

# Compaction of nucleic acids

---

**Marion, Sanjin**

**Doctoral thesis / Disertacija**

**2017**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:217:999923>

*Rights / Prava:* [In copyright](#)/[Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2024-07-06**



*Repository / Repozitorij:*

[Repository of the Faculty of Science - University of Zagreb](#)





University of Zagreb

FACULTY OF SCIENCE  
DEPARTMENT OF PHYSICS

Sanjin Marion

**Compaction of nucleic acids: physical  
mechanisms and biological relevance**

DOCTORAL THESIS

Zagreb, 2017.





University of Zagreb

FACULTY OF SCIENCE  
DEPARTMENT OF PHYSICS

Sanjin Marion

**Compaction of nucleic acids: physical  
mechanisms and biological relevance**

DOCTORAL THESIS

Supervisor:  
dr. sc. Antonio Šiber

Zagreb, 2017.





Sveučilište u Zagrebu

PRIRODOSLOVNO - MATEMATIČKI FAKULTET  
FIZIČKI ODSJEK

Sanjin Marion

**Sažimanje nukleinskih kiselina: fizikalni  
mehanizmi i biološka relevantnost**

DOKTORSKI RAD

Mentor:  
dr. sc. Antonio Šiber

Zagreb, 2017.



# Acknowledgements

My gratitude goes to all the people I've had the pleasure to work with. Special thanks to my supervisor Antonio Šiber for showing me what science is, and what isn't. I would also like to thank all the nano&bio people at the Institute of Physics. To all who guided me through Science and Life: Damir, Siniša, Darko, Marko and Tomislav, and to Tom for providing new opportunities, I am grateful.

Special thanks goes to my family, who've supported me through this journey. To Ida, you are the joy of my life. Without you I wouldn't have made the journey.





# Summary

**Keywords:** physical virology, nucleic acids, confinement, DNA condensation, protein-DNA mixtures, crowding

Packing of nucleic acids inside (naturally occurring) confined spaces presents an intriguing problem of compacting a long and highly charged polymer into a small space possibly crowded with other particles (proteins). For example, viruses have a large amount of genomic information that is encoded in nucleic acids packed in small spaces resulting in high densities of matter. The arising interactions are coupled to the confinement giving a more complex phase diagram than expected in bulk. In this work we study the problem of packing nucleic acids in confined spaces in the context of physical virology. First, we study compacted states of DNA including condensed DNA in cells and confined DNA in bacteriophage capsids. We apply polymer and liquid crystal theory along with mean field approximations for the bending energy to characterize the state of DNA. The resulting framework is used to explain *in vivo* ejection of DNA from a bacteriophage into a Gram-positive bacteria based only on thermodynamic considerations, without invoking any active cellular mechanisms. The packing mechanism for DNA with condensing proteins in adenoviruses is studied by comparing Langevin dynamics simulations of effective particle models, representing condensing proteins, with experimental data. The DNA is found to act as an effective medium for condensing core protein interactions. A backbone of DNA linking the condensing proteins is not needed to explain the experimental results. To further explain such systems, we construct a full model of packed polymer and condensing proteins inside spherical confinement using Langevin dynamics. Internal organization of condensing particles shows that they tend to cover themselves with the DNA polymer which provides an effective medium for interactions with other condensers, confirming the applicability of our effective model for core particle organization in adenoviruses. Crowding of the viral interior and confinement influences the conformation of the DNA and protein, facilitating more direct contacts between the DNA polymer and the condensing particles, and modifying the interactions between them. Our model is able to explain the general internal organisation of adenovirus cores, and provide insight into packing of genetic material in similar systems.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Background . . . . .	1
1.1.1	Nucleic acids as polymers . . . . .	1
1.1.2	Electrostatics in solutions . . . . .	3
1.1.3	Virology . . . . .	4
1.2	Overview . . . . .	5
1.3	Outline of work . . . . .	7
<b>2</b>	<b>Compacted DNA</b>	<b>9</b>
2.1	Condensation of DNA . . . . .	9
2.1.1	Ubbink-Odijk continuum model . . . . .	10
2.1.2	Optimizing the DNA condensate shapes . . . . .	14
2.1.3	Minimal packed length for condensate stability . . . . .	16
2.2	Phases of confined DNA . . . . .	17
2.3	Bending of confined DNA . . . . .	19
2.3.1	Mean-field bending model . . . . .	20
2.3.2	Loop bending model . . . . .	24
2.4	Confined DNA in the isotropic regime . . . . .	28
2.5	Liquid crystalline DNA phase . . . . .	32
2.6	Viral ejection in vivo . . . . .	38
2.6.1	DNA inside the cell . . . . .	39
2.6.2	DNA in the capsid . . . . .	41
2.6.3	Tug of war . . . . .	42

---

<b>3</b>	<b>Nucleic acids and condensing proteins in confinement</b>	<b>47</b>
3.1	DNA and core protein organization in adenoviruses . . . . .	48
3.1.1	Cryo-EM of Adenovirus cores . . . . .	49
3.1.2	Statistical analysis of core particle positions . . . . .	51
3.1.3	Modelling the core proteins . . . . .	53
3.1.4	Core organization as a mixture of effective particles . . . . .	59
3.2	Mixtures of a polymer and condensing particles in confinement . . . . .	61
3.2.1	Molecular dynamics simulations . . . . .	61
3.2.2	Statistical indicators . . . . .	66
3.2.3	Pressures of packing . . . . .	69
3.2.4	Internal organization . . . . .	77
3.2.5	Opening the capsid . . . . .	90
3.2.6	Implications for viral packing . . . . .	92
<b>4</b>	<b>Conclusions</b>	<b>97</b>
	<b>Prošireni sažetak</b>	<b>99</b>
1	Uvod . . . . .	99
2	Sažimanje DNK . . . . .	101
2.1	Kondenzirana DNK . . . . .	101
2.2	Faze zatočene DNK . . . . .	101
2.3	Savijanje zatočene DNK . . . . .	102
2.4	Zatočenje u izotropnom režimu . . . . .	102
2.5	Faza tekućeg kristala . . . . .	103
2.6	Viralna ejakcija u bakteriju . . . . .	103
3	Nukleinske kiseline i kondenzirajući proteini u prostornom zatočenju . . . . .	104
3.1	Organizacija DNK i jezgrenih proteina u adenovirusima . . . . .	104
3.2	Mješavine polimera i kondenzirajućih proteina u prostornom zatočenju	106
4	Zaključak . . . . .	108
	<b>Bibliography</b>	<b>111</b>
	<b>Curriculum vitae</b>	<b>123</b>

## CHAPTER 1

# Introduction

## Prologue

Applying physical concepts to biological problems is not a new endeavour. Still, due to the complexity of physical phenomena of Life, there is ample opportunity for new developments. It is within this overlap of physics and biology that we find the problem of compaction of nucleic acids. The problem that all organisms face, from viruses and cells to humans, is how to store the largest amount of information in the smallest space possible, be it genetic information or the entirety of knowledge accumulated by mankind. What interests us is how simple organisms compact this information, and store it for future use.

The simplest organism we can find is probably a virus, although it is debatable if it is alive. Consisting of barely a container and its genetic information, it is the prototype of one of the simplest "nano-machine" Nature has produced. If we would aspire to make nano-machines of our own, it is in viruses we should find inspiration.

## 1.1 Background

### 1.1.1 Nucleic acids as polymers

A nucleic acid is the basis of Life, containing in itself the instructions for the assembly of all proteins necessary for functioning of organisms. The nucleic acids DNA (deoxyribonucleic acid) and RNA (ribonucleic acid), are polymers made from monomers known as nucleotides. A pair of nucleotides in a double stranded DNA forms one monomeric unit called a base pair. They form the building blocks of the DNA double helix, and contribute to the folded structure of both DNA and RNA. Our primary interest lies in the compaction of either double stranded DNA (dsDNA) or single stranded RNA (ss-RNA). Double stranded DNA, in the most common B-form, has base pairs (bp) of length  $L_{bp} = 0.34$  nm and base radius  $D \approx 1$  nm. Single stranded RNA on the other hand, has a radius of about  $D \approx 0.5$  nm.

The simplest model of a polymer is that of an ideal chain [1]. An ideal chain can be represented as a random walk of a step of length  $a$ , the monomer unit length, in a three-dimensional space. For example, RNA can be represented as a freely rotating chain with taking into account that the angle  $\theta$  between two neighbouring bond vectors will depend on the chemical structure of the bond. This enables the calculation of many statistical parameters of such a polymer [1].

The most important difference between RNA and DNA will be in the stiffness of such molecules. While, RNA is well represented by a freely jointed chain with a monomer size  $a = 1$  nm consisting of  $\sim 3$  nucleotides, DNA is usually better represented as a **worm like chain** (the Kratky-Porod model) [1]. In contrast to the ideal chain where each bond is rigid, the worm like chain is not completely rigid and can fluctuate and bend. The length scale on which the natural fluctuations of the worm like chain overcome the inherent chain stiffness is called the persistence length  $L_p$ . The length scale after which the correlation of tangent vectors displaced by  $l$  on the contour reduces by a factor  $e^{-1}$ :

$$\langle u(\mathbf{r})u(\mathbf{r} + \mathbf{l}) \rangle_{\mathbf{r}} = e^{-\frac{l}{L_p}} \quad (1.1.1)$$

is by definition the persistence length  $L_p$ . This allows one to use the Kuhn length, which is twice the persistence length  $l_b = 2L_p$ , to renormalise the behaviour of the worm like chain into an ideal chain, albeit with a new bond length of  $l_b$  [2]. In addition to governing statistical behaviour of the chain,  $L_p$  is also a measure of the energy required to bend a DNA of length  $L$  into a curve with a radius of curvature  $R$  [3]:

$$F_b = \frac{1}{2}k_B T \frac{L L_p}{R^2}, \quad (1.1.2)$$

where  $L_p = EI/k_B T$  with  $E$  the Young modulus and  $I$  the moment of inertia of the cross section of DNA.

A realistic chain can not cross itself, as each segment excluded a volume in space, thus forming a self avoiding random walk. This excluded volume  $v$  is a result of the interaction potential  $U(\mathbf{r})$  between two chain segments. In essence,  $v$  is related to the Boltzmann factor for finding the segment at any point in space. The probability of finding a point in space occupied by an particle is given via the Mayer  $f$ -function [1]:

$$f(\mathbf{r}) = e^{-U(\mathbf{r})/k_B T} - 1, \quad (1.1.3)$$

such that the excluded volume is the integral of  $f(r)$  over all available volume:

$$v = - \int f(\mathbf{r}) d^3r. \quad (1.1.4)$$

The Flory theory of a polymer in a good solvent [1] then gives that the interaction energy

of a polymer with itself, or segments from another polymer, is given by the excluded volume as

$$F_{int} = k_B T \frac{vn^2}{V} \quad (1.1.5)$$

with  $n$  the number of chain beads (segments) of excluded volume  $v$  in volume  $V$ .

## 1.1.2 Electrostatics in solutions

Electrostatics in water solutions containing salt ions is governed mostly by classical physics. The existence of a solution of ions complicates matters by giving a background medium which screens all electrostatic interactions, as we will now show. In order to determine the electrostatic potential  $\phi$  from a charge distribution in space, one can use Poisson's equation connecting the electrostatic potential and the local charge density  $c = \sum_i c_i$  of all the charge species  $i$  in the medium.

$$\nabla^2 \phi(\mathbf{r}) = -\frac{4\pi}{\epsilon} c(\mathbf{r}), \quad (1.1.6)$$

and connect it with the Boltzmann distribution

$$c_i = c_i^0 e^{z_i \phi(\mathbf{r})/k_B T} \quad (1.1.7)$$

for finding a concentration  $c_i$  of charged species with valency  $z_i$  at a point in space. This Poisson-Boltzmann equation can be linearised to obtain the famous Debye-Hückel (DH) equation [4] and further simplified for monovalent salts ( $c_1 = c_2 = c_0$ ,  $z_1 = -z_2 = 1$ ):

$$\nabla^2 \phi = \lambda_D^{-2} \phi(\mathbf{r}) \quad (1.1.8)$$

where  $\lambda_D$  is the Debye-Hückel screening length:

