

# Modeling project: DNA packaging in a virus

October 2023

## Abstract

In order to infect a host cell, 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 ignited. 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 viruses is thus an essential step in their life cycle. In this article, we describe a simple physical model explaining this packaging process for diverse shaped capsids.

## 1 Introduction: Orders of magnitudes

The dynamic of the packing of DNA in a bacteriophage and viruses has been investigated experimentally and theoretically and is now quite elucidated (see [Pur+05] for a complete review). These organisms are forms by a icosahedral capsid that encloses a double stranded beam of DNA. Inside the capsid the DNA is just solvated in an ionic solution. There is no tertiary structure formed of histones and nucleosomes as in eucariot cells. The life cycle of the virus is as follow. It interacts via its spike protein with the membrane of the cell and is able to introduce its genome inside the cell. The DNA of the viruse is then replicated by the cell and expressed to produce the proteins necessary to forme other capsids. The formation of a new viral capsid take typically 10 to 15 minutes for phage  $\lambda$  or T4. When a capsid is formed and the genome completely replicated, the portal motor of the capsid, will package the DNA inside the the capsid. This process is typically realised in 5 minutes. After approximately one hour, the cell tends to be full of viruses and dies. This is the lisis that releases the viruses (or phages) in the medium.

The capside has the shape of a sphere or a cylinder with a semi-sphere at each end-cap. When the DNA is condensed inside the capsid, it takes the shape of a spool. The strand organise themselves in hexagon with an interstrand space of the order of  $25 \text{ \AA}$  [EH77]. The capsid radius is smaller than the persistence length of the DNA strand. which means that the DNA is highly constrained inside the capsid. The capsid is impermeable to water. But it is permeable to Polyethilen Glycole (PEG). This property have been used by Evilevitch to apply an osmotic presure against the extraction of the genom of the phage [evilevitch2003]. They report that the ejection of the DNA strand is prevented by a pressure of  $20 \text{ atm}$ .

It is believed that the DNA inside the capside organises it self as a spool. But the first experimental measurement of the spatial organisation of DNA inside the capsid, during the packing process has been observed only in 2008 by Comoli and al. [comoli2008]. They showed that contrarily to what was believed before (for instance in [prohit2005; Pur+05], DNA tends to organise in the spool structure very late in the packaging process. Until 78% of the genome is packed, the DNA inside the capsides stays in a desordered phase.

Name of the variable	Symbol	Value	Source
Diameter of the capside	$d_{cap}$	$42 \times 54 \text{ nm}$	Phage $\phi 29$ [Tao+98]
Diameter of the DNA	$d_{dna}$	$1 \text{ nm}$	[Pur+05]
Length of the DNA strand	$L_{tot}$	$5.5.10^3 \text{ nm}$	Phage $\phi 29$ [Tao+98]
Persistence length	$\xi_p$	$50 \text{ m}$	[STS01]
Gyration radius	$R_G$	$300 \text{ nm}$	phage $\phi 29$ [Pur+05]
Motor force	$F_{mot}$	$10 \text{ pN}$	[Pur+05]
Typical pressure	$p_{poly}$	$219 \text{ atm.}$	will be explained later
Pressure of non-release	$p_{nr}$	$20 \text{ atm.}$	will be explained in the text

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

## 2 Dimensional analysis model

To understand the relation between the packing percentage and the pressure inside the capsid, let us start with a small scaling argument. If the DNA is rolled in a spool like structure made of 1 fold and  $N$  hoops in each fold (like a solenoid of  $N$  turns). The length of the DNA in this spool is :  $N \sim \frac{L}{2\pi R}$ . Assuming that the inter-strand spacing is  $r_{IA}$  and assuming that the capsid is a cylinder of height  $h$ , we typically have  $\frac{\sqrt{3}}{2} N r_{IA} \approx h \approx NR$  (assuming that in fact the capsid is almost spherical and  $h \sim R$ ). An other way to evaluate the scaling law is to suppose that in fact the DNA is almost fully condensed and the cylinder is full of DNA. In this case, if we take a cut of the DNA spool parallel to the axis of the capsid, the surface occupied by DNA is typically:  $N\pi r_{IA}^2 \sim S$  where  $S \sim \pi R^2$ . Using once again the fact that  $N \sim \frac{L}{2\pi R}$  we can write  $r_{IA} \sim \frac{R^{\frac{3}{2}}}{\sqrt{L}}\sqrt{2}$ . Then to deduce a pressure, we use the scaling law that have been pointed out experimentally by Rau et al. [RP92]:  $P \sim e^{-\frac{d_s}{c}}$  (where  $c$  is typically just here to adimensionalise  $d_s$ ). We obtain the two following scaling laws:

$$P \sim e^{\frac{-2\pi R^2}{cL}} \text{ valid for small percentage of packing} \quad (1)$$

$$P \sim e^{\frac{-\sqrt{2}R^{\frac{3}{2}}}{\sqrt{cL}}} \text{ valid for high percentage of packing} \quad (2)$$

