the capsid), and ρ is the density of spools of radius R. This density gives the number of spools with radius in between R and $R + dR$. If we denote as $N(R)$ the number of circles of DNA that can be at radius R, we have $\rho(R) = \frac{N(R)}{\frac{\sqrt{3}}{2} d_s}$ because in the horizontal direction, there is one layer every $\frac{\sqrt{3} d_s}{2}$. In the vertical direction, one can relate $N(R)$ to the shape of the capsid. We define $h(R)$ as the height of the capsid at radius R. We thus have $N(R) = \frac{h(R)}{d_s}$ because in the vertical direction, there is one strand every d_s (see figure 1).

---

[1]We only consider capsids that are invariant under rotation and are cylindrical, cylindrical or caped-cylindrical (see figure 3

[2]T=300 Kelvin for instance

[3]This typically means that if we define as $\vec{t}$ the tangent vector of the polymer on s and $\vec{t'}$ the tangent vector of the polymer at s' we typically have $\langle \vec{t}.\vec{t'} \rangle \propto \frac{k_B T}{2} \exp \left(\frac{|s - s'|}{\xi_p} \right)$ where $\langle \dots \rangle$ is an average over the configurations of the polymer
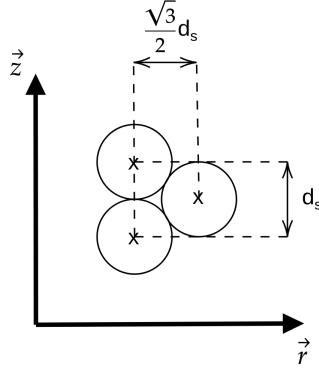
Figure 1: Schematic of the packing of DNA strands.

Putting all these expressions together, we obtain bending energy:

$$G_{bend} = \frac{2\pi\xi_p k_B T}{\sqrt{3}d_s^2} \int_{R_{int}}^{R_{out}} \frac{h(R)}{R} dR \tag{1}$$

We see that the bending term takes into account the rigidity of the polymer and the shape of the capsid. The latter will be encoded in the h function. We come back to the computations for different shapes later, for now let us look at the repulsion term.

2.2.2 DNA-DNA repulsion

The other term that must be taken into account in the energy of the system is the repulsion between DNA strands. In fact, each base pair is charged with a charge $+2$ [Alb+02] and DNA is a strongly charged molecule $\sim 7C.nm^{-1}$. Due to this charge, DNA could not be stable if it was not in a solution containing counter ions [SS15]. In this model we will only consider interaction between nearest neighbours. Once again this choice is based on the assumption that the spool is always very well organized.

The force has been determined experimentally based on measurements of osmotic pressure. The force between two unit length DNA strands seperated by a distance d_s is : $g_{el}(d_s) = \frac{F_0 d_s}{\sqrt{3}} exp\left(\frac{-d_s}{c}\right)$, where F_0 is homogeneous to a force per unit length and c to a length. They are parameters that should be determined from fits on experimental data. The form of this expression comes from the fact that the ions in the solution tend to screen the charge of the DNA strands. That is why we do not have a behaviour in $\frac{1}{d_s^2}$ as one might expect for coulomb repulsion. No direct measurement of F_0 and c have been done to our knowledge [PKP03]. g_{el} represent the potential energy of interaction of two DNA strands of unit length at a distance d_s from one to another. To compute the potential energy of interaction of two unit length of strand we have to compute the energy we bring two strands that are separated by an infinity distance, to d_s between them. By doing so we compute the energy of one link between two DNA strands.

$$
\begin{aligned}
G_{el}^{link} &= \int_{\infty}^{d_s} f_{el}(d_s') \, dd_s' \\
&= \int_{\infty}^{d_s} \frac{F_0 d_s'}{\sqrt{3}} exp\left(\frac{-d_s'}{c}\right) dd_s' \\
&= \frac{F_0}{\sqrt{3}} \left(c^2 + cd_s\right) exp\left(\frac{-d_s}{c}\right) \quad \text{(we made an intergration by part)}
\end{aligned}
$$

As each strand has six neighbours and as the link is shared between to strands (see figure 2), to compute the contribution by unit of length on one strand we have to put a factor 3. Then we have to multiply this energy by unit length along full length of the DNA strand:

$$G_{charge} = 3LG_{el}(ds) = L\sqrt{3}F_0\left(c^2 + cd_s\right) exp\left(\frac{-d_s}{c}\right) \tag{2}$$
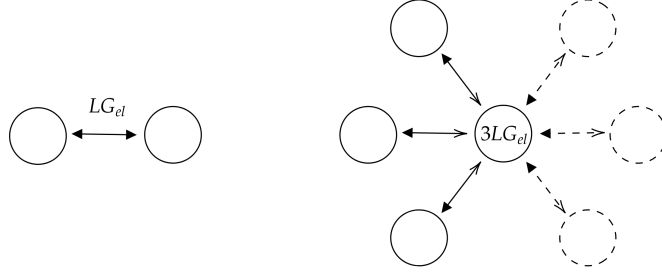
3

Figure 2: Repulsion between DNA strand.