$$\lambda_D = \sqrt{\frac{\epsilon k_B T}{8\pi e^2 c_0}}. \quad (1.1.9)$$

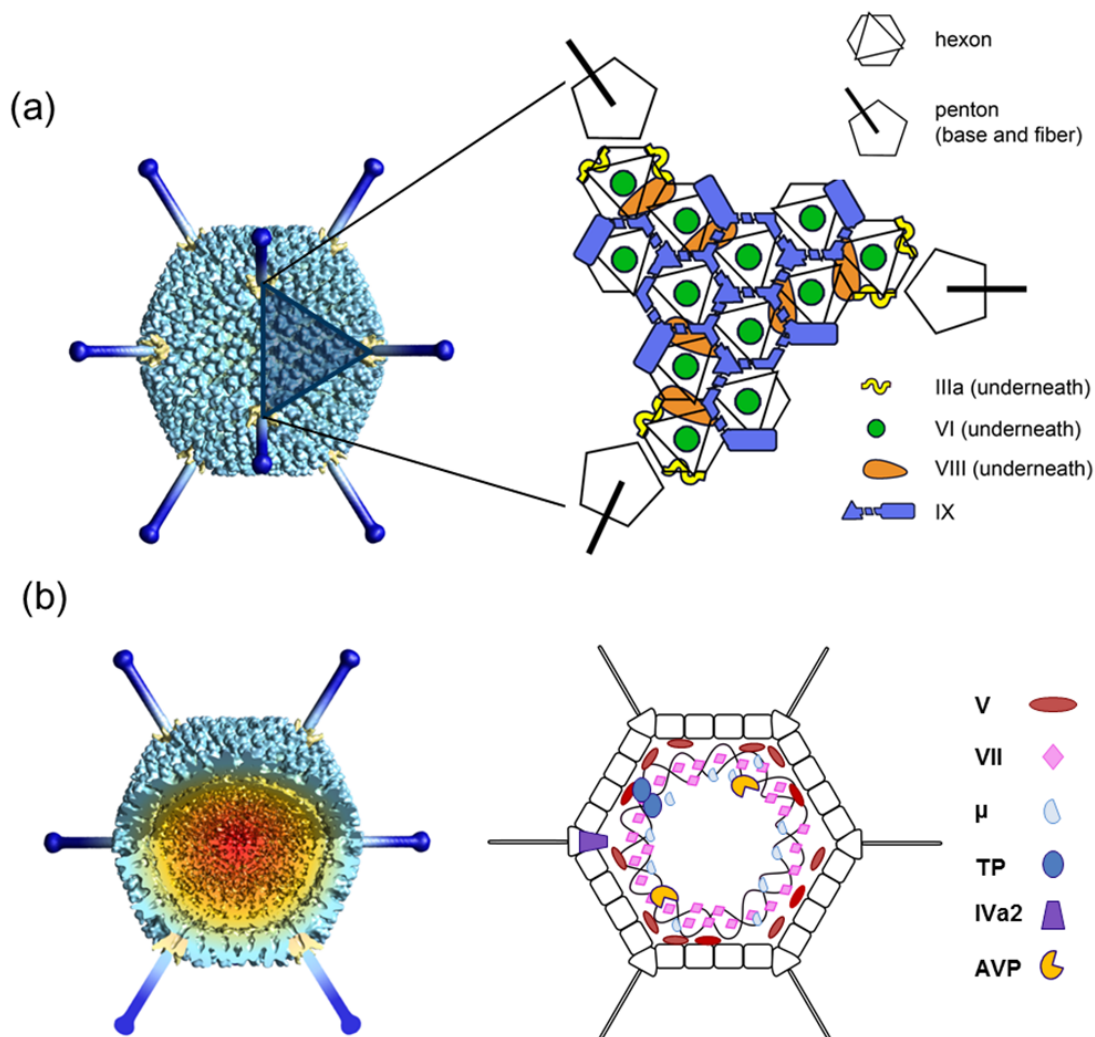
The solution of this equation for a point charge gives the most important result of the DH equation – counterions screen the electrostatic potential so that it acquires the Yukawa form [4]:

$$\phi(r) \sim \frac{1}{r\kappa_D} e^{-\kappa_D r}. \quad (1.1.10)$$

The result is that all electrostatic interactions at length scales smaller than the screening length  $\kappa$  are practically unchanged, while interactions are exponentially suppressed at larger length scales.



## 1.1.3 Virology

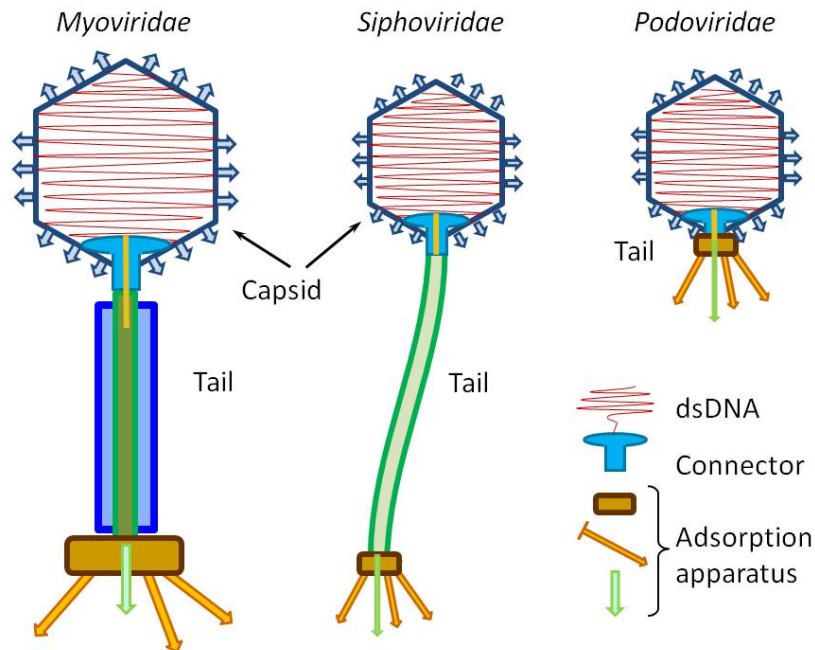


**Figure 1.1: Sketch of an virus with icosahedral symmetry – an adenovirus.** a) Icosahedral shell organization with a model showing the roles of different capsid proteins. (b) Core structure of the shell from cryo-EM with a schematic representation of the core contents, including DNA and core proteins (V, VII and  $\mu$ ). Taken from Ref. [5].

As the basic application of our work lies in better understanding how viruses work, we shall present a physicist’s caricature of what a virus is. A virus consists of at least a genome, either DNA or RNA, and a container made out of protein protecting it (the capsid). The taxonomy of viruses is diverse, e.g. one can base it on the type of nucleic acid or the shape of the container [6]. We will focus only on two groups of viruses relevant in our work: a) tailed bacteriophages, and b) viruses with icosahedral symmetry. In discussing icosahedral viruses, we will additionally limit ourselves to types that are known to use condensing agents to compact their genome in the capsid (adenovirus, polyomavirus).

The basic building block of viral capsids with icosahedral symmetry are repeating proteins which triangulate the viral capsids. It is into this capsid that the genome is packed, either during assembly of the capsid shell or afterwards via cellular motors, and

onto which different protein receptors are located [6]. A typical example of icosahedral capsid structure can be found in the adenovirus (Fig. 1.1).



**Figure 1.2: A tailed bacteriophage.** Sketches of tailed bacteriophages with marked capsids containing the genome (double stranded DNA) and tails of various lengths. Taken from Ref. [7].

Bacteriophages are viruses which specifically attack bacteria. Their structure is simple (Fig. 1.2), they have a capsid of either icosahedral or prolate shape protecting their genome, a protein tail (in most cases), and receptors enabling the ejection of their genome inside bacteria. Tailed phages contain dsDNA and represent 96% of all known bacteriophages [7]. Tailed bacteriophages are represented with great variety: typical tail lengths are in the ranges 10-800 nm, capsids sizes 30-160 and genome lengths in the range of 14 – 498 kbp (thousands of base pairs) [8].

## 1.2 Overview

Packing of nanoparticles, synthetic polyelectrolytes and genomes in a confined space is fundamentally and technologically important. From one aspect, it is a study of the interactions in a macromolecular complex with molecular signatures of "life", thus important for understanding life processes [9]. On the other side, viruses are evolved nano-machines with a "purpose" to penetrate the cellular membrane for delivery of their "cargo", the genome [6]. Any modifications to this cargo, e.g. by using a designed macromolecule or nano-particle, carries a significant potential for applications in personalised medicine (targeted drug or genome delivery). Still, there is a lack of fundamental understanding of

the state of DNA in a viral capsid at different densities and thus the physics behind the viral packaging of genetic material and subsequent infection process [9, 10].

Although the problem of packing a polymer, either flexible or semi-flexible, into confined spaces of different geometries (slits, cavities) has been extensively studied [2], the intricacies brought about by a three-dimensional cavity have only been approached recently [11]. Depending on the flexibility of the polymer, its degree of confinement, density and any (excluded volume) interactions a rich phase diagram is expected [12]. The phase diagram is expected to be traversed by DNA during its ejection from a bacteriophage [10].

The understanding of DNA phase states and the polymorphism of DNA conformation in confinement [10] requires an approach combining and bridging different theoretical models. The isotropic (disordered) state at low densities is known to have a phase transition to a liquid crystal state. This happens in bulk liquid crystals due to the inherent anisotropy in the excluded volume of long polymers [13]. Still, experiments find intricacies not expected in bulk encouraging new approaches to modelling confined liquid crystals [14, 15]. The liquid crystal phase covers roughly one third of the density diagram [10, 12] and ends in a dense condensed phase [16, 17]. It is known that the formation of dense phases of DNA (condensed DNA) depends on a combination of electrostatic and hydrophobic interactions between different segments of the negatively charged DNA backbone in a background of mono- and multi-valent counterions [9, 17]. In this sense the densely packed DNA in viruses, e.g. bacteriophages, is a relatively simple case well suited to study confined DNA phases without the complications of a cellular (*in vivo*) environment [17].

An open question regarding DNA packaged into viruses is what are the mechanisms of its release into the cell. The explanation of all the relevant (thermodynamical) forces guiding the ejection of dsDNA from tailed bacteriophages into bacterial cells is, in spite of 50 years of research, still missing [18, 19]. The ejection starts as a release of DNA from the fully packed viral capsid (protein coating of the bacteriophage). The DNA is packed to extreme densities and exerts a pressure of 25 – 100 atm on the capsid [9, 20]. Models developed and tested *in vitro* (see [18] and references therein) predict that the ejecting force resulting from even such a dense packing is insufficient to completely transfer the DNA into the cell interior. Although cells have smaller turgor pressures than fully packed bacteriophages [21], the ejecting force (and pressure) in the capsid drops sharply as it empties [22]. A recent single molecule Hershey-Chase experiment [23] hints that *in vivo* ejection is controlled not by the amount of DNA left inside the capsid but by the amount ejected into the cell. This means that once the pressure built-up in the bacteriophage is spent on the DNA ejection, a cellular mechanism takes over. There have been various proposed mechanisms for completion of the ejection but it appears that none of these models give a definite answer while experiments suggest a coexistence of many different mechanisms.

Packing of DNA in confinement has been studied both experimentally and theoretically for some time, but what happens when interacting proteins are added to the mix is unknown. There are reasons to believe that some viral proteins may be evolutionary adapted to encapsidate nucleic acids [24]. This may be relevant for encapsidation of synthetic cargo [25]. A good example are adenoviruses which package dsDNA into the viral capsid accompanied by viral DNA-binding proteins, which may help in charge neutralization and condensation [5]. There is no high resolution structural data for any of these proteins, and the exact conformation of this mixture of polymer (DNA) and nano-particles (proteins) is still not known. The DNA, being partially bound or neutralized by the proteins, may be modelled as an effective medium for the DNA-binding proteins. Here the effects of the viral capsid may induce a degree of ordering in the distribution of these proteins. Similar effects have been recently seen in the chromatin organization in capsids of SV40 virus where the coupling between packing proteins and the confinement influences the state of the packaged DNA material [26].

The crowded environment of polymers/polyelectrolytes and nanoparticles is also interesting from the biological viewpoint as it parallels with the crowded cell interior. Crowding has been shown to lead to many different phenomena, e.g. renormalization of bare interactions [27] and anomalous diffusion [28]. It is known to be relevant for both nucleic acid compaction [29, 30, 31, 32] and protein folding [33]. But the exact nature of this mechanism and the repercussions it may have on our understanding of cellular functions are not known.

## 1.3 Outline of work

Chapter 2 "Compacted DNA" covers the topic of condensed and confined DNA. We explain parts of the phase diagram for confined DNA in 3D confinement and apply it to understanding DNA phases in bacteriophages, the interplay of various free energy terms and the ejection process from a phage. We contribute to understanding the problem of in vivo DNA ejection from bacteriophages by giving a thermodynamic model explaining ejection into Gram-positive bacteria.

Chapter 3 "Nucleic acids and condensing proteins in confinement" deals with the subject of how condensing proteins interact with nucleic acids in a confined environment. We will discuss basics of packaging nucleic acids in confinement, and how the confinement and crowding influence elementary interactions. We apply this models to explain the organization of viral capsid interiors containing condensing core proteins and DNA.



## CHAPTER 2

# Compacted DNA

A part of the work presented in this chapter has been previously published in:

[34] S. Marion and A. Šiber, “Ejecting Phage DNA against Cellular Turgor Pressure,” *Biophys. J.* **107**,1924–1929 (2014).

DNA is considered to be compacted if its spatial extent has been reduced due to an external influence. In its native state, when located inside a physiological solution, dsDNA has the conformation of a self-avoiding walk in space [35], i.e. a random coil. Two major compaction mechanisms which will be considered are: a) condensation of DNA in a dense phase, and b) confinement of DNA inside a small volume. Specifically, we will consider confinements into three-dimensional cavities [11]. The motivation for considering states of compacted DNA is to understand the process of genome ejection from bacteriophages into cells but also the packing of DNA in viruses. During viral ejection from a bacteriophage into a cell, the DNA moves from one container into another, from a viral capsid into a crowded cellular interior. We will first try to understand how these different compacted states behave, before combining our understanding to explain *in vivo* ejection from a bacteriophage. Although *in vitro* ejection has been successfully resolved [36, 37, 38], the *in vivo* case with all its biological intricacies is still lacking a solid explanation [18].

## 2.1 Condensation of DNA

Condensation of DNA is known to take place in various conditions [39]. In a dilute solution, DNA-DNA interactions are strongly repulsive resulting in DNA performing a self avoiding walk in space, producing a so called random coil. But, if the strong electrostatic repulsive interaction between parts of the DNA is suppressed, the DNA as any polymer might condense in globules [2], or specifically for DNA, even in dense hexagonal phases [40, 41]. The hexagonal phases of DNA are known to be induced by multivalent cations, which mediate a net attractive interaction between the DNA strands. Other methods for

condensation include neutral polymers, where by using the osmotic pressure produced by Polyethylene glycol (PEG) one can measure the forces resisting the osmotic compaction, and simultaneously measure the unit cell of the hexagonal packings [40, 40]. Examples of multivalent cations that act as condensing agents include cobalt hexamine (3+) [40], and compounds present *in vivo* like polylysine [30] and spermidine (3+) [29].