Using the fact that a pressure is proportional to a force time a surface. Here the small surface of a cut of the DNA, we can try to check our scaling law on the data provided in the subject. We took the points from the graph and plotted the pressure as a function of the inverse of the percentage of genome packed. We fitted this curve with an exponential function:  $x = e^{\gamma x}$ . This gives us the result plotted in the figure. To check that the scaling is satisfactory, we can extract back the typical radius of the capsid witch is expressed:

$$R_{typ} = \sqrt{\frac{\gamma}{2\pi}} = 6,66 \text{ for the law at low packing density} \quad (3)$$

$$R_{typ} = 1.414\gamma^{0.6666} = 32.13 \text{ for the high packaging density} \quad (4)$$

We can conclude that at least qualitatively speaking, it seems that the DNA is compact in the denser form we have proposed.

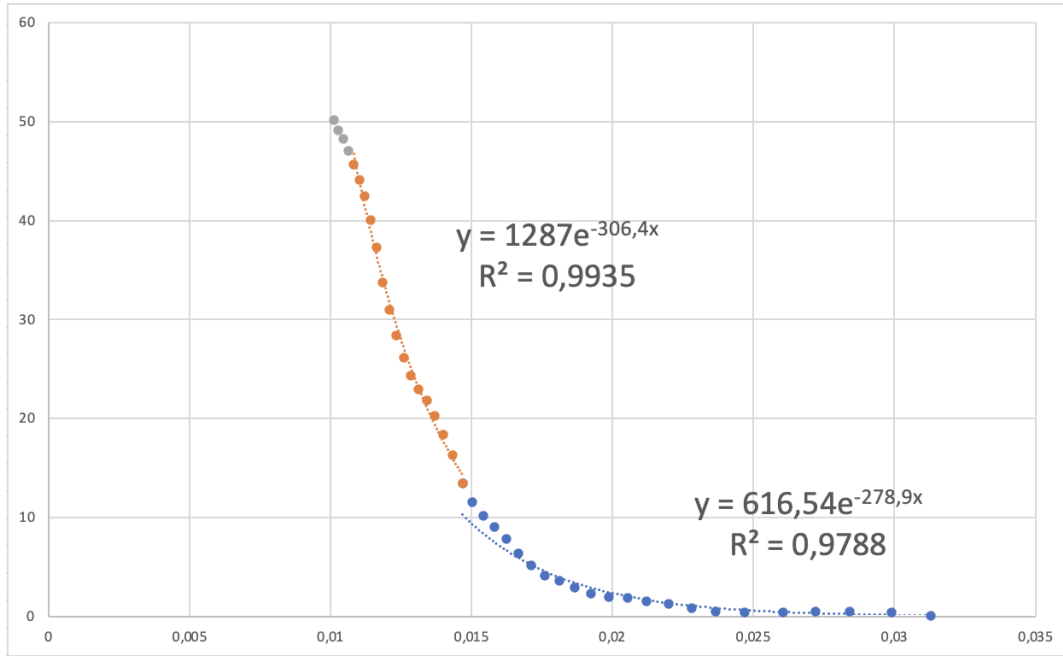


Figure 1: Force of the motor as a function of the percentage of DNA packed. The horizontal axis is the inverse of the percentage of the DNA packed while the vertical axis is the force. We fitted the blue points with the law (1) and the orange points with the law (2). The exponent  $\gamma$  of each law can be used to compute the value of the radius of the capsid.