The expression might seem very simple. In fact it also contain a dependence in the geometry of the capsid. This dependence is hidden in L. The length of DNA in the capsid reads (see the appendix equation 37):

$$L = \frac{4\pi}{\sqrt{3}d_s^2} \int_{R_{int}}^{R_{out}} Rh(R)dR$$

We have derived the explicit expression for each term in the hamiltonian. Now what we want to do is put everything together and minimize the hamiltonian. This is done in the next section.

2.3 The equation on d_s

In this section we are going to minimize analytically the expression of G. This approach rely on the fact that at mechanical equilibrium the DNA strands are just lying in the configuration such that the energy is minimised. As the length of DNA inside the capsid is fixed, the only parameter that can be optimised is d_s. Let us write the full expression of G:

$$G_{tot} = \frac{2\pi\xi_p k_B T}{\sqrt{3}d_s^2} \int_{R_{int}}^{R_{out}} dR \frac{h(R)}{R} + L\sqrt{3}F_0 \left(c^2 + cd_s\right) \exp\left(\frac{d_s}{c}\right) \tag{3}$$

Writing the condition $\frac{dG_{tot}}{dd_s} = 0$ leads to the equation:

$$\sqrt{3}F_0 \exp\left(\frac{-d_s}{c}\right) = \frac{\xi_p k_B T}{d_s^2 R_{int}^2} - \frac{\xi_p k_B T}{2d_s^2} \frac{\int_{R_{int}}^{R_{out}} \frac{h(R)}{R} dR}{\int_{R_{int}}^{R_{out}} Rh(R)dR} \tag{4}$$

We see that this is a closed hierarchy and that it can in principle be solved. Analytical solving does not seem to be at hand. So for this we need to rely on numerical solving of this equation.

2.4 The force during packing

Now let us imagine that we have found the value of d_s^* that minimises the free energy. We can reintroduce it in G and deduce the motor force via $F = \frac{dG}{dL}$. We deduce from equation (3) that (see appendix 4 for detailed derivation):

$$F = \frac{\xi_p k_B T}{2R_{int}^2} + F_0\sqrt{3}\left(c^2 + cd_s^*\right) \exp\left(\frac{-d_s^*}{c}\right) \tag{5}$$

This is exactly the expression that have been given in [PKP03] (see equation [19] of the reference). As a conclusion, we have seen how to derive the expression for the force of the motor. Now we want to find a more explicit form for this force. We would like to express the force as a function of the percentage of DNA packed inside the capsid. We will see in the next section that for that we need to specialise to a given geometry of capsid.

3 Computing of the force in various geometries

In this section we will show that for typical geometry of capsid, we can express the force as a fucntion of the percentage of DNA packed inside the capsid, the radius R_{out} and the rate of compaction $\rho_{pack} = \frac{\Omega_{dna}}{\Omega_{caps}}$ (it the ratio of the volume occupied by the DNA when it is fully packed divided by the volume of the capsid). We remark here that $\Omega_{dna} = \frac{\sqrt{3}d_s^2}{2}L$. It is interesting to use these three parameters corresponds to quantity that are easy to access experimentally [Pur+05; PKP03]. We will dsitinguish three geometry of capsid : (1) *cylindrical*, when the capsid is a cylindre, (2) *spherical*, when the capsid is sphere, (3) *spherical-capped*, when the capsid is a cylinder capped with two hemispheres (see figure 3).
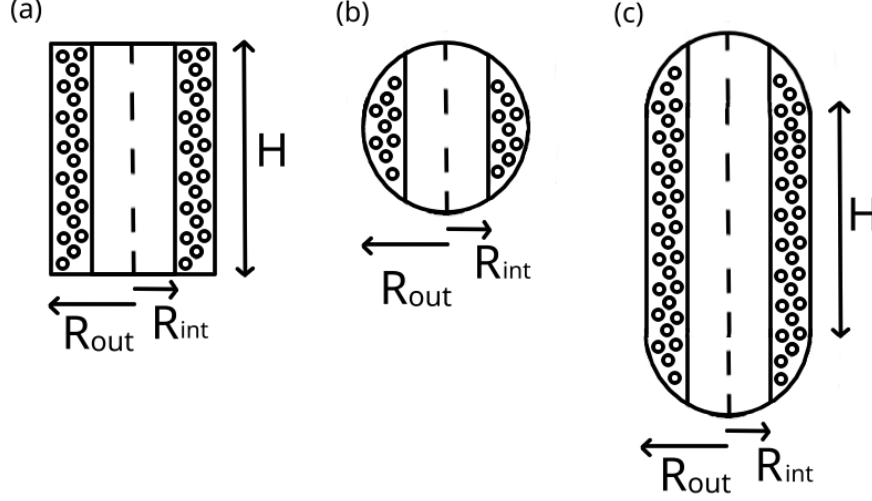

Figure 3: Geometries that will be studied analytically: (a) is a simple cylindrical capsid, (b) is a spherical capsid and (c) is a capped cylindrical capsid.