Condensation of DNA with multivalent ions has been demonstrated in controlled biological conditions. Electron microscopy shows that DNA collapses in globules, toroids and rods, with toroids being a preferable state in biological conditions (See references in [42] and [41]). It was also shown that in the presence of condensing agents like spermidine, bacteriophages eject DNA which forms toroidal condensates, even when confined in liposomes [43]. The interior of the cell is known to be a crowded environment with DNA condensing agents, both causing condensation of DNA [29, 30, 31, 32]. Thus, we study how DNA condensation can be modelled using the continuum model of Ubbink and Odijk [16, 42] in order to understand condensed phases of DNA, especially those found in cells, as a basis for explaining *in vivo* ejection of DNA from bacteriophages.

### 2.1.1 Ubbink-Odijk continuum model

The continuum model of Ubbink and Odijk (U-O) [16, 42] examines the free energy of a DNA condensate in a dense hexagonal phase from the three-dimensional shape of the condensate. The model treats the free energy of the condensate  $F = F_v + F_s + F_b$  as having three distinct contributions coming from: the bulk contribution  $F_v$  proportional to the volume  $V$ , the surface free energy contribution  $F_s$  and the elastic energy contribution  $F_b$  due to the bending of the "DNA bundle". The resulting problem is to obtain the shape of the condensate that corresponds to the free energy minimum arising from an interplay of surface and bending effects.

In our consideration, we limit our modelling to systems with rotational symmetry around the  $z$ -axis. The total free energy  $F_{tor}$  is minimal when the condensate has the shape of a torus [16, 37, 42], but there are other possibilities like rods [30] which are not relevant in physiological conditions [44]. We will optimize the shape of this condensate by finding the closed curve which produces the final surface of revolution. The free energy of the DNA condensate will be given as a functional  $F[C]$  of the curve  $C$  which defines this cross section (Fig. 2.1). Numerically, the curve  $C$  is treated as an ordered set of  $N_0$  points  $\{r_i, h_i\}$ .

We will obtain the free energy per base pair of the condensed DNA as a function of its length  $L = V/A_0$ . Here  $V$  is the condensate volume and  $A_0$  the area of the base of the hexagonal lattice  $A_0 = \sqrt{3}d^2/2$  of DNA strands and  $d$  the closest neighbour distance between DNA strands. In this section we will take the value  $d = 4$  nm as in the work of

Ubbink and Odijk [16, 42]. Other parameters are the base pair length  $L_{bp} = 0.34$  nm and the persistence length  $L_p = 50$  nm valid for physiological monovalent salt concentrations [45].

The **volume contribution** to the free energy is given as a negative (attractive) part proportional to the DNA volume  $-\gamma V$  or when integrated over a rotational body at a distance  $r$  from the axis of symmetry with a unit area of  $drdh$

$$F_v = -k_B T \gamma \iint_{\{r,h\} \text{ in } C} 2\pi r dr dh. \quad (2.1.1)$$

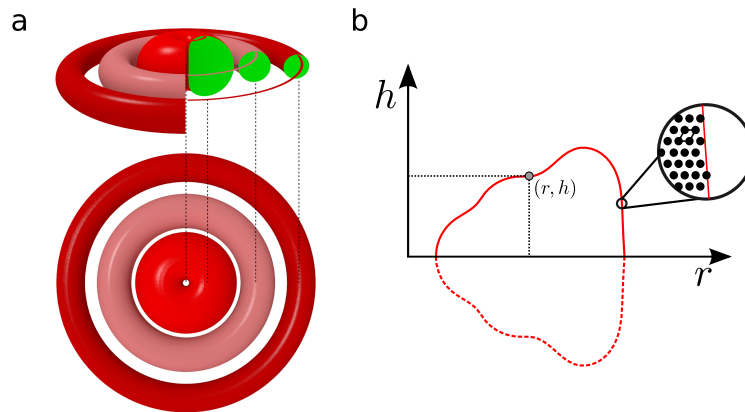
where  $\gamma$  is an effective condensing pressure,  $r$  is the radial coordinate and  $h$  the rotation axis coordinate in the cylindrical coordinate system (Fig. 2.1b). The **surface contribution** is given by a positive contribution proportional to the surface area  $\sigma S$  represented by the integral

$$F_s = k_B T \sigma \oint_C 2\pi r dl, \quad (2.1.2)$$

where  $\sigma$  is the surface tension. The **total bending energy** is given as  $L_p/2R^2$  per unit length of a circular loop of DNA wound at a radius  $R$  [3] with persistence length  $L_p \approx 50$  nm in physiological conditions [45]. The total binding energy is obtained from an integral of thin circular fibres over the cross section area with a cylindrical symmetry:

$$F_b = k_B T \frac{L_p L}{2V} \iint_{\{r,h\} \text{ in } C} \frac{2\pi r}{r^2} dr dh. \quad (2.1.3)$$

Before proceeding to minimizing the resulting free energy, we must examine the physical content of the constants  $\gamma$  and  $\sigma$ .



**Figure 2.1:** a) Cross-sectional areas (in green) for various DNA condensates of the same volume, b) Coordinate system used assuming cylindrical symmetry around the axis of rotation for the torus and inversion symmetry for axis  $h$ . Slice shows the hexagonal DNA ordering.



## Non-dimensionalisation of the Ubbink-Odijk equation

In the U-O model a non-dimensional parameter  $\alpha$  [16] is used to quantify the relative impact of the surface versus bending free energy contributions. It is obtained by scaling all the spatial variables in  $F = F_b + F_s$ :

$$\tilde{F} = \int_{R-r}^{R+r} \frac{H(X)}{X} dX + \alpha \int_C X dL(x) \quad (2.1.4)$$

where  $\alpha$  is defined as

$$\alpha = \frac{4\sigma V^{\frac{4}{3}}}{L_p L} = \frac{4\sigma A_0^{\frac{4}{3}} (L_{bp} N_{bp})^{\frac{1}{3}}}{L_p}. \quad (2.1.5)$$

The non-dimensional parameter  $\alpha$  will define the shape of the condensate. Because  $\alpha_0$  is a function of the length of DNA in the condensate, we will define  $\alpha_0 = \alpha(N_{bp}^0)$  as the  $\alpha$  when all the DNA base pairs  $N_{bp}^0$ , e.g. from a virus, are in the condensate. The shape of the condensate, and not just its volume, will depend on the amount of DNA  $L = L_{bp} N_{bp}$  in the condensate:

$$\alpha(N_{bp}) = \left( \frac{N_{bp}}{N_{bp}^0} \right)^{\frac{1}{3}} \alpha_0. \quad (2.1.6)$$

When optimising the shape of the toroid, we will change the length of DNA but keep the values of the surface and volume energy density constant, in essence we start from a certain  $\alpha = \alpha_0$  which is then reduced if the quantity of condensed DNA is decreased. Because  $\alpha_0$  is a coupling parameter of the total surface and total bending energy of a toroid, by changing the length we will vary the ratio of these contributions.

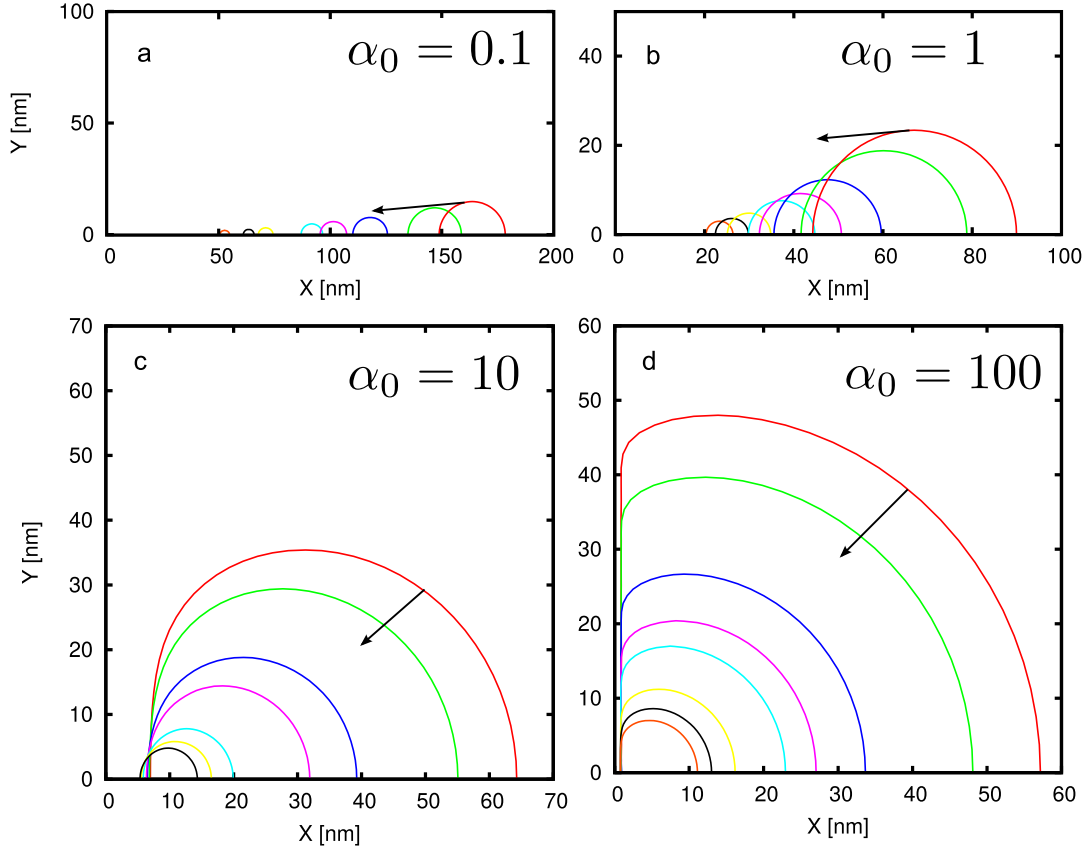
### Connection between $\sigma$ and $\gamma$

The volume contribution of the free energy divided by the volume  $F_v/V$  can be obtained by summing the contribution  $f_0$  (a calculation of  $f_0$  is given in Ref. [42]) from all the pairs of interacting strands times the density of base pairs:

$$-\frac{F_v}{Vk_B T} = 6f_0 \frac{N_{bp}}{LA_0} = \frac{6f_0}{L_{bp} A_0} = \gamma. \quad (2.1.7)$$

For the surface contribution, the free energy divided by the surface area will be proportional to the number of neighbours ( $\sim 2$  for the hexagonal lattice) lost due to the reduced coordination of the surface and the number of base pairs on the surface divided by the surface area:

$$\frac{F_s}{Sk_B T} = \frac{2f_0 N_{bp}^{2D}}{Ld} = \frac{2f_0}{L_{bp} d} = \sigma. \quad (2.1.8)$$



**Figure 2.2:** Cross sections of the toroid (only upper halves shown) for different starting  $\alpha_0$ . Each sub-figure shows the effect of reducing the length (volume) of the condensed DNA on the shape while keeping  $\sigma$  (and  $\gamma$ ) constant. Arrows show the direction of reducing the total condensate volume. Take note of different scales on the  $x$  and  $y$  axes of the sub-figures.

It follows that  $\sigma$  and  $\gamma$  are related as:

$$\gamma = \frac{2\sqrt{3}}{d}\sigma. \quad (2.1.9)$$

We can now represent the volume free energy contribution via the surface (tension) constant  $\sigma$

$$\frac{F_v}{k_B T N_{bp}} = \frac{1}{N_{bp}} \gamma V = \frac{1}{N_{bp}} \left( \frac{2\sqrt{3}}{d} \sigma \right) (A_0 L_{bp} N_{bp}) = 3d L_{bp} \sigma \quad (2.1.10)$$

or vice versa.

## 2.1.2 Optimizing the DNA condensate shapes

In order to obtain the optimal shapes for DNA we minimize the free energy function with contributions from Eq. (2.1.2), (2.1.3) and (2.1.10) to obtain

$$\frac{F_{tor}}{k_B T} = \frac{L_p L}{2V} \iint_{\{r,h\} \text{ in } C} \frac{2\pi r}{r^2} dr dh + \sigma \oint_C 2\pi r dl - \frac{2\sqrt{3}}{d} \sigma V, \quad (2.1.11)$$

for a fixed DNA length  $L = V/A_0$ . This is achieved via numerical optimization of  $F_{tor}(C)$ , i.e. by a minimization of a multidimensional function  $F_{tor}(r_i, h_i)$  (Fig. 2.1). The optimization was performed in Python using the Numba JIT compiler [46] with 20 coordinate pairs  $(r_i, h_i)$  representing the upper half of the base of the rotational body.

A representative example of minimization of the free energy given by Eq. (2.1.11) for four starting  $\alpha_0$  with different numbers of DNA base pairs  $N_{bp}$  is shown on Fig. 2.2). The figure represents how a DNA condensate's shape changes as the amount of DNA becomes smaller. When  $\alpha_0$  is small (Fig. 2.2a,b), tori with shorter DNA present a tendency to only reduce the toroid major axis while maintaining a circular cross section. For larger  $\alpha_0$  (Fig. 2.2c,d), tori no longer have a circular cross section. As the bending contribution is smaller for large  $\alpha(L)$ , the condensate collapses in a globule with a small cylindrical "void" near  $r = 0$  due to the bending energy diverging there. A crossover between these two behaviours (the "thin torus" and "fat sphere") occurs when  $\alpha \sim 10$ .