### 3 Long polymer model of DNA inside a viral capsid

An other model that is interesting to understand the dynamic of a viral capsid is the thermodynamic model of a long polymer under confinement. The idea is simply to use what we know from tutorial 1 about the physics of polymer chains. Which is the following:

$$G_{bend} = \int_0^L \frac{k_B T \xi_p}{2} (\partial_s^2 \mathbf{r}(s))^2 ds \quad (5)$$

For a polymer which is rolled with a regular radius, and no tension, this leads to the expression

$$G_{bend} = \int_0^L \frac{k_B T \xi_p}{2} \frac{1}{R^2(s)} ds \quad (6)$$

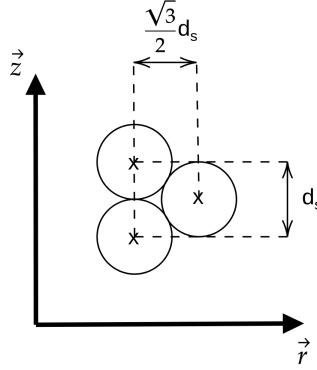


Figure 2: Schematic of the packing of DNA strands.

When we perform the integral over  $ds$ , shell by shell, we write  $N(R)$  the number of turns of radius  $R$ . Thus the radial density of turns is :  $\frac{2N(R)}{\sqrt{3}d_s}$  where  $d_s$  is the distance between two strands of DNA. The factor  $\frac{\sqrt{3}}{2} = \sin(\frac{\pi}{3})$  account to the fact that the DNA stands are organised in a triangular lattice.

TODO : add a figure with the triangular lattice and the understand space

We end up with the expression:

$$G_{bend} = \frac{k_B T \xi_p}{2} \int_{R_{in}}^{R_{out}} \frac{N(R)}{\sqrt{3}d_s} \frac{1}{R^2} dR \int_0^{2\pi R} ds \quad (7)$$

$$G_{bend} = \frac{\pi k_B T \xi_p}{\sqrt{3}d_s} \int_{R_{in}}^{R_{out}} \frac{N(R)}{R} dR \quad (8)$$

There is also room for twisting energy stored in the DNA strangds, this effect will appear as a third order derivative in the hamiltonian density. This leads to the contribution to the energy :

$$G_{twist} = \frac{\xi_{tw} k_B T}{2} \int_0^L (\partial_s^3 \mathbf{r})^2 ds \quad (9)$$

$$(10)$$

We chose to neglate this term as the current state of the art doesn't prove wether the strand can twist or not (i.e. wether the motor spins freely, is fixed, or has a specific spin patern). We thus assume by neglating this term that the motor turns freely, thus allowing the strand to un-twist.

We also note that we don't take any tension into account as the tension or compression energy is the one deriving from the pushing force of the motor, which we are trying to figure out (at mechanical equilibrium, including the tension force and thus all forces would naturally result in a derivative of the total energy equal to 0, thus not giving us the motor pushing force).

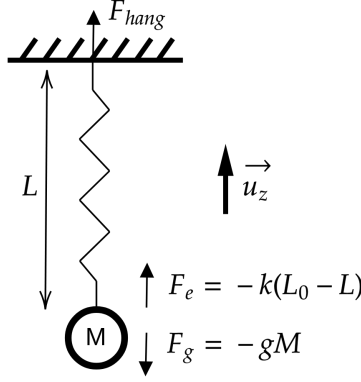


Figure 3: Simple example of forces cancelling out.

To simplify our argument we take this simple hanging mass as an example: We have here  $F_e + F_g = 0$ . But for the sake of argument and for a valid comparison let's say that we want to figure out the force  $F_{hang}$  holding the system (analogous to the pushing force of the motor in our example) from the total energy.

If we include the energy of the spring (which is analogous to the tension/compression energy of the DNA; similarly it is the way the force is transmitted), we obtain a total energy  $G_{tot} = gMh - kh^2$ , but at equilibrium forces even out, thus  $Mg = kh$ . We thus have  $G_{tot} = 0$  and  $\frac{\partial G_{tot}}{\partial h} = 0$ .