To compute the force we will need to express the following quantities (see the equations below) as a function of d_s, ρ_{pack}, p which is the percentage of DNA that have been packed. These quantity can be expressed as a function of R_{int} but it hides the dependency in d_s. We will repress this quantidy in terms of the variables d_s, ρ_{pack}, p.

$$\int_{R_{int}}^{R_{out}} 2\pi h(R) R dR \quad = \quad -\frac{d_s^2 \sqrt{3}}{2} L \tag{6}$$

$$\int_{R_{int}}^{R_{out}} dR \frac{h(R)}{R} \tag{7}$$

$$\rho_{pack} \quad = \quad \frac{\Omega_{dna}}{\Omega_{caps}} \tag{8}$$

$$A \quad = \quad \frac{\xi_p k_B T}{R_{int}^2} - \xi_p k_B T \frac{\int_{R_{int}}^{R_{out}} \frac{h(R)}{R} dR}{\int_{R_{int}}^{R_{out}} R h(R) dR} \tag{9}$$

$$\text{Which is useful to compute the right hand side of (4)} \tag{10}$$

$$G_{bend} \tag{11}$$
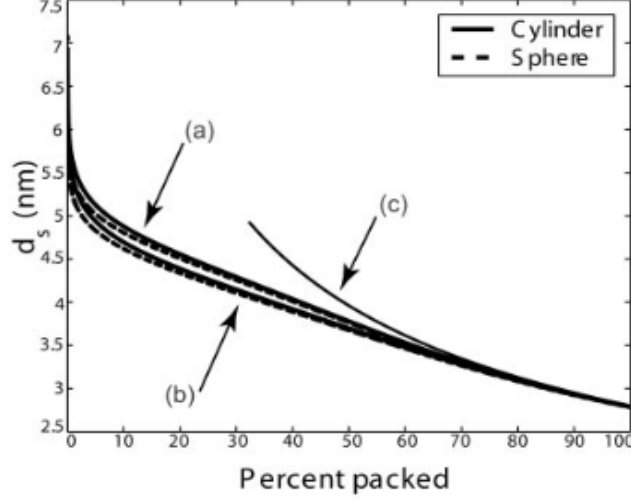
$$G_{charge} \tag{12}$$

Figure 4: Spacing between DNA strands as a function of the percent of genome packed for cylindrical and spherical capsid geometries. We assume repulsive solvent conditions with $F0 = 255,000 \; pN/nm^2$ (a) or $F0 = 4.1 \times 55,000 \; pN/nm^2$ (b). The capsid dimensions are chosen so their volumes coincide with the ϕ 29 virus. Curve (c) is the spacing between strands assuming a uniform packing of the capsid. Figure and legend taken from: [PKP03].

3.1 Cylindrical capsid

For the cylindrical capsid, the starting point is $h(R) = h$. We deduce:

$$\int_{R_{int}}^{R_{out}} h2\pi R dr = \frac{d_s^2 \sqrt{3}}{2} L \tag{13}$$

$$R_{int} = R_{out}\sqrt{1 - \frac{\sqrt{3}d_s^2 L}{2\pi h R_{out}^2}} \tag{14}$$

$$\rho_{pack} = \frac{d_s^2 \sqrt{3} L}{2} \frac{1}{\pi h R_{out}^2} \tag{15}$$

$$\int_{R_{int}}^{R_{out}} dR \frac{h(R)}{R} = -\frac{h}{2} \ln\left(1 - \rho_{pack}\right) \tag{16}$$

$$A = \frac{\xi_p k_B T}{R_{out}^2}\left[\frac{1}{(1 - \rho_{pack}p)} - \frac{\ln\left(1 - \rho_{pack}p\right)}{\rho_{pack}p}\right] \tag{17}$$

$$G_{bend} = \frac{2\pi k_B T h}{\sqrt{3}d_s^2} \ln\left(\frac{R_{out}}{R_{int}}\right) \tag{18}$$

$$G_{charge} = L\sqrt{3}\left(c^2 + cd_s\right)\exp\left(\frac{-d_s}{c}\right) \tag{19}$$

Now to compute the force during the packing, we need to varie p from 0 to 1. For each value of p we compute d_s (20). We can deduce the expression of the force via (21).