When the bending energy dominates (see Fig. 2.2a) we expect a **thin torus** with a circular cross section whose free energy can be approximated using a variational approach. We first approximate the free energy (without the volume contribution) as:

$$\frac{F_{slim}}{k_B T} = \frac{L_p L}{2V} \frac{V}{R^2} + \sigma S. \quad (2.1.12)$$

To obtain the major radius  $R$  of the torus, we minimize the free energy with respect to  $R$  while keeping the volume  $V$  fixed:

$$R = \left( \frac{L_p}{\sqrt{2}\pi A_0 \sigma} \right)^{\frac{2}{5}} V^{\frac{1}{5}}. \quad (2.1.13)$$

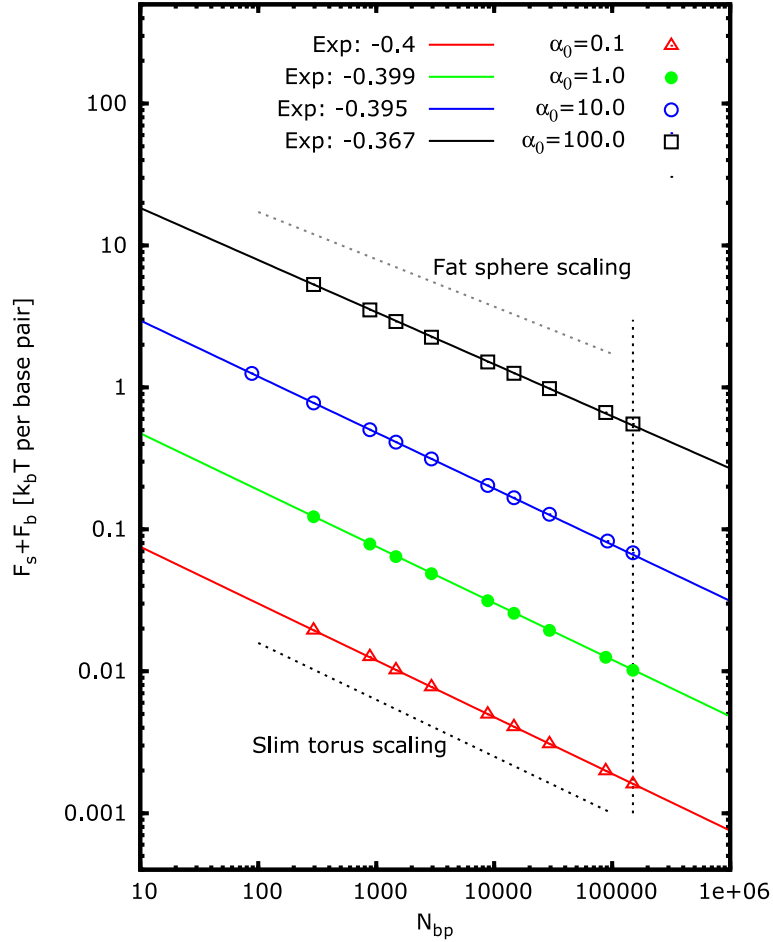
Now, the total free energy per base pair is obtained as:

$$\frac{F_{slim}}{k_B T N_{bp}} = 5 \left( \frac{L_p \sigma^4 \pi^4 A_0^2 L_{bp}^3}{8} \right)^{\frac{1}{5}} N_{bp}^{-\frac{2}{5}}. \quad (2.1.14)$$

In the **"fat sphere" limit** (see Fig. 2.2d) the bending contribution is considered

negligible, so only the surface energy has a noticeable contribution:

$$\frac{F_{fat}}{k_B T N_{bp}} = \sqrt[3]{4\pi\sigma} (3A_0 L_{bp})^{\frac{2}{3}} N_{bp}^{-\frac{1}{3}}. \quad (2.1.15)$$



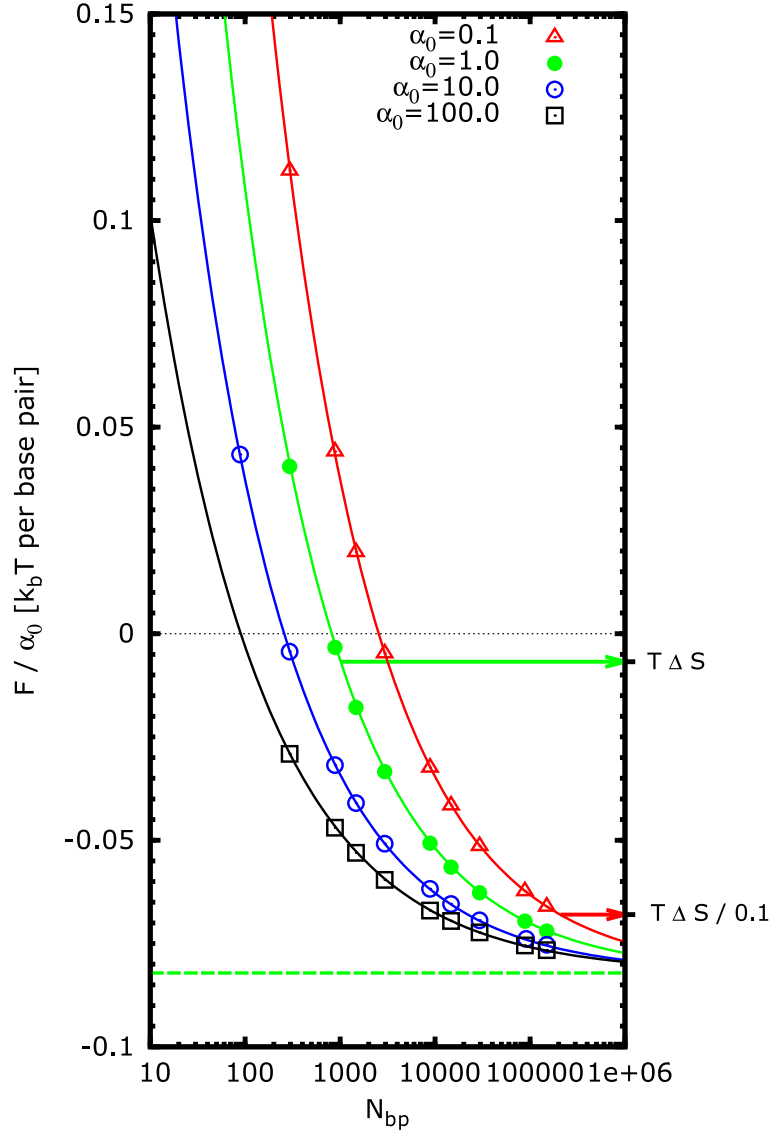
**Figure 2.3:** Free energy per base pair in the U-O model as a function of the number of base pairs  $N_{bp}$  in the condensate; only surface and bending contributions are included.

The points are values obtained numerically while the lines are power law fits. The vertical dashed line corresponds to the length for which the U-O surface-bending coupling constant  $\alpha_0$  is defined for a DNA of 169 kbp (T4 DNA). Different  $\alpha$  parameters were obtained by fixing the DNA persistence length at  $L_p = 50$  nm and varying  $\sigma$  (and thus  $\gamma$ ) using Eq. (2.1.5). The points on this curve correspond to the starting shapes in Fig. 2.2 The two limiting scaling exponents are shown.

We calculate the sum of surface and bending energies for different  $\alpha$  in Fig. 2.3. The dependence of the free energy on the number of base pairs with the included surface and bending terms is between the limiting power laws of the thin torus and the fat sphere. We note that both  $\alpha_0 = 0.1$  and  $\alpha_0 = 1$  curves are in the thin torus regimes, while the  $\alpha_0 = 100$  curve is approaching the fat sphere regime. In the fat sphere regime there is an exclusion cylinder at  $x = 0$  due to a divergence in the bending free energy.

### 2.1.3 Minimal packed length for condensate stability

For DNA to be in the condensed state, the total free energy  $F_{tor}$  must be negative. When we add the volume contribution to the previously calculated surface and bending contributions (Fig. 2.3) we see that all  $\alpha_0$ -curves follow the same trend (see Fig. 2.4): there exists a minimal length for stable condensation ( $F < 0$ ).



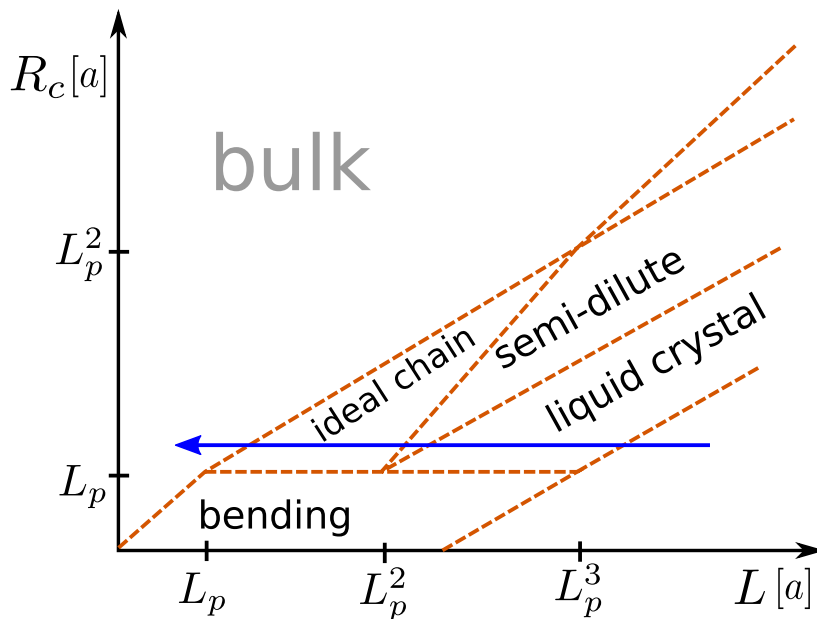
**Figure 2.4:** Free energy per base pair in the U-O model as a function of the number of base pairs  $N_{bp}$  in the condensate; the values have been scaled by division with  $\alpha_0$  (at maximal packing). Thus, the  $y$ -axis corresponds directly to the curve  $\alpha_0 = 1$ . The horizontal dashed line corresponds to the asymptote for extremely large  $N_{bp}$ . The entropy contribution  $T\Delta S$  for  $\alpha_0 = 1$ , and the scaled  $T\Delta S/\alpha_0$  corresponding to the  $\alpha_0 = 0.1$  curve, are marked on the left  $y$ -axis with arrows. Parameters are the same as in Fig. 2.3.

In the fat sphere limit, the minimal base pair quantity for stable packaging can be obtained from Eq. (2.1.15) and (2.1.10). Thus, without bending the Ubbink-Odijk model requires at least  $N_{min} \approx 40$  base pairs. In the thin torus limit for the  $\alpha_0 = 0.1$  curve, we

obtain from Eq. (2.1.14) and (2.1.10) that at least  $N_{min} \approx 2500$  base pairs are needed for a stable condensate. This matches the interpolated numerical value obtained from Fig. 2.4.

We now analyse what happens if entropy is treated separately from the free energy densities included in the model. The entropy penalty for condensing a semi-flexible random coil can be approximated by using the free energy from undulations of a semi-flexible polymer [47] as  $\Delta S/N_{bp} = L_{bp}/L_p = 0.0068$ . Comparing this entropy penalty with the values given in the model we find that the condensation in the  $\alpha_0 = 0.1$  case is not stable, while for the  $\alpha_0 = 1$  case it increases the minimum number of base pairs by roughly 10%, and is negligible for higher  $\alpha_0$  (note the markings on the right  $y$ -axis at Fig. 2.4). Bloomfield [48] suggests that the minimal length for DNA condensation is between 150 and 400 base pairs for a dilute DNA solution. Also, he gives a value for the condensation energy on the order of  $-0.07 k_B T$  per bp corresponding to our  $\alpha_0 = 1$  curve, consistent with the previous discussion and other similar approaches [49].

## 2.2 Phases of confined DNA

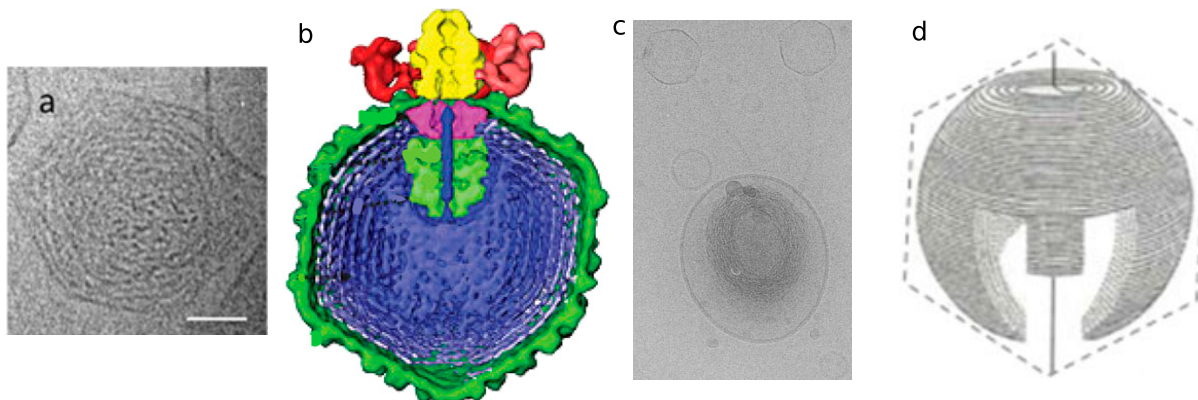


**Figure 2.5: Phase diagram of a confined polymer.** Different regimes for a confined semi-flexible polymer are shown. The  $x$ -axis shows the total length of polymer inside confinement with different persistence lengths  $L_p$  marked (for details see Ref. [12]). Note, the polymer monomer size  $a$  is set  $a = 1$  so that both the capsid length and persistence length are given as non-dimensional quantities in the units of the monomer size. The  $y$ -axis shows the radius  $R_c$  of the confinement. The arrow (blue) shows the approximate position and direction of the proposed ejection path for a bacteriophage (see text for details). This figure has been adapted from Ref. [12].