If we correctly chose to exclude the tension energy of the spring, we could have had  $G_{tot} = hMg$ , thus we correctly obtain  $F_{hang} = -\frac{\partial G_{tot}}{\partial h} = Mg$  ( $= -F_g = F_e$ )

The other term that must be taken into account in the energy in the system is the repulsion between DNA strands. In fact, each base pair is charged with a charge +2 for this purpose, 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. The force has been determined experimentally based on measurement of osmotic pressure. The force between two unit length DNA strands at distance  $d_s$  is:  $f_{el}(d_s) = \frac{F_0 d_s}{\sqrt{3}} \exp\left(\frac{-d_s}{c}\right)$ .  $F_0$  is homogeneous to a force and  $c$  to a length. They are parameters that should be determined from fits. No direct measurement have been done to our knowledge [PKP03].

$$G_{el} = \int_{-\infty}^{d_s} f_{el}(d'_s) dd'_s \quad (11)$$

$$= \int_{-\infty}^{d_s} \frac{F_0 d'_s}{\sqrt{3}} \exp\left(\frac{-d'_s}{c}\right) dd'_s \quad (12)$$

$$= \sqrt{3}F_0 (c^2 + cd_s) \text{Lexp}\left(\frac{-d_s}{c}\right) \text{ (we made an integration by part)} \quad (13)$$

In fact it has been discussed that this expression must be modified in the presence of trivalent ions. Phillips and al. derive the expression [Pur+05]:

$$G_{el} = \sqrt{3}F_0 L \left[ (c^2 + cd_s) \exp\left(\frac{-d_s}{c}\right) - (c^2 + d_s) - \frac{1}{2} (d_0^2 + d_s^2) \right] \quad (14)$$

This expression leads to a small attractive potential at long distance and strong repulsive potential for distance smaller of the order of  $d_0$ . As experimental evidences shows that when the DNA is packed, it is in this regime. This allows to use the expression 13. As the parameter  $d_0$  and  $c$  and  $F_0$  13 and 14 are determined throw a fit on experimental data of parallel DNA strands, it naturally incorporates electric repulsion and entropic effects and maybe other effect. However it does not include any contribution from binding energy.

## 4 Impact of capsid geometry on the packing pressure

We will study a variety of capsid geometries to assess the impact of this geometry on the packing pressure.

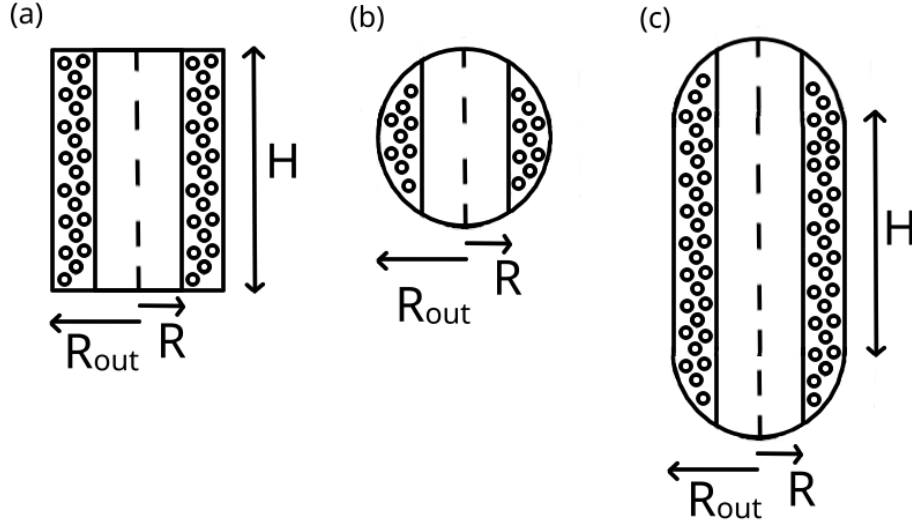


Figure 4: Geometries that will be studied analytically: (a) is a simple cylindrical capsid, (b) is a spherical capsid and (c) is a capped cylindrical capsid. We will only model cylindrical and capped cylindrical capsids as spherical capsids can be modeled from a capped-cylindrical capsid with  $H = 0$ .

#### 4.1 Cylindrical capsid

We assume that DNA strands are separated by a constant  $d_s$  distance.

We also assume that strand at a distance  $R$  from the centre are simply looping at a curvature  $\partial_z^2 \mathbf{r} = \frac{1}{R}$ .

We have at a radius  $R$   $N(R) = \frac{h(R)}{d_s} = \frac{H}{d_s}$  DNA strands per  $d_s$  unit of radius (with  $h(R)$  being the length of the vertical section at radius  $R$ , which for the simple cylindrical capsid is simply constant and equal to  $H$ ).