$$0 = \sqrt{3}F_0 \exp\left(\frac{-d_s}{c}\right) - \frac{\xi_p k_B T}{R_{out}^2 d_s^2}\left[\frac{1}{(1 - \rho_{pack}p)} - \frac{\ln\left(1 - \rho_{pack}p\right)}{\rho_{pack}p}\right] \tag{20}$$

$$F = \frac{2\pi\xi_p k_B T}{2R_{out}^2 \left(1 - \frac{d_s^2 \sqrt{3}L}{2} \frac{1}{\pi h R_{out}^2}\right)} + F_0\sqrt{3}\left(c^2 + cd_s\right)\exp\left(\frac{-d_s}{c}\right) \tag{21}$$
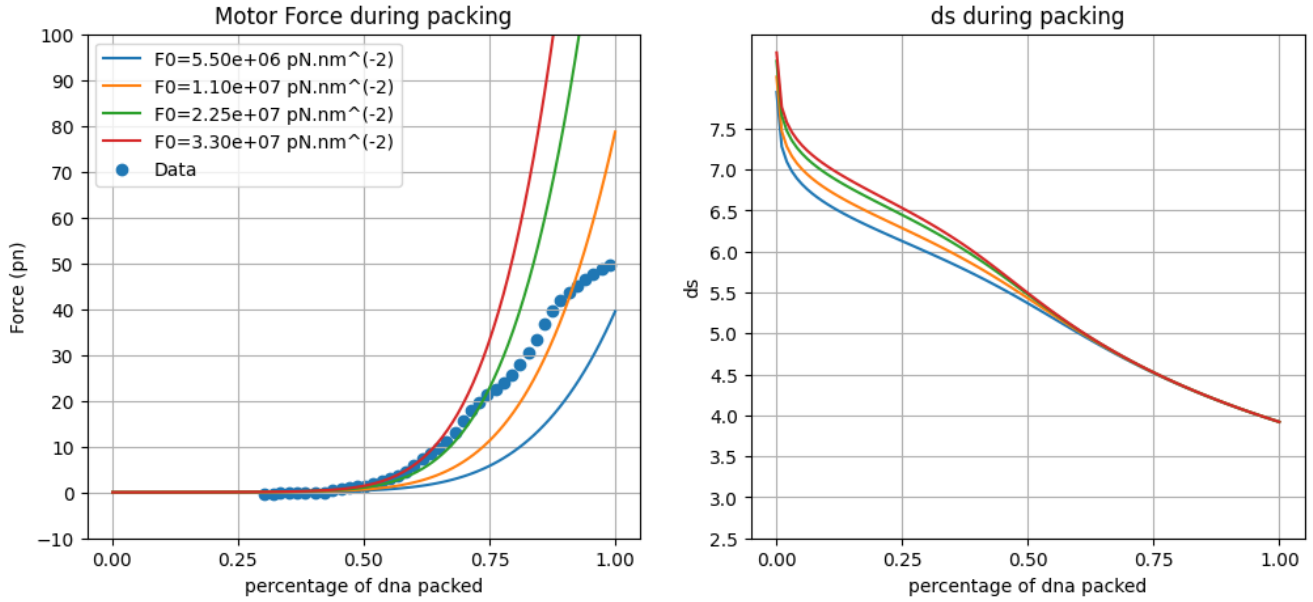
6

Figure 5: We implemented these expressions in python. For each step, we set the parameter p (percentage of DNA packed) determine d_s via a the routine `scipy.optimize.root()`. We used the `'lm'` (Levenberg-Marquart) method (other method where not achieving to converge). Then, usind d_s and p one can compute the force. The parameters we used are : $k_B = 1.380.10^{-23} J K^{-1}$, $T = 310 K$, $R_{out} = 47 nm$, $h = 37.9 nm$, $\xi_p = 25 nm$, $L = 3 * 6.584e3 nm$, $c_0 = 0.27 nm$ (close to [PKP03])

We where able to implement these expression numerically and find back at a qualitative level the plot presented in [PKP03]. Here we have considered that L is the length of the DNA strand (in full) which means that when it is not fully packed, one should replace L by $L \times p$ (this is true in all the above expressions).

We observe an evolution of d_s that seems pseudo-exponential around 0. This is because at first the strand will minimise their energy by being packed as far away from each other as possible, but will very quickly pack closer to each other to then be able to have a large enough bend radius; thus moving to the next regime we observe.

We then observe a seemingly linear decrease in d_s with the packing percentage, as decreasing d_s is a trade-off between allowing more strand to have a larger bending radius thus minimising the bending energy, and having strands closer together thus increasing the repulsion energy.

We also observe a slight change of regime in the evolution of d_s between $p \in [5\%, 45\%]$ where the trajectory of d_s is affected by $F0$, and $p \in [45\%, 100\%]$ where the trajectory of d_s is independent (and smilingly in continuity with the trajectory found at small $F0$).

We can thus suppose that for small packing percentages ($p \in [5\%, 45\%]$) the repulsion energy (G_{charge} dependent in $F0$) has a significant impact, whereas for large packing percentages ($p \in [45\%, 100\%]$) the bending energy (G_{bend}) dominates the determination of d_s. This can be qualitatively understood by the fact that for high packing percentages, smaller and smaller bend radius occur.

Finally we observe that our model for the force properly describe the evolution of the experimentally observed packing force for $F0 = 2.25.10^7 pN.nm^{-2}$ for packing percentages $p \leq 75\%$. For packing percentages greater than 75% our model doesn't describe the inflection of the force.