Confined DNA differs from bulk DNA solutions. Depending on the interplay of the DNA persistence length  $L_p$ , capsid confinement radius  $R_c$  and the total length  $L$  of encapsidated DNA we expect several regimes. When scaling arguments are applied to a DNA molecule confined in a capsid of radius  $R_c$  one can expect four distinct phases [12] before a dense crystalline-like packing is reached at high densities (Fig. 2.5). When the confinement size is smaller than the persistence length of DNA, the DNA molecule will be stiff, and will be pushed against the interior of the confinement. We will call this the bending dominated regime. In the opposite case, although the DNA will be bent inside, this bending energy will not dominate over excluded volume interactions. Within this regime, we expect a confined ideal polymer chain followed by a semi-dilute regime and finally a liquid crystal state [12]. At high density, in all the regimes the DNA will enter a dense hexagonal state [40].

The simplest regime is the bulk regime, as here the dimension of the DNA random walk in space is smaller than the size of the confinement, and requires no special treatment. After the DNA coil spatial extend starts to probe the confinement, the DNA enters the ideal chain regime ("dilute") in confinement as the DNA can not significantly interact with the capsid, but it determines the local concentration of DNA. The transition into this regime happens at  $\sqrt{L_{bp}N_0L_p} \approx 2R_c$  [50] which for  $L_p = 50$  nm and  $R_c = 30$  nm amounts to  $N_0 = 200$  bp. At larger lengths the DNA enters the confined ideal chain regime – the dominant free energy contribution is a confinement penalty coming from bending. After the ideal chain the DNA enters the (mean field) semi-dilute regime – here the bending contributions are overcome by excluded volume self-interactions. At even higher densities, the excluded volume and bending influence the local director, the tangent on the curve of the DNA, to become correlated with that of the nearest neighbours to facilitate denser packing. The DNA locally orders into a liquid crystal phase [13, 51]. It is known that the packing progresses firstly through a nematic then cholesteric regime and ending in a dense hexagonal packing manifest as an inverse spool (see Fig. 2.6) [10]. This inverse spool has an axis of symmetry left from packing of DNA through a portal in the viral capsid. The completely packed DNA strand has a cylindrical symmetry around this axis passing through the tail and center of the capsid.

The bending regime is the least studied, as in physiological conditions the persistence length of DNA is just above the threshold for the transition in a bending dominated regime. Many experiments find the DNA inside the capsid as being in a liquid crystalline state [10]. The transition from an isotropic (disordered) to liquid crystal (ordered) state seems to happen at about 30% packing in the case of phage, and the transition in a hexagonal state at about 70% [10]. But, if the temperature of the partially filled viral capsid is reduced by 20 degrees, the DNA adopts a conformation of an inverse spool (Fig. 2.6d), the signature of a bending dominated regime [10]. So, it seems that packed DNA



**Figure 2.6: Examples of different DNA packing regimes.** a) Cholesteric liquid crystal DNA ordering inside partially filled capsids. Figure adapted from Ref. [10], b) Reconstruction of dense ordered organization inside epsilon15 phage capsids obtained from icosahedrally averaged electron micrographs. Figure adapted from Ref. [52], c) Micrograph of two bacteriophages ejecting DNA inside a liposome in the presence of the spermidine condensing agent. The resulting DNA torus has hexagonal ordering (as in Fig. 2.2). Figure adapted from Ref. [43], d) DNA inverse spool. The DNA enters the capsid portal (tail), and packs as a spool with cylindrical symmetry starting from the outer layers inwards. Figure adapted from Ref. [53].

is just outside the bending dominated regime, which is initiated by enhancement in the stiffness at low temperatures. The ejection of DNA from a bacteriophage capsid is thus expected to follow through several packing regimes (representing a horizontal path through the phase diagram on Fig. 2.5). *In vivo*, it will progress from a high density hexagonal phase, through the liquid crystal, and finally the semi-dilute state before almost none is inside and the force of ejection drops to zero. In the bending regime, ejection is expected to be simpler, with only a change of the ordering and packing organization as the length of the encapsidated DNA is reduced.

In what follows, we will first discuss how the bending energy can be calculated in both the bending-dominated and the liquid crystal regimes, before continuing to study the free energy of packed DNA in the ideal chain, and the liquid crystal regimes.

## 2.3 Bending of confined DNA

The bending of DNA is especially relevant in viral capsids, as the typical persistence length of the DNA ( $L_p \approx 50$  nm) is comparable to the size of viral capsids ( $R_c \sim 30 - 50$  nm) [54]. The bending energy is affected by the confinement, effectively limiting the maximal radius of curvature of a DNA contour to the size of the capsid  $R_c$ . In this regime the best approach is to model the DNA as a worm-like chain with a dominant contribution from bending. The bending energy of a single strand of DNA can be expressed as an



integral of its local radius of curvature  $R$  over the full contour length [3]

$$F_b = \frac{1}{2}L_p k_B T \int \frac{1}{R(\mathbf{r})^2} dl(\mathbf{r}) \quad (2.3.1)$$

where  $dl$  follows the contour. The bending energy in confinement becomes large for tightly wound DNA as the minimal bending energy is limited with the capsid radius  $R_c$  as  $\sim R_c^{-2}$ . When the Kuhn length of DNA  $2L_p$ , which represents one step in the random walk of the DNA, is larger than the diameter of confinement  $2R_c$  the DNA random walk becomes frozen by the confinement [12]. We can see this if we compare the thermal energy available to one persistence length segment to the corresponding bending contribution. DNA inside a confining sphere of radii  $R_c$  has a minimal free energy when its radius of curvature is equal to the radius of the confinement. A length  $L_p$  of such DNA would have a free energy of  $\sim \frac{1}{2}k_B T L_p L_p / R_c^2$ , comparable to the thermal energy of  $\sim \frac{1}{2}k_B T$  per degree of freedom. The conformation of such DNA is effectively dominated by bending.

First, we shall discuss a simple mean field bending model, that is applicable if the DNA is not tightly bent (not in the bending regime in Fig. 2.5), as a basis for future discussions. Afterwards, we will discuss an effective loop model for tightly bent ordered DNA [55] and extend it to include also transition to a disordered state [56].

### 2.3.1 Mean-field bending model

The bending energy of a DNA strand in a spherical confinement of volume  $V_c$  can be written as

$$\frac{F_b}{k_b T} = \frac{L_p}{2} \int_{\text{sphere}} \rho(\mathbf{r}) K^2(\mathbf{r}) d^3 \mathbf{r} \quad (2.3.2)$$

where  $K(\mathbf{r}) = 1/R(\mathbf{r})$  is the local curvature, the integration is limited to the inside of the sphere, and  $\rho(\mathbf{r})$  is the local density of DNA. We will make an approximation of a mean local bending energy, such that there exist an expected value of the bending  $\langle K^2 \rangle$  in place of the true  $K^2(\mathbf{r})$  together with the assumption that the density  $\rho = L_{bp} N_{bp} / V_c$  inside the confining volume  $V_c$  is homogeneous.

We know that

$$K(\mathbf{r}) = |\nabla_s \mathbf{u}(\mathbf{r})| \quad (2.3.3)$$

where  $\nabla_s$  is applied along the curve of the DNA strand, and  $\mathbf{u}(\mathbf{r})$  is the tangent vector for a DNA strand at position  $\mathbf{r}$ .  $\mathbf{u}(\mathbf{r})$  can be decomposed in a spherical coordinate system with its center coinciding with the center of the spherical confinement:

$$\mathbf{u} = u_r \hat{r} + u_\theta \hat{\theta} + u_\phi \hat{\phi}. \quad (2.3.4)$$

As we expect rotational symmetry inside the confinement, we can simplify the calculation

of the local curvature. We will consider a simplified case where there is no  $\theta$  or  $\phi$  dependence in the projections of  $\mathbf{u}$  on  $\hat{r}, \hat{\theta}, \hat{\phi}$ ; only a dependence on a order parameter, and possibly a dependence of  $u_r$  on  $r$ . Here, we neglect correlations between neighbouring parts of the strand inside the capsid, except radial correlations due to boundary conditions on the surface of the sphere. Thus, we can imagine the capsid consisting of infinitesimal boxes with a given probability distribution  $f(\mathbf{u})$  for finding a piece of DNA in a given direction  $\mathbf{u}$ ; this probability distribution can be either isotropic or give a measure of ordering.

We calculate the changes of the unit vectors of the coordinate system as we locally move from one point to the other to obtain the local curvature:

$$K(\mathbf{r}) = \left| \hat{r} \left( \frac{\partial u_r}{\partial r} + \frac{u_\theta}{r} - \frac{u_\phi}{r} \right) + \hat{\theta} \left( \frac{u_r}{r} - \frac{\cos \theta}{r \sin \theta} u_\phi \right) + \hat{\phi} \left( \frac{u_r}{r} + \frac{\cos \theta}{r \sin \theta} u_\theta \right) \right|. \quad (2.3.5)$$

Here, we must have  $u_r(r = R_c) = 0$  due to boundary conditions; the strand cannot penetrate the capsid walls. But, to simplify the problem we will neglect the additional effects of the boundary by taking  $u_r$  to have a constant value inside the capsid. This violates the boundary conditions, but is consistent with either an inverse spool or liquid crystal ordering. In the inverse spool the DNA smoothly transitions between layers with a minimal  $u_r$ . If we were to take  $u_r$  then the distribution function  $f_{\mathbf{n}}$  needs to also be dependent on the azimuthal angle and possibly the distance from the center, requiring a more complex model.

From

$$K^2(\mathbf{r}) = \frac{1}{r^2} (\bar{u}_\theta - \bar{u}_\phi)^2 + \frac{1}{r^2} \left( \bar{u}_r + \frac{\cos \theta}{\sin \theta} \bar{u}_\theta \right)^2 + \frac{1}{r^2} \left( \bar{u}_r - \frac{\cos \theta}{\sin \theta} \bar{u}_\phi \right)^2 \quad (2.3.6)$$

after some simplifications and partial integration over  $r$  and  $\phi$  we obtain

$$\frac{F_b}{k_b T N_{bp}} = \frac{3L_p L_{bp}}{8\pi R_c^3} 4\pi R_c \left( (u_\theta - u_\phi)^2 + \int_{\theta_0}^{\frac{\pi}{2}} \left( u_r + \frac{\cos \theta}{\sin \theta} u_\theta \right)^2 \sin \theta d\theta + \int_{\theta_0}^{\frac{\pi}{2}} \left( u_r - \frac{\cos \theta}{\sin \theta} u_\theta \right)^2 \sin \theta d\theta \right) \quad (2.3.7)$$

where the integrals diverge for  $\theta_0 \rightarrow 0$ . This is because the integrals can be interpreted as the winding of loops of increasingly smaller radius around the symmetry axis (analogy to the inverse spool model). But, due to the finite dimensions of the DNA strand we can enter a minimal distance that is achievable. Thus, we introduce a lower cut-off length of  $R_0$  which is much smaller than  $R$  so  $R_c \theta_0 \approx R_0$ . It follows after neglecting higher orders

of  $\theta_0$  ( $\theta_0 \ll 1$ ):

$$\frac{F_b}{k_b T N_{bp}} = \frac{3L_p L_{bp}}{2R_c^2} \left[ (u_\theta - u_\phi)^2 + 2u_r^2 - 2u_r(u_\phi - u_\theta) + (u_\phi^2 + u_\theta^2)(-1 + \log 2 - \log(\theta_0)) \right]. \quad (2.3.8)$$

Here, we identify  $R_0$  as the "correlation hole" in the inverse spool model which should depend on the length of DNA inside. At smaller packed lengths, the DNA will prefer to fill the outer parts of the capsid as filling the center increases the bending energy [9]. Note that the derived formulas are only valid when the density of DNA is homogeneous inside the confinement. Our approximations are not valid at low densities, e.g. near the end of DNA ejection from a bacteriophage.