We thus have the bending energy  $G_{bend}$ :

$$G_{bend} = \int_R^{R_{out}} \frac{k_B T \xi_p}{2\sqrt{3}} N(r) \left( \int_0^{2\pi r} \frac{1}{r^2} dl \right) \frac{dr}{d_s} = \frac{\pi k_B T \xi_p}{\sqrt{3} d_s^2} \int_R^{R_{out}} \frac{h(r)}{r} dr \quad (15)$$

$$G_{bend}^{cylinder} = \frac{\pi k_B T \xi_p}{\sqrt{3} d_s^2} \ln \left( \frac{R_{out}}{R} \right) H \quad \text{There must an addition } \frac{1}{\sqrt{3}} \text{ factor.} \quad (16)$$

And also the electronic repulsion energy  $G_{charge}$ , assuming that we have hexagonal packed parallel strands of length  $2\pi r$ :

$$G_{charge} = \int_R^{R_{out}} 6\pi r N(r) G_{el}(d_s) dr = \frac{6\pi G_{el}(d_s)}{\sqrt{3} d_s} \int_R^{R_{out}} r h(r) dr = 3G_{el}(d_s) L_{inserted} \quad (17)$$

$$G_{charge}^{cylinder} = \frac{\sqrt{3}\pi G_{el}(d_s)}{d_s} (R_{out}^2 - R^2) H \quad (18)$$

Identifying the inserted length  $L_{inserted}$ :

$$L_{inserted} = \frac{2\pi}{\sqrt{3} d_s} \int_R^{R_{out}} r h(r) dr \quad (19)$$

$$L_{inserted}^{cylinder} = \frac{\pi}{\sqrt{3} d_s} (R_{out}^2 - R^2) H \quad (20)$$

We eliminate the variable  $d_s$  in the above equations by minimizing  $G_{tot}(d_s) = G_{bend} + G_{charge}$ :

$$\frac{\partial G_{tot}}{\partial d_s} = 0 \quad (21)$$

$$\frac{\partial G_{tot}}{\partial d_s} = \frac{\partial G_{bend}}{\partial d_s} + \frac{\partial G_{charge}}{\partial d_s} \quad (22)$$

$$\frac{\partial G_{tot}^{cylinder}}{\partial d_s} = -\frac{2\pi k_B T \xi_p}{\sqrt{3}d_s^3} \ln\left(\frac{R_{out}}{R}\right) H \quad (23)$$

$$+ \sqrt{3}\pi \left[ \frac{f_{el}(d_s)}{d_s} - \frac{G_{el}(d_s)}{d_s^2} \right] (R_{out}^2 - R^2) H \quad (24)$$

$$\frac{\partial G_{tot}^{cylinder}}{\partial d_s} = -\frac{2\pi k_B T \xi_p}{\sqrt{3}d_s^3} \ln\left(\frac{R_{out}}{R}\right) H + \pi F_0 \left[ 1 - 3\frac{c}{d_s} - 3\frac{c^2}{d_s^2} \right] \exp\left(\frac{-d_s}{c}\right) (R_{out}^2 - R^2) H \quad (25)$$

$$d_{s,min}^{cylinder} = TODO \quad (26)$$

## 4.2 Capped cylindrical capsid

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

$$G_{bend}^{capped\_cyl.} = \frac{\pi k_B T \xi_p}{3d_s^2} \int_{R_{int}}^{R_{out}} \frac{h + 2R\sqrt{1 - \frac{r^2}{R_{out}^2}}}{r} dr \quad (27)$$

$$= \frac{\pi k_B T \xi_p}{3d_s^2} \left[ h \ln\left(\frac{R_{out}}{R_{in}}\right) + 2R_{out} \ln\left(\frac{\sqrt{R_{out}^2 - R_{int}^2} + R_{out}}{R_{int}}\right) - 2R_{out} \sqrt{\left(1 - \frac{R_{in}^2}{R_{out}^2}\right)} \right] \quad (28)$$

$$L_{inserted}^{capped\_cyl.} = \frac{\pi}{\sqrt{3}d_s} (R_{out}^2 - R_{in}^2) H + \frac{8\pi}{3\sqrt{3}d_s^2} (R_{out}^2 - R^2)^{\frac{3}{2}} \quad (29)$$

To obtain  $G_{charge}$  we can simply remark that it is the sum of the  $G_{el}$  in the spherical capsid plus the one in the cylindrical capsid. We obtain:

$$G_{charge}^{capped\_cyl.} = \frac{\sqrt{3}\pi G_{el}(d_s)}{d_s} \left[ (R_{out}^2 - R^2) + \frac{2R_{out}}{3} \left( 1 - \frac{R_{in}^2}{R_{out}^2} \right)^{\frac{3}{2}} \right] \quad (31)$$