3.2 Spherical capsid

We can do the same job for a spherical capsid.

$$h(R) = 2R_{out}\sqrt{1 - \frac{R^2}{R_{out}^2}} \tag{22}$$

$$R_{int} = \left(1 - \left(\frac{d_s^2 L\sqrt{3}}{2\frac{4\pi}{3}R_{out}^3}\right)^{\frac{2}{3}}\right)^{\frac{1}{2}} R_{out} \tag{23}$$

$$\rho_{pack} = \frac{\frac{\sqrt{3}}{2}d_s^2 L}{\frac{4}{3}\pi R_{int}^3} \tag{24}$$

$$\int_{R_{int}}^{R_{out}} dR\frac{h(R)}{R} = -2R_{out}\left[\ln\left(\left(1 - \rho_{pack}^{2/3}\right)^{-1/2} - \frac{\rho_{pack}^{1/3}}{\sqrt{\left(1 - \rho_{pack}^{2/3}\right)}}\right) + \rho_{pack}^{1/3}\right] \tag{25}$$

$$\int_{R_{int}}^{R_{out}} dRh(R)R = \frac{2}{3}R_{out}^3\left(1 - \frac{R_{int}^2}{R_{out}^2}\right)^{3/2} \tag{26}$$

$$A = \frac{\xi_p k_B T}{R_{out}^2\left(1 - \rho_{pack}^{2/3}\right)} - \frac{3\xi_p k_B T}{2R_{int}}\left[\frac{\ln\left\{\left(1 - \rho_{pack}^{2/3}\right)^{-1/2} - \frac{\rho_{pack}^{1/3}}{\sqrt{\left(1 - \rho_{pack}^{2/3}\right)}}\right\} + \rho_{pack}^{1/3}}{\rho_{pack}}\right] \tag{27}$$

$$G_{bend} = \frac{-4\pi\xi_p k_B T R_{out}}{\sqrt{3}d_s^2}\left[\ln\left\{\left(1 - \rho_{pack}^{2/3}\right)^{-1/2} - \rho^{1/3}\left(1 - \rho_{pack}^{2/3}\right)^{-1/2}\right\} + \rho_{pack}^{1/3}\right] \tag{28}$$

$$F = \frac{\xi_p k_B T}{2R_{out}^2\left(1 - \rho_{pack}^{2/3}\right)} + F_0\sqrt{3}\left(c^2 + cd_s\right)\exp\left(\frac{-ds}{c}\right) \tag{29}$$

We weren't able to do the same numerical analysis as for the cylindrical capsid as our numerical solver didn't obtain a stable solution for d_s given the properties of the energy for this capsid geometry.

However looking as the equation we obtained in comparison with the one obtained for the cylindrical capsid, we can mainly observe that the force has a different order in p_{pack} ($-\frac{2}{3}$ for the spherical capsid versus -1 for the cylindrical capsid).

This difference can be qualitatively explained by the fact that the bend radius will decrease quicker for the spherical capsid as the vertical section of the capsid for a given radius is decreasing with a decreasing radius.

We could hypothesis given the insights of the numerical analysis for the cylindircal capsid that this would lead to a transition to a regime where $F0$ isn't determinant in the dynamic of d_s for smaller packing radiuses.

3.3 Capped cylindrical capsid

We will now do the same analysis as in the previous section, but using a different geometry:

$$h(R) = H + 2Rout\sqrt{1 - \frac{R^2}{R_{out}^2}} \tag{30}$$

$$\rho_{pack} = \frac{\frac{4}{3}\pi R_{out}^3 \left(1 - \frac{R_{int}^2}{R_{out}^2}\right)^{3/2} + \pi H R_{out}^2 \left(1 - \frac{R_{int}^2}{R_{out}^2}\right)}{\frac{4}{3}\pi R_{out}^3 + \pi H R_{out}^2} \tag{31}$$

$$G_{bend} = \frac{\pi k_B T \xi_p}{3 d_s^2} \left[H \ln\left(\frac{R_{out}}{R_{int}}\right) + 2R_{out} \ln\left(\frac{\sqrt{R_{out}^2 - R_{int}^2} + R_{out}}{R_{int}}\right) - 2R_{out}\sqrt{\left(1 - \frac{R_{int}^2}{R_{out}^2}\right)} \right] \tag{32}$$

$$\frac{\sqrt{3}d_s^2}{2} L = \pi H R_{out}^2 \left(1 - \frac{R_{int}^2}{R_{out}^2}\right) + \frac{4\pi}{3} R_{out}^3 \left(1 - \frac{R_{int}^2}{R_{out}^2}\right)^{\frac{3}{2}} \tag{33}$$

$$A = \frac{\xi_p k_B T}{R_{int}^2} + \frac{\xi_p k_B T}{2} \frac{H \ln\left\{\left(\frac{R_{out}}{R_{int}} - R_{out}\right) - \sqrt{1 - \frac{R_{int}^2}{R_{out}^2}}\right\} - \ln\left\{\frac{R_{out}}{R_{int}}\right\} - \sqrt{\frac{R_{out}^2}{R_{int}^2} - 1}}{\frac{4\pi}{3} R_{out}^3 \left(1 - \frac{R_{int}^2}{R_{out}^2}\right)^{3/2} + \pi H R_{out}^2 \left(1 - \frac{R_{int}^2}{R_{out}^2}\right)} \tag{34}$$

This very last expression could in principle be used to express R_{int} it is doable analytically with help of a formal calculus program like Wolfram. But the expression obtained are so cumbersome that we will not even try to present them here. Expressing everything in terms of ρ_{pack} seems hopeless.