We approximate the bending energy by the expected value of the bending energy for a given distribution function  $f(\beta)$  such that

$$F_b \rightarrow \langle F_b \rangle = \int F_b f(\beta) d\Omega \quad (2.3.9)$$

The projections of the tangent vector on the spherical coordinate system axis are

$$\begin{aligned} u_r &= \sin \beta \sin \psi \\ u_\phi &= \sin \beta \cos \psi \\ u_\theta &= \cos \beta \end{aligned} \quad (2.3.10)$$

where we have redefined the values of  $\beta$  and  $\psi$  to be in a new spherical coordinate system where  $\hat{z}$  is in the direction of the director field which we take to be  $\hat{\phi}$  (to recover the inverse spool model in the completely ordered state). Now, from the property of the distribution function that  $f(\beta) = f(\pi - \beta)$  and its independence on the azimuthal angle we can show that the expected value of the projections are zero by definition

$$\begin{aligned} \langle u_r \rangle &= 0 \\ \langle u_\phi \rangle &= 0 \\ \langle u_\theta \rangle &= 0 \end{aligned} \quad (2.3.11)$$

and that the mixed terms are also zero

$$\begin{aligned} \langle u_r u_\phi \rangle &= 0 \\ \langle u_r u_\theta \rangle &= 0 \\ \langle u_\theta u_\phi \rangle &= 0 \end{aligned} \quad (2.3.12)$$

Using the preceding equation we can simplify Eq. (2.3.8) to obtain

$$\frac{F_b}{k_b T N_{bp}} = \frac{3L_p L_{bp}}{2R_c^2} \left[ \langle u_\theta^2 \rangle + \langle u_\phi^2 \rangle + 2\langle u_r^2 \rangle + (\langle u_\phi^2 \rangle + \langle u_\theta^2 \rangle) \left( \ln \left( \frac{2R_c}{R_0} \right) - 1 \right) \right]. \quad (2.3.13)$$

By using trigonometric expressions and taking into account that the distribution is not dependent on the azimuthal angle  $\psi$

$$\frac{F_b}{k_b T N_{bp}} = \frac{3L_p L_{bp}}{2R_c^2} \left[ 1 + \frac{1}{2} \langle \sin^2 \beta \rangle + (1 - \frac{1}{2} \langle \sin^2 \beta \rangle) \left( \ln \left( \frac{2R_c}{R_0} \right) - 1 \right) \right] \quad (2.3.14)$$

or by introducing the order parameter from liquid crystal theory  $S[f] = \langle P_2(\cos \beta) \rangle$ :

$$\frac{F_b}{k_b T N_{bp}} = \frac{L_p L_{bp}}{R_c^2} \left[ 1 - S + (1 + \frac{1}{2} S) \ln \left( \frac{2R_c}{R_0} \right) \right] \quad (2.3.15)$$

In the isotropic state  $S = 0$ , so:

$$\frac{F_b}{k_b T N_{bp}} = \frac{L_p L_{bp}}{R_c^2} \left( \ln \left( \frac{2R_c}{R_0} \right) + 1 \right) \quad (2.3.16)$$

and in the fully ordered state  $S = 1$ , so

$$\frac{F_b}{k_b T N_{bp}} = \frac{3}{2} \frac{L_p L_{bp}}{R_c^2} \ln \left( \frac{2R_c}{R_0} \right) \quad (2.3.17)$$

which corresponds to the expression obtained by integrating the bending over an inverse spool [55]. We can use standard values typical for a  $\lambda$ -phage to obtain an estimate of the difference in bending energy between these two states. The bending energy changes from the isotropic to the completely ordered phase by a factor of  $\approx 5\%$  (full packing) to  $\sim 25\%$  (low packing density) depending on the length of encapsidated DNA.

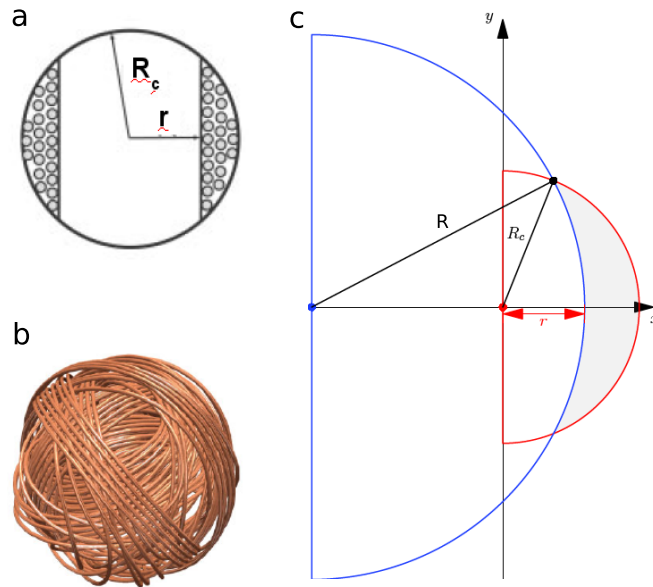
Note the term  $\sim \ln R_c$  present in the isotropic state is also seen in the work by Morrison and Thirumalai [57] – although their  $\ln R_c$  term is not extensive and lacks an equivalent lower cut-off  $R_0$ . The parameter  $R_0$  is easily identified as the inaccessible region in the center of the spool like configuration, i.e. DNA cannot fill the central volume of the spool due to its finite size and the corresponding bending energy. When the bending energy is small with respect to electrostatics,  $R_0$  can be obtained from a variational approach at high density [9]:

$$R_0(N) \approx \frac{10R_c}{\sqrt{N}}. \quad (2.3.18)$$

which is of use when the system is in the liquid crystal regime and beyond, but is not useful at low densities. In the disordered case, we can obtain a similar scaling at large packing by variationally treating the competition of excluded volume and bending energy, albeit this is of limited use as the disordered state is known to transition into an inverse spool at high density [22]. Alternatively we can use a constant value comparable to the width of a DNA base pair  $R_0 = 2.5$  nm.

### 2.3.2 Loop bending model

In the bending dominated regime, the DNA persistence length is larger than the confining radius  $L_p > R_c$ . The major contribution to the free energy comes from bending which is determined by the possible conformations the DNA strand can have inside the confinement. We approach the problem by finding an elementary DNA conformational unit which we can use as an effective quasi-particle akin to Kuhn segments in bulk [1]. The DNA is expected to be bent into the shape of the confinement so it will layer around the interior surface. We will define our effective unit as one loop of DNA which forms a circle. This method was first introduced by Purohit [22, 55] to model DNA in the inverse spool packing organization inside viral capsids, and was later studied for elastic wires packed in spherical confinement [56, 58].



**Figure 2.7:** a) Schematic for the inverse spool model with loops. Taken from Ref. [59]. b) Reconstruction of stiff wire packing experiment inside spherical confinement. Adapted from Ref. [56]. c) Drawing of the integration method used for determining the order  $S$  dependence of the loop radius  $r$  scaling with the total length  $L$  of confined DNA. The shaded area represents the volume filled with DNA which is obtained from the intersection of the large circle of radius  $R = r/(1 - S)$  (blue) and the capsid sphere of radius  $R_c$  (red).

We can now define basics elements of the DNA loop model. The bending energy change  $dF$  for an infinitesimal increase  $dl$  of packaged DNA length from  $L$  to  $L + dl$  can be represented as

$$dF = \frac{1}{2} k_B T L_p \frac{dl}{r^2(L)} \quad (2.3.19)$$

where  $r(L)$  is radius of curvature forced on the length  $dl$  of DNA by the available free volume. This internal radius  $r$  will become smaller as more DNA is packed inside and layered on the interior surface, also leading to a non-linear increase of bending energy.

Also, depending on the symmetry of packing,  $r$  will have different functional dependences. We can construct two simple cases (See Fig. 2.7a,b): a) DNA packages with a cylindrical symmetry ("inverse spool"), b) DNA packages randomly, filling equally the interior surface ("onion").

In the case of an inverse spool packing, the internal radius  $r$  corresponds to the cylindrical radial distance. As the length of DNA  $L$  is increased inside the capsid,  $r$  changes according to [55]

$$r(L) = R_c \sqrt{1 - (\xi L)^{\frac{2}{3}}} \quad (2.3.20)$$

with  $\xi = \frac{3A_0}{4\pi R_c^3}$  where  $R_c$  is the spherical confinement radius and  $A_0$  is the cross sectional area per DNA strand. The cross sectional area will depend on the DNA-DNA electrostatic repulsion, and will for full packing represent the area per DNA strand in hexagonal packing. This results in a total bending energy of:

$$\frac{F_b}{k_B T} = \frac{3L_p}{2R_c^2 \xi} \left[ -\sqrt[3]{\xi L} + \frac{1}{2} \ln \left( \frac{1 + \sqrt[3]{\xi L}}{1 - \sqrt[3]{\xi L}} \right) \right] \quad (2.3.21)$$

In the "onion" like regime the internal radius goes as:

$$r(L) = R_c \sqrt[3]{1 - \xi L} \quad (2.3.22)$$

leading to a total bending energy of:

$$\frac{F_b}{k_B T} = \frac{3L_p}{2R_c^2 \xi} \left[ 1 - (1 - \xi L)^{\frac{1}{3}} \right] \quad (2.3.23)$$

We see that the bending energy per unit of length grows with  $L$  as smaller loops carry a larger bending contribution in both regimes, which is a necessity as experiments clearly show that the ejection force is length dependent [22].

The packing of DNA inside viral capsids is expected to progress first as a disordered "onion" packing corresponding to Eq. (2.3.22) and at longer packaged lengths to change into an ordered "inverse spool" configuration. The bending energy of an onion configuration we calculate is expectedly lower than in the ordered case as it does not introduce an energetic penalty for loops to cross each other. Each loop crossing will produce stronger electrostatic repulsion and require some degree of bending. We can make an estimate of such effects by introducing a scaling relationship accounting for direct collisions of loop pairs.

The free energy penalty coming from collisions of two DNA loops should be proportional to the square of the local loop density  $\rho_{loop}^2$  (probability of finding two particles at one place), the total volume  $V_c$  in which the collisions could happen, the excluded volume for interactions  $v$  and the probability  $P_{collision}$  of the collisions happening. We can thus

write a scaling relationship:

$$F_{collision} \approx k_B T \rho_{loop}^2 \times V_c \times v \times P_{collision}. \quad (2.3.24)$$

The probability of the collision will depend on the way DNA is packed. If packing is progressing in the inverse spool, then collisions (crossings) between two loops are non-existent. On the other hand, if all DNA loops are disordered the collisions will happen always. We are here led to define an order parameter  $S$ , in line with liquid crystal theory [60], such that  $S = 1$  in the inverse spool and  $S = 0$  in the onion model. With a mixture of two packing regimes the order parameter will be in the range  $S \in [0, 1]$ . The collision probability for any two DNA loops to cross each other is a sum of the probability of one being ordered and the second disordered, with a multiplicity of two, and of both being disordered, leading to:

$$P_{collision} = 2 \times S \times (1 - S) + (1 - S) \times (1 - S) = 1 - S^2. \quad (2.3.25)$$

A rough estimate for the excluded volume  $v$  would be to take it as being equal to the maximal excluded volume of two DNA rods of persistence length  $L_p$  and effective width  $D$ ,  $v = 2L_p^2 D$ . This gives the basic scaling of for the energetics of loop packing.

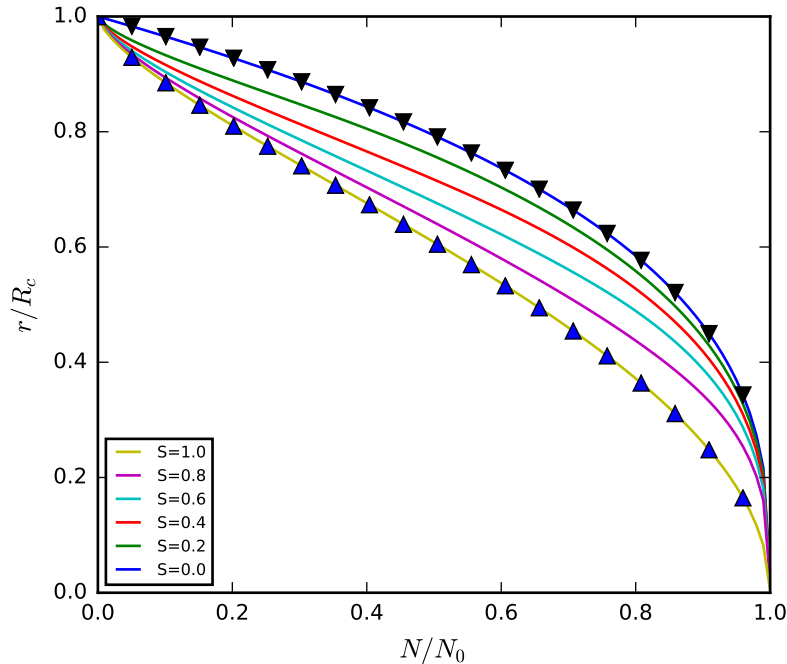
To calculate the collision free energy, we determine the number of loops  $n_h(L)$  in confinement. The number of loops will grow with  $L$  non-linearly depending on the packing type. For an increase of DNA length of  $dL$  the number of loops increases as

$$dn_h = \frac{dL}{2\pi r(L)} \quad (2.3.26)$$

from which one can calculate the number of loops in both regimes by simply integrating using eq. (2.3.22) or (2.3.20). This gives sufficient information to calculate the total bending energy in both cases.

In reality, DNA will not always be completely ordered or completely disordered inside the system. The DNA will transition between these regimes during viral ejection, similarly to the way it transitions from the liquid crystal state with lower DNA stiffness [10]. To study this effect, one can construct an effective model of packing between the  $S = 1$  (inverse spool) and  $S = 0$  (onion) regimes. The simplest model for a transition between these two packings for which it is possible to calculate how  $r$  scales on  $L$  is by producing an interpolation between these two regimes. To obtain how  $r(L)$  scales, we construct a geometric model (Fig. 2.7c) of a cross section of our confining sphere and another sphere of radius  $R$ . The spheres are placed such that for  $S = 1$  the inner sphere radii  $R \rightarrow \infty$  so that it represents a cylinder, and for  $S = 0$  the inner sphere radii  $R \rightarrow r$  so that it represents the inner surface of the packaged DNA ("onion shell"). Assuming  $R = r/(1 - S)$

allows us to calculate the volume  $V(r, S)$  of the cross section of the two spheres. It is this volume that is filled as more DNA is inserted. Because of the shape of volume, an effective radius of the current loop  $r$  is defined. Numerically one obtains how  $r(L)$  depends on the order  $S$  (Fig. 2.8). For  $S = 0$  and  $S = 1$  we obtain the analytical results, while for a varying degree of order we obtain a smooth transition between these regimes. This allows the calculation of how the free energy in the model depends on the amount of packaged DNA.