We eliminate the variable  $d_s$  in the above equations by minimizing  $G_{tot}(d_s) = G_{bend} + G_{charge}$ :

$$TODO \quad (32)$$

## 4.3 Capsid geometry comparison

Name of the variable	cylindrical capsid (a)	spherical capsid (b)	capped cylindrical capsid (c)
$h(R)$	$H$	$2\sqrt{R_{out}^2 - R^2}$	$H + 2\sqrt{R_{out}^2 - R^2}$
$L_{inserted}$	$\frac{\pi}{\sqrt{3}d_s} (R_{out}^2 - R^2) H$	$\frac{8\pi}{3\sqrt{3}d_s^2} (R_{out}^2 - R^2)^{\frac{3}{2}}$	$\frac{\pi}{\sqrt{3}d_s} (R_{out}^2 - R^2) H + \frac{8\pi}{3\sqrt{3}d_s^2} (R_{out}^2 - R^2)^{\frac{3}{2}}$
$G_{bend}$	$\frac{\pi k_B T \xi_p}{d_s^2} \ln\left(\frac{R_{out}}{R}\right) H$	$\frac{-4\pi \xi_p k_B T}{\sqrt{3}d_s} [\sqrt{R_{out}^2 - R^2} + \ln(\frac{R_{out} - \sqrt{R_{out}^2 - R^2}}{R})]$	TODO
$G_{charge}$	$\frac{3\pi G_{el}(d_s)}{d_s} (R_{out}^2 - R^2) H$	$\frac{8\pi}{\sqrt{3}d_s^2} F_0(c^2 + d_s)(R_{out}^2 - R^2)^{\frac{3}{2}} \exp(-\frac{d_s}{c})$	TODO
$G_{tot}$	TODO	TODO	TODO
$\frac{G_{tot}}{L_{inserted}}$	TODO	TODO	TODO

Table 2: Analytical analysis of the various capsid geometries.

Blabla...

## 5 Estimation of the pressure inside the capsid using the charge of DNA

Right what I have in my notes.

## 6 Entropic Considerations

Let us modelise here the DNA as chaine of rods-link representing a polymer. A configuration is fully specified by giving the value of 6 coefficients  $(k_i)_{i \in [1,6]}$  and  $N = \sum_i k_i$ . Then the number of configurations is:

see page 371 of Phillips

Why is it different from the expression derived phenomenologicaly in the paper ? Can we bridge the gap in some manner ?

$$W(k_i, N) = \frac{N!}{\prod_{i=1}^6 k_i!} \quad (33)$$

And the entropy then reads:

$$S = -k_B T \ln \left( \frac{N!}{\prod_{i=1}^6 k_i!} \right) \quad (34)$$

I have never seen this contribution taken into account anywhere. In the fully condensed phase maybe we can neglect it. But when the DNA is less than 78% condensed, the disorder might play an important role.

It would be great to understand why is it ignored in th papper we have reard ?

## References

- [Pur+05] Prashant K. Purohit et al. In: *Biophysical Journal* 88(2) (Feb. 2005), pp. 851–866.
- [EH77] W. C. Earnshaw and S. C. Harrison. “DNA arrangement in isometric phage heads”. In: *Nature* 268 (Aug. 1977), pp. 598–602. DOI: <https://doi.org/10.1038/268598a0>.
- [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](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 29 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>.
- [RP92] D.C. Rau and V.A. Parsegian. “Direct measurement of the intermolecular forces between counterion-condensed DNA double helices. Evidence for long range attractive hydration forces”. In: *Biophysical Journal* 61.1 (1992), pp. 246–259. ISSN: 0006-3495. DOI: [https://doi.org/10.1016/S0006-3495\(92\)81831-3](https://doi.org/10.1016/S0006-3495(92)81831-3). URL: <https://www.sciencedirect.com/science/article/pii/S0006349592818313>.
- [PKP03] Prashant K. Purohit, Jane Kondev, and Rob Phillips. “Mechanics of DNA packaging in viruses”. In: *Proceeding of the nationla academy of sciences USA* 18 (Mar. 2003), pp. 3173–8. DOI: [10.1073/pnas.0737893100](https://doi.org/10.1073/pnas.0737893100).