In the impossibility of doing a numerical analysis, we can qualitatively say that the capped cylindrical capsid has two limits: for $R_{out} \gg H$ we end up with a capsid that can be considered spherical. In the opposite limit, for $H \gg R_{out}$ we end up with a capsid that will be dominated by its thin cylindrical body, and can thus be analysed in this limit.

4   Conclusion

We were able through a physical description of the DNA packing dynamic, and through analytical computation to describe the different terms contribution to the energy taken to pack DNA.

We were then capable of using theses computation to numerically obtain results for the simpler cylindrical capsid geometry, and get a qualitative understanding of these results which we could carry over to the other capsid geometry which we were not able to numerically analyse.

We were also able to separate the packing process between two regime were different contribution would decide the inter-strand spacing. We were however unable to explain the inflection observed in the real packing force as our model's packing force has a continuous exponential growth in the force with the packing percentage.

References

[Pur+05]   Prashant K. Purohit et al. In: *Biophysical Journal* 88(2) (Feb. 2005), pp. 851–866.

[PH08]   Anton S Petrov and Stephen C Harvey. "Packaging double-helical DNA into viral capsids: structures, forces, and energetics". en. In: *Biophys. J.* 95.2 (July 2008), pp. 497–502.

[Ber+14]   Zachary T Berndsen et al. "Nonequilibrium dynamics and ultraslow relaxation of confined DNA during viral packaging". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 111.23 (June 2014), pp. 8345–8350.

[Com+08]   Luis R. Comolli et al. "Three-dimensional architecture of the bacteriophage 29 packaged genome and elucidation of its packaging process". In: *Virology* 371.2 (2008), pp. 267–277. ISSN: 0042-6822. DOI: https://doi.org/10.1016/j.virol.2007.07.035. URL: https://www.sciencedirect.com/science/article/pii/S0042682207004990.

[PKP03]   Prashant K. Purohit, Jane Kondev, and Rob Phillips. "Mechanics of DNA packaging in viruses". In: *Proceding of the nationla academy of sciences USA* 18 (Mar. 2003), pp. 3173–8. DOI: 10.1073/pnas.0737893100.

[Tao+98]   Yizhi Tao et al. "Assembly of a Tailed Bacterial Virus and Its Genome Release Studied in Three Dimensions". In: *Cell* 95.3 (1998), pp. 431–437. ISSN: 0092-8674. DOI: https://doi.org/10.1016/S0092-8674(00)81773-0. URL: https://www.sciencedirect.com/science/article/pii/S0092867400817730.

[STS01]   D. Smith, S. Tans, and S. et al Smith. "The bacteriophage ϕ29 portal motor can package DNA against a large internal force". In: *Nature* 413 (Mar. 2001), pp. 748–752. DOI: https://doi.org/10.1038/35099581.

[Alb+02]   B. Alberts et al. *Molecular Biology of the Cell. 4th edition.* New York: Garland Science, 2002. Chap. The Structure and Function of DNA, see figure 4-5. URL: https://www.ncbi.nlm.nih.gov/books/NBK26821/.

[SS15]   Amar Singh and Navin Singh. "Effect of salt concentration on the stability of heterogeneous DNA". In: *Physica A: Statistical Mechanics and its Applications* 419 (2015), pp. 328–334. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2014.10.029. URL: https://www.sciencedirect.com/science/article/pii/S037843711400867X.

Appendix

Details of the computations of section 2

Length of DNA in the capsid.

$$L = \int_0^L ds \tag{35}$$

$$= \int_{R_{int}}^{R_{out}} 2\pi R \frac{2N(R)}{\sqrt{3}d_s} dR \tag{36}$$

$$= \frac{4\pi}{\sqrt{3}d_s^2} \int_{R_{int}}^{R_{out}} Rh(R) dR \tag{37}$$

Minimization of G

The conformation that DNA will take at steady state is the one corresponding to d_s minimal. We thus have to minimize G_{tot}. Let us recall the expression :

$$G_{tot} = \frac{2\pi\xi_p k_B T}{\sqrt{3}d_s^2} \int_{R_{int}}^{R_{out}} \frac{h(R)}{R} dR + L\sqrt{3}F_0 \left(c^2 + cd_s\right) \exp\left(\frac{-d_s}{c}\right)$$

We have to take the total derivative of this expression, in d_s when L is constant. The part on the right of the plus sign, containing the electric repulsion between the strands leads to:

$$\frac{dG_{charge}}{dd_s} = -L\sqrt{3}F_0 d_s \exp\left(\frac{-d_s}{c}\right) \tag{38}$$

Then we have to compute the derivative of the left side of the "+" sign. We have to take care that we must not derivate the L but we have to take the derivativ of the integral term which depends implicitly on d_s via R_{int}.