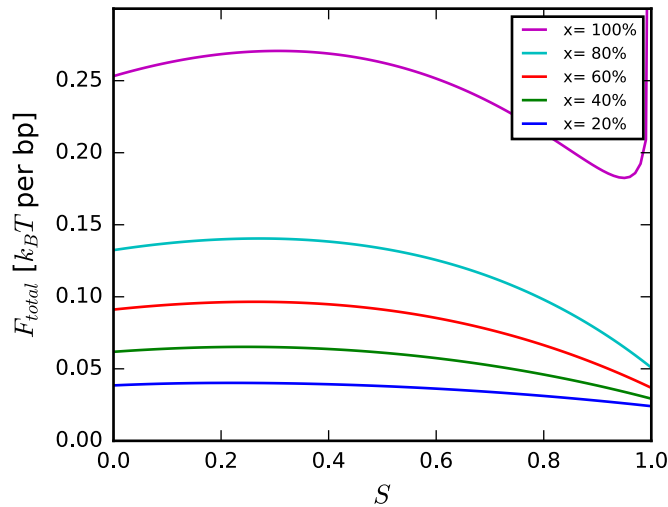


**Figure 2.8:** Dependence of loop radius  $r(N)$  as a function of packing percentage for several different order parameters  $S$ . Symbols mark the inverse spool ( $S = 1$ ) and onion-like ( $S = 0$ ) packing limits given by Eq. (2.3.20) and (2.3.22), respectively.

The total free energy can be estimated using our scaling relationship based on the numerically obtained  $r(N, S)$  (Fig. 2.8) with both the bending energy and collision energy depending on the ordering inside, i.e. by using our previous approximation for the contribution of collisions  $F_{tot}(N, S) = F_{collision}(N, S) + F_b(N, S)$  (Fig. 2.9). We see there are two local free energy minima. First is at  $S = 0$ , corresponding to the onion model. The second, at  $S = 1$  represents the inverse spool model, and is a global minima because there are no loop collisions. Interestingly, there is an energy barrier separating the ordered ( $S = 1$ ) and disordered ( $S = 0$ ) regimes, which becomes larger as the packing is increased from a larger number of loops. For full packing, we see that  $S = 1$  becomes globally unfavourable as a result of our model not treating high density phases correctly.

The loop model presented here gives interesting insight into packing of stiff DNA inside confinement. First, it predicts that the inverse spool phase is more stable, but





**Figure 2.9:** Total free energy in loop model with contributions from bending  $F_b(N, S)$  and collisions  $F_{collisions}(N, S)$  as a function of the order parameter for different lengths of packed DNA (here represented as packing fractions  $x = N/N_0$ ). At full packing a dense hexagonal phase with DNA strand separation of  $d = 2.8$  nm is assumed.

it also predicts that it is possible that packaging could remain stuck in the disordered phase, because of the energy barrier. This could produce a hysteresis between packing and unpacking, when the disordered phase can persist for long lengths during packaging, but will be suppressed during ejecting the material. We can also relate our model to packing of stiff wires in spheres [56, 58], where it could be applied to give analytical fits to experimental data of partially ordered packings.

Torsion can be relevant for packing, even leading to a lower degree of ordering when there is no mechanism for relaxation [56]. In reality, packing in the disordered phase consists of small ordered groupings of wires. Whenever the coiling radius becomes too small, a new loop is started at a different orientation [56]. So each layer seems to be ordered in respect to itself, but with a different orientation between different layers. This seems to minimize the crossing penalty. At higher densities, Stoop *et al.* [56] see a transition to a "inverse spool" packing regime, as predicted by the loop model.

## 2.4 Confined DNA in the isotropic regime

After discussing the DNA bending, and specifically the bending dominated regime of DNA confinement, we will continue with the regimes where the bending contribution to the free energy is not dominant. Going in the opposite direction to that of *in vivo* ejection (Fig. 2.5), we start in the isotropic regime, and follow into the liquid crystal regime. The DNA at low packing densities, below 30% of packed total length, is considered to be in the isotropic, or disordered, phase [10]. This phase has two distinct regimes [12]: a) the

first part is the ideal chain regime, when the whole DNA does not feel the confinement, b) the second is when the correlation length between different DNA segments become comparable to the size of confinement, and behaves as if being part of a semi-dilute bulk solution [11]. As DNA is a strongly charged polyelectrolyte, we will focus on contributions coming from angle dependent excluded volume interactions, which will dominate the free energy [13, 14]. In this section, we will explain the state of the DNA, its free energy, and use the resulting information to give estimates on pressures and forces involved during the last parts of DNA ejection.

When the DNA behaves as an ideal chain, i.e. when it can be represented as a random walk within confinement, we can renormalizes the length scales to reduce the case to that of an ideal chain [2]. We use a coarse graining scheme due to Khokhlov and Semenov [61]. The DNA is treated as an ideal polymer with a new renormalized monomer size (the Kuhn length) of  $2L_p$ , thus consisting of  $n = L/(2L_p)$  segments. This representation includes the "self-interaction" of the DNA strands. We must make a distinction between  $2L_p$ , which is the Kuhn length and becomes a measure of spatial extent, and the persistence length  $L_p$ , which is related to the mechanical stiffness.

In order to model the DNA in the isotropic regime, we must first determine the interactions between two segments of the DNA. Due to the confinement, these segments are forced to exist much closer to each other than in free space. This spatial exclusion coming from screened electrostatic interactions is called the excluded volume, and gives a measure of the crowdedness inside the capsid.

## DNA-DNA interactions

When two parts of a DNA molecule are in close proximity, they exert a repulsive force on each other. Effectively, each part of the DNA excludes a volume around it depending on the strength of the interaction. When the density of particles is such that there is more free volume than excluded volume, these particles are consider to be barely interacting. As the density increases, the excluded volume becomes a good measure of the free energy required to obtain such a packing. We will use measured DNA-DNA interaction parameters from the seminal work by Rau and Parsegian [40] to obtain how the excluded volume behaves, taking into account the mutual orientation and distance of two DNA segments.

The dominant part of the excluded volume for two DNA molecules comes from screened electrostatic interactions. Due to its local rigidity DNA behaves as a long rod with a linear charge density  $\lambda$  and effective width  $D$  inserted into a salt solution with electrostatic screening length  $\kappa^{-1}$ . A solution of the linearised Poisson-Boltzmann equation (Debye-Hückel) for the interaction of two long charged DNA rods with electrostatic screening ( $L_p \gg \kappa, D$ ) which have their points of closest separation at  $\mathbf{r}_1$  and  $\mathbf{r}_2$  and a mutual angle

of  $\gamma$  between direction vectors (directors)  $\mathbf{n}_1$  and  $\mathbf{n}_2$  is given by [62]

$$U_{12}(\mathbf{r}_1, \mathbf{r}_2) = \frac{\pi\lambda^2}{\epsilon\kappa} \frac{e^{-\kappa|\mathbf{r}_1-\mathbf{r}_2|}}{\sin\gamma} \quad (2.4.1)$$

where  $\lambda$  is the effective linear charge density related to the effective charge per monomer  $\nu_{eff}$  as

$$\lambda = \frac{\nu_{eff}eN_{bp}}{L_{bp}N_{bp}} = \frac{\nu_{eff}e}{L_{bp}}. \quad (2.4.2)$$

The bare charge of DNA is  $\nu_0 = 2$  elementary charges per base pair, but in physiological conditions due to Manning-Oosawa counterion condensation [63] the effective charge is reduced from its bare value  $\nu < \nu_0$ .

The angle dependent electrostatic excluded volume (second virial coefficient) is then calculated as [13, 51]

$$\beta(\mathbf{n}_1, \mathbf{n}_2) = \frac{1}{V} \int \int (e^{-\frac{U_{12}(\mathbf{r}_1, \mathbf{r}_2)}{k_B T}} - 1) d\mathbf{r}_1 d\mathbf{r}_2 \quad (2.4.3)$$

which for the case of two charged rods of length  $L_p$  at a closest separation of  $x = |\mathbf{r}_1 - \mathbf{r}_2|$  reduces to

$$\beta(\mathbf{n}_1, \mathbf{n}_2) = 2L_p^2 \sin\gamma \int_D^\infty (1 - e^{-\frac{U_{12}(x, \gamma)}{k_B T}}) dx \quad (2.4.4)$$

Because of the ambiguity in choosing the effective linear charge density due to correlations and screening we opt to use an experimental fit which includes both hydration repulsion and screened electrostatic interactions between two parallel rods as measured by Strey *et al.* [64]

$$U(r) = a \frac{e^{-\kappa_H r}}{\sqrt{\kappa_H r}} + b \frac{e^{-\kappa r}}{\sqrt{\kappa r}}. \quad (2.4.5)$$

where  $\kappa_H^{-1}$  and  $\kappa^{-1}$  are the hydration and electrostatic decay lengths, respectively. Following the outline for skewed rods [62], with a mutual angle of  $\gamma$  at a minimal distance  $x$ , we obtain for the angle dependent potential:

$$U(x, \gamma) = \sqrt{\frac{\pi}{2}} \frac{a}{\kappa_H} \frac{e^{-\kappa_H x}}{\sin\gamma} + \sqrt{\frac{\pi}{2}} \frac{b}{\kappa} \frac{e^{-\kappa x}}{\sin\gamma}. \quad (2.4.6)$$

From this we can numerically calculate the angle dependent excluded volume  $\beta(\mathbf{n}_1, \mathbf{n}_2)$  in (2.4.3). Representative values for a 100mM monovalent salt solution are  $\kappa_H^{-1} = 0.3$  nm,  $\kappa^{-1} = 0.974$  nm,  $a = 1.7 \cdot 10^4$   $k_B T/\text{nm}$  and  $b = 85$   $k_B T/\text{nm}$  [64].

## Modelling the DNA in the isotropic regime

When the end-to-end distance of the encapsitated DNA is < comparable to the capsid size, it starts to interact with its confinement. The free energy of the DNA will be

dominated by bending with excluded volume interactions becoming relevant at higher densities. Using pairwise interactions, a Flory scaling approach [2] when  $L_p^2 < R_c a$  [12] gives that the free energy scales with the number of DNA monomers as  $F \sim n^2$ . Scaling arguments give the results up to a constant of order unity. In contrast, numerical simulations combined with mean field modelling give the free energy at low density for a worm like chain confined inside a spherical surface as [57]

$$F = 0.56 \frac{L_p L}{R_c^2} - 1.1 \frac{L_p}{R_c} + 3 \ln(R_c). \quad (2.4.7)$$

which also gives a bending contribution, with terms  $L_p/R_c$  and  $\ln(R_c)$ , not obtained in typical scaling arguments [12] but it neglects strong electrostatic interactions. We can not return to Eq. (2.3.15) for the mean field bending as it assumes a homogeneous packing which is not yet established. The loop bending model as given by Eq. (2.3.23) in the disordered (or ordered) regime provides a better estimate.

In order to construct the free energy, we will use our loop bending Eq. (2.3.23) and combine it with pairwise interactions of the excluded volume type. When the worm like chain starts to intersect we add the excluded volume interactions inside a spherical cavity [2]

$$F_v = k_B T v \frac{n_p^2}{V_c} \quad (2.4.8)$$

where the relevant quantity is not the number of monomers  $N_{bp}/V$  but the number of "persistent segments" inside the confinement:

$$n_p = \frac{L_{bp} N_{bp}}{2L_p}. \quad (2.4.9)$$

Here,  $v$  is the isotropic excluded volume obtained by averaging the angle dependent excluded volume from Eq. (2.4.4) over all possible mutual (random) angles between two persistence segments of DNA:

$$v = \frac{1}{2} \int \int \beta(\mathbf{n}_1, \mathbf{n}_2) \frac{1}{4\pi} \frac{1}{4\pi} d\Omega_1 d\Omega_2. \quad (2.4.10)$$

We can integrate the excluded volume

$$v = \frac{1}{8} \int_0^\pi \int_0^\pi \beta(\sin \gamma) \sin \theta_1 d\theta_1 \sin \theta_2 d\theta_2 \quad (2.4.11)$$

where  $\gamma$  is the angle between the current directors  $\theta_1$  and  $\theta_2$ . We can change the origin of the first coordinate system so that it always coincides with the second, thus:

$$v = \frac{1}{4} \int_0^\pi \beta(\gamma) \sin \gamma d\gamma \quad (2.4.12)$$

leading to

$$v = L_p^2 D \left( \frac{\ln A + \gamma_e}{\kappa D} - 1 \right) - \frac{L_p^2}{2\kappa} \int_0^\pi |\sin \gamma| \ln(|\sin \gamma|) \sin \gamma d\gamma. \quad (2.4.13)$$

After integration we obtain the final equation for the isotropic excluded volume [13]:

$$v = L_p^2 D \left( \frac{\ln A + \gamma_e}{\kappa D} - 1 \right) - \frac{L_p^2}{\kappa} (\ln 2 - 1). \quad (2.4.14)$$

Here  $D$  is the effective width of the DNA,  $\gamma_e \approx 0.577$  a numerical constant and  $\kappa^{-1}$  the electrostatic decay length. We can make an estimate of the excluded volume in viral capsids by taking typical parameters ( $D \approx 3$  nm and  $\nu_{eff} = 1.9$ ) and an electrostatic decay length of 0.977 for 100 mM salt [64] and evaluate to obtain  $v \approx 2200$  nm<sup>3</sup>. For a viral capsid of radius  $R_c = 30$  nm, we obtain  $V_c \approx 50v$ . Thus, when 50 persistence length segments (about 2.5  $\mu$ m of DNA) is inside the capsid, all the available volume is excluded. This gives us a good estimate of the range of validity of models which take into account only the pairwise term in the virial expansion.