Let us first derive some intermediary relation that will show up to be very usefull afterwards. First let us express the volume of DNA inside the capsid.

$$\int_{R_{int}}^{R_{out}} 2\pi R h(R) dR = \frac{d_s^2 \sqrt{3}}{2} L \tag{39}$$

$$\frac{d}{dR_{int}} \int_{R_{int}}^{R_{out}} 2\pi R h(R) dR = -2\pi R_{int} h(R_{int}) $$

$$\frac{d}{dR_{int}} \int_{R_{int}}^{R_{out}} R h(R) dR = \frac{d_s \sqrt{3}L}{2\pi} \frac{dd_s}{dR_{int}} \tag{40}$$

We use equation (40) to express:

$$\frac{dR_{int}}{dd_s} = \frac{-d_s \sqrt{3}L}{2\pi R_{int} h(R_{int})} \tag{41}$$

We deduce from equations (41) that:

$$\frac{d}{dd_s} \int_{R_{int}}^{R_{out}} dR \frac{h(R)}{R} = \frac{\sqrt{3}L}{2\pi R_{int}^2} \tag{42}$$

We can reintroduce this expression in G_{bend} we have:

$$\frac{dG}{dd_s} = \frac{-2\pi\xi_p k_B T}{\sqrt{3}d_s^3} \int_{R_{int}}^{R_{out}} dR \frac{h(R)}{R} + \frac{k_B T L}{\sqrt{3}R_{int}^2} \tag{43}$$

11

We can put equations (38, 43).

$$\frac{dG_{tot}}{dd_s} = \frac{\xi_p k_B T}{d_s R_{int}^2} L - \frac{2\pi \xi_p k_B T}{\sqrt{3} d_s^3} \int_{R_{int}}^{R_{out}} dR \frac{h(R)}{R} - L\sqrt{3} F_0 d_s \exp\left(\frac{-d_s}{c}\right)$$

$$\sqrt{3} F_0 \exp\left(\frac{-d_s}{c}\right) = \frac{\xi_p k_B T}{d_s^2 R_{int}^2} - \frac{2\pi \xi_p k_B T}{\sqrt{3} d_s^4} \frac{\int_{R_{int}}^{R_{out}} dR \frac{h(R)}{R}}{L} \tag{44}$$

Using the expression of equation (39) we have the expression:

$$\sqrt{3} F_0 \exp\left(\frac{-d_s}{c}\right) = \frac{\xi_p k_B T}{d_s^2 R_{int}^2} - \frac{\xi_p k_B T}{d_s^2} \frac{\int_{R_{int}}^{R_{out}} \frac{h(R)}{R} dR}{\int_{R_{int}}^{R_{out}} R h(R) dR} \tag{45}$$

Computation of the force

In this appendix we drive the expression (5). We compute the derivativ of G in L assuming d_s constant. This assumption is valid, because the variation dd_s associated to dL is of higher order. This is not trivial to prove that but it is a reasonable assumption given the shape of the equation of d_s (4).

$$\frac{dG_{tot}}{dL} = \frac{2\pi k_B T \xi_p}{\sqrt{3} d_s^2} \left(-\frac{h(R_{int})}{R_{int}}\right) \frac{dR_{int}}{dL} + F_0 \sqrt{3} \left(c^2 + cd_s\right) \exp\left(\frac{-d_s}{c}\right) \tag{46}$$

To compute $\frac{dR_{int}}{dL}$ do as in equation (39, 41) but this time we derive by L keeping d_s constant.

$$-R_{int} h(R_{int}) = \frac{d_s^2 \sqrt{3}}{4\pi} \frac{dL}{dR_{int}}$$

$$\frac{dR_{int}}{dL} = \frac{-d_s^2 \sqrt{3}}{4\pi R_{int} h(R_{int})}$$

Reintroducing this result in (46) we obtain (5).

Computation of the force for different geometries

Cylindric capsid

1. Derivation of the force:

$$F_{mot} = -\frac{\partial G_{tot}}{\partial L} = -\left(\frac{\partial G_{charge}}{\partial L} + \frac{\partial R_{int}}{\partial L} \frac{\partial G_{bend}}{\partial R_{int}}\right)$$

$$F_{mot} = -3G_{el}(d_s) - \left(\frac{\partial L}{\partial R_{int}}\right)^{-1} \frac{\partial G_{bend}}{\partial R_{int}}$$

$$\frac{\partial L^{cylinder}}{\partial R_{int}} = -\frac{4\pi}{\sqrt{3} d_s^2} H R_{int}$$

$$\frac{\partial G_{bend}^{cylinder}}{\partial R_{int}} = -\frac{2\pi \xi_p k_B T}{\sqrt{3} d_s^2} \frac{H}{R_{int}}$$

$$F_{mot}^{cylinder} = -3G_{el}(d_s) + \frac{\xi_p k_B T}{2R_{int}^2} = -\sqrt{3} F_0(c^2 + cd_s) \exp(-\frac{d_s}{c}) + \frac{\xi_p k_B T}{2(R_{out}^2 - \frac{\sqrt{3} d_s^2 L}{2\pi H})}$$

2. Derivation of the optimal inter-strand spacing:

Using the general equation (39), we obtain the following condition for d_s:

$$\sqrt{3}F_0 \exp\left(\frac{-d_s}{c}\right) = \frac{\xi_p k_B T}{d_s^2 R_{int}^2} - \frac{2G_{bend}^{cylinder}}{L d_s^2}$$

$$\sqrt{3}F_0 \exp\left(\frac{-d_s}{c}\right) = \frac{\xi_p k_B T}{d_s^2 R_{int}^2} - \frac{2\xi_p k_B T}{d_s^2} \frac{\ln(\frac{R_{out}}{R_{int}})}{R_{out}^2 - R_{int}^2}$$

Spherical capsid

1. Derivation of the force:

$$\frac{\partial G_{charge}^{sphere}}{\partial L} = \sqrt{3}F_0(c^2 + cd_s)\exp(-\frac{d_s}{c})$$

$$\frac{\partial G_{bend}^{sphere}}{\partial L} = \left(\frac{\partial L}{\partial R_{int}}\right)^{-1} \frac{\partial G_{bend}^{sphere}}{\partial R_{int}}$$

$$\left(\frac{\partial L}{\partial R_{int}}\right)^{-1} = \frac{\sqrt{3}d_s^2}{8\pi R_{int}} \frac{1}{\sqrt{R_{out}^2 - R_{int}^2}}$$

$$\frac{\partial G_{bend}^{sphere}}{\partial R_{int}} = \frac{4\pi\xi_p k_B T}{\sqrt{3}d_s}\left[\frac{R_{int}}{\sqrt{R_{out}^2 - R_{int}^2}} - \frac{R_{out}^2}{R_{int}\sqrt{R_{out}^2 - R_{int}^2}}\right] = \frac{4\pi\xi_p k_B T}{\sqrt{3}d_s} \frac{R_{out}^2 - R_{int}^2}{R_{int}\sqrt{R_{out}^2 - R_{int}^2}}$$

$$\Rightarrow \frac{\partial G_{bend}^{sphere}}{\partial L} = -\frac{\xi_p k_B T}{2R_{int}^2} = -\frac{\xi_p k_B T}{2(R_{out}^2 - (\frac{3\sqrt{3}d_s^2 L}{8\pi})^{\frac{2}{3}})}$$

where we used the equation relating the internal radius $R_{int} = \sqrt{R_{out}^2 - (\frac{3\sqrt{3}d_s^2 L}{8\pi})^{\frac{2}{3}}}$ to the packaged length L.

The total force is thus:

$$F_{tot}^{sphere} = -\frac{\partial G_{tot}^{sphere}}{\partial L} = -\sqrt{3}F_0(c^2 + cd_s)\exp(-\frac{d_s}{c}) + \frac{\xi_p k_B T}{2(R_{out}^2 - (\frac{3\sqrt{3}d_s^2 L}{8\pi})^{\frac{2}{3}})}$$

2. Derivation of the optimal inter-strand spacing:

Using the general equation (4), we obtain the following condition for d_s:

$$\sqrt{3}F_0 \exp\left(\frac{-d_s}{c}\right) = \frac{\xi_p k_B T}{d_s^2 R_{int}^2} - \frac{2G_{bend}^{sphere}}{L d_s^2}$$

$$\sqrt{3}F_0 \exp\left(\frac{-d_s}{c}\right) = \frac{\xi_p k_B T}{d_s^2 R_{int}^2} - \frac{3\xi_p k_B T}{d_s^2}\left[\frac{1}{R_{out}^2 - R_{int}^2} + \frac{R_{out}}{(R_{out}^2 - R_{int}^2)^{\frac{3}{2}}} \ln\left(\frac{R_{out} - \sqrt{R_{out}^2 - R_{int}^2}}{R_{int}}\right)\right]$$

Capped cylindrical capsid

For a capped cylinder, our starting point is the height of the capsid as a function of R.

$$h(R) = h + 2R_{out}\sqrt{1 - \frac{R_{int}}{R_{out}}} \tag{47}$$

$$\frac{\sqrt{3}d_s^2 L}{2} = \left[\frac{4\pi}{3}R_{out}^3\left(1 - \frac{R_{int}^2}{R_{out}^2}\right)^{\frac{3}{2}} + hR_{out}^2\pi\left(1 - \frac{R_{int}^2}{R_{out}^2}\right)\right] \tag{48}$$

(49)