We can finally write the free energy for the whole isotropic regime as

$$\frac{F_d}{k_B T} = \frac{3L_p}{2R_c^2 \xi} \left[ 1 - (1 - \xi L)^{\frac{1}{3}} \right] + v \frac{3}{4\pi R_c^3} \left( \frac{L}{L_p} \right)^2. \quad (2.4.15)$$

This provides one equation for estimating the free energy before the DNA enters a liquid crystal regime. Notably, for extremely short lengths  $L \rightarrow 0$ , the bending term reduces to the exact formula. Alternatively, if we have reason to believe that there is a strong tendency for ordered packing, e.g. as simulated [57], we can use the spool bending model given by Eq. (2.3.21) [59].

## 2.5 Liquid crystalline DNA phase

The liquid crystal phase is identified by local ordering of the DNA strands such that local DNA strand direction vectors (directors) become correlated on a global scale. This transition from an isotropic phase to a liquid crystal is known to closely resemble the same transition in concentrated unconstrained DNA solutions [65]. We base our modelling of the liquid crystal phase on adding small perturbations to bulk models. These perturbations are caused by spherical confinement and manifest themselves as a limit on the possible director field in the ordered phase and the addition of bending free energy.

As the DNA is one long strand confined inside a viral capsid, the confinement limits the possibilities for the liquid crystal global director to those that do not penetrate the capsid and have a minimal amount of tight bends. Additionally, we will assume that there

is no positional variation in the local ordering, i.e. that we have nematic ordering [60]. This leads us to approximate the director for a liquid crystal nematic order as the unit vector for the polar angle  $\hat{\phi}$  with  $f(\beta)$  the probability distribution for finding a DNA rod at an angle  $\beta$  towards  $\hat{\phi}$ . A flat distribution of  $f(\beta)$  means that all directions in space are equally probable (disordered state), while if  $f(\beta) = \delta(\beta)$  it would mean that all DNA strands are directed as  $\hat{\phi}$  (ordered state). A similar approach was adopted by Oskolkov *et. al.* [14] as they also studied the effects of confinement on the spatial dependence of polymer density and ordering with an Onsager approach [51]. They use only steric interactions and take into account the first correction to the entropy due to spatial and directional inhomogeneities. But they do not explicitly include the connectivity of the polymer contour and the resulting bending energy, which we expect to modify the isotropic to nematic liquid crystal transition.

### Onsager model for a liquid crystal

We use a modification of the Onsager approach for rods with excluded volume effects to treat the liquid crystal ordering. We extend this approach by addition of the mean field bending given by Eq. (2.3.15) calculated before and study how it affects the liquid crystal transition. The Onsager approach [51] we adapt here [13] predicts the liquid crystal transition well, although it does not account for other non-nematic orderings.

In the Onsager approach, the free energy  $F_{lc}(f(\beta))$  is a functional of the probability distribution function  $f(\beta)$  for finding a local DNA strand oriented at an angle  $\beta$  to the the director  $\hat{\phi}$  and can be written as [13]

$$\frac{F_{lc}}{k_B T N_{bp}} = \frac{1}{N_{bp}} S - \frac{1}{N_{bp}} \Delta S_{or}[f_{\mathbf{n}}] + \frac{F_{bend}[f_{\mathbf{n}}]}{k_B T N_{bp}} + \frac{1}{2} \frac{n^2}{N_{bp} V_c} \int \int \beta(\mathbf{n}_1, \mathbf{n}_2) f_{\mathbf{n}_1} f_{\mathbf{n}_2} d\Omega_1 d\Omega_2 \quad (2.5.1)$$

where  $n_p$  is the number of persistent segments  $n_p = \frac{L_{bp} N_{bp}}{2L_p}$  but all coming from the *same* DNA in our case, which is in contrast to the standard Onsager approach.

The content of the terms in  $F_{lc}$  is the following. The last term given in Eq. (2.5.1) is the anisotropic excluded volume given by Eq. (2.4.4) and represents the first term in the virial expansion for DNA-DNA interactions. The resulting integral averages over all director orientations of two DNA strands with directors  $\mathbf{n}_1$  and  $\mathbf{n}_2$ . The orientational entropy can be obtained from the director distribution function [66]

$$\Delta S_{or}[f_{\mathbf{n}}] = -\frac{L}{2L_p} \sigma(f_{\mathbf{n}}) = -\frac{L}{2L_p} \int f^{\frac{1}{2}}(\cos\theta) \Delta f^{\frac{1}{2}}(\cos\theta) d\Omega \quad (2.5.2)$$

with  $\Delta$  given as:

$$\Delta = \frac{1}{\sin\theta} \frac{\partial}{\partial\theta} \sin\theta \frac{\partial}{\partial\theta}. \quad (2.5.3)$$

The bending contribution we take is from our mean field bending model given by Eq. (2.3.15), which was averaged over all the confined interior. This approximates the bending energy in a homogeneous mixture of volume  $V_c$ :

$$\frac{F_{bend}[f_{\mathbf{n}}]}{k_B T N_{bp}} = \frac{L_p L_{bp}}{R_c^2} \left[ 1 - S[f_{\mathbf{n}}] + \left(1 + \frac{1}{2} S[f_{\mathbf{n}}]\right) \ln \left( \frac{2R_c}{R_0} \right) \right]. \quad (2.5.4)$$

This term is not present in the standard liquid crystal approach to DNA [13], and allows us to study the effects of bending on the transition point. From experiments (see Ref. [10] and references therein) we know that in monovalent salt conditions there is no exclusion hole in the center of the capsid thus justifying our homogeneous approximation.

In order to solve the resulting equation we use the Onsager ansatz [51]:

$$f(\theta) = \frac{\alpha}{4\pi \sinh \alpha} \cosh(\alpha \cos \theta) \quad (2.5.5)$$

dependent on a parameter  $\alpha$ . Instead of having to solve for a continuous function  $f(\theta)$  we can now solve for the parameter  $\alpha$ . As a measure of the order, we will use the nematic order parameter  $S = \langle P_2(\cos \theta) \rangle$  [60]. The preceding ansatz allows us to directly connect the order parameter  $S$  and the parameter  $\alpha$ .

Finally, the order dependent part of the free energy can be written as:

$$\frac{\Delta F}{k_b T N_{bp}} = \frac{L_{bp}}{2L_p} \sigma(\alpha) + A_b \langle P_2(\cos \theta) \rangle + \left( \frac{L_{bp}}{L_p} \right)^2 \frac{2L_p^2}{\kappa V} [(\ln A + \gamma_e - \kappa D) \rho(\alpha) - \rho_{twist}(\alpha)] \quad (2.5.6)$$

where the constants are

$$A_b = \frac{L_p L_{bp}}{R_c^2} \left[ \frac{1}{2} \ln \left( \frac{2R_c}{R_0} \right) - 1 \right]. \quad (2.5.7)$$

and

$$A = \frac{\pi \nu_{eff}^2 l_b}{L_{bp}^2 \kappa}. \quad (2.5.8)$$

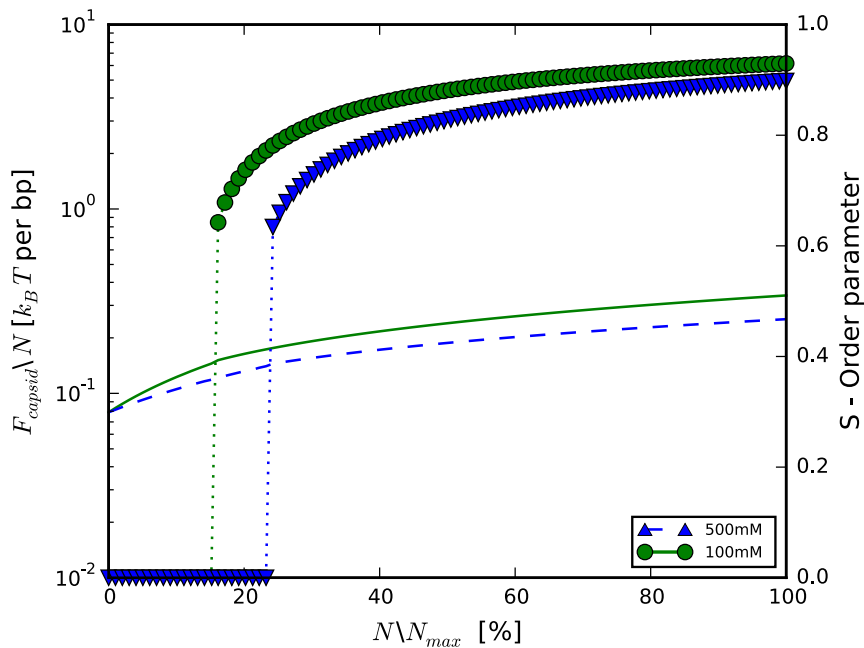
The order independent part amounts to:

$$\frac{F_{bend}[f_{\mathbf{n}}]}{k_B T N_{bp}} = \frac{L_p L_{bp}}{R_c^2} \left[ 1 + \ln \left( \frac{2R_c}{R_0} \right) \right] \quad (2.5.9)$$

For the bending contributions we take  $R_0 = 2.5$  nm as a representative value as it approximates the radius of the inaccessible area in the middle of a spool-like packing with the diameter of DNA.

## Application to the state of DNA inside viral capsids

One can proceed to minimize the free energy for the Onsager model as a function of the number of base pairs present inside the capsid. A full calculation for the free energy given by Eq. (2.5.1) with the possibility of ordering into a nematic liquid crystal is shown in Fig. 2.10. We use three representative values of the monovalent salt concentration 150 mM and 500 mM with experimentally measured potential parameters [64]. We see that the phase transition, manifest as an abrupt increase in the order parameter  $S$ , happens at  $N_c/N_{max} \sim 20\%$  of packing, with smaller salt concentrations expectedly inducing an earlier transition due to stronger electrostatic interactions.



**Figure 2.10:** Free energy per base pair (chemical potential) in the Onsager approach as given by Eq. (2.5.1) for one DNA molecule encapsidated in a spherical capsid of radius  $R_c = 30$  nm with  $L_p = 50$  nm. Lines show the results of calculations for two different monovalent salt concentrations (100 mM, 500 mM) as a function of the number of DNA base pairs  $N$  inside the capsid. The connected symbols are corresponding values of the order parameter  $S$  (right  $y$ -axis). The maximal number of packed base pairs was  $N_{max} = 58.6$  kbp.

Figure 2.11 shows the transition point, via the fraction of total packed DNA, for the transition into a liquid crystal state as a function of persistence length with and without the bending contribution. We note that the persistence length contributes both as a measure of bending stiffness and a length scale for the renormalization of the polymer random walk, so that without bending we return to a bulk liquid crystal transition. As expected there is a clear trend for earlier ordering as the persistence length is increased. The added bending contribution significantly influences the transition point at low salt conditions.

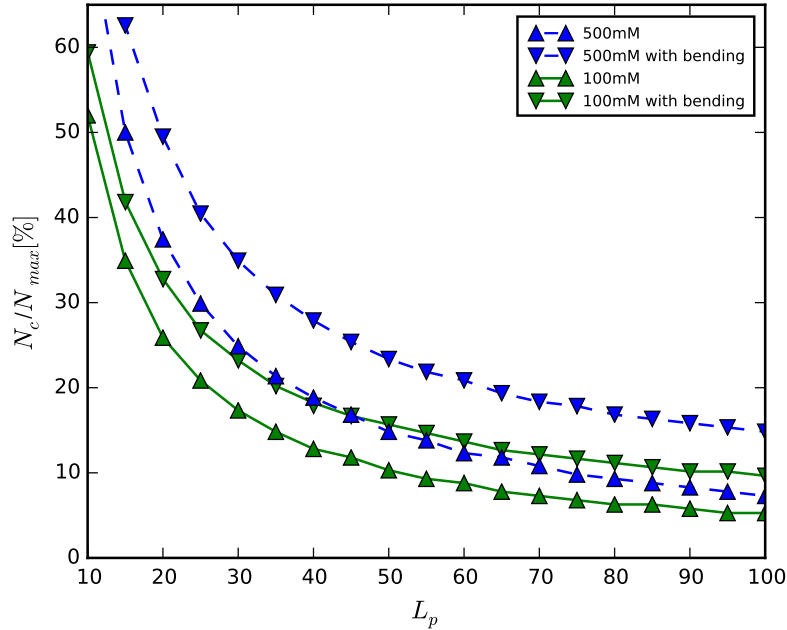


A comparison with experiments which give a transition at 30% encapsulated length, e.g. in T5 [65], would indicate that the effective persistence length is somewhat lower than  $L_p = 50$  nm, or that we missed some important term like twist effects [67]. Comparing the transition points, without and with bending, we see that bending suppresses the transition, possibly because in the ordered case there are tight bends near the center of the capsid that are energetically costly.

Let us discuss the values of the persistence length in the context of the transition point. It is expected to be around 50 nm in the bulk with only monovalent salts to become larger when the salt concentration is decreased [45, 68]. A small amount of multivalent ions remaining inside the capsid could however lower the persistence length. Also, recent experiments on the flexibility of DNA on length scales shorter than its persistence length show that the worm-like chain model overestimates the bending energy at small length scales [69], although this is not yet certain [70]. Thus, we use  $L_p = 50$  nm as a reference for the flexibility of DNA and test how a change of DNA stiffness alters the transition point. Notably, temperature can also influence the persistence length. A change of ambient temperature from 40°C to 5°C has been shown to cause an increase of the persistence length of  $\sim 10$  nm [71]. According to Fig. 2.11, a change of about 10 nm, could cause a transition to a disordered state. This could explain experimental results on partially filled capsids where the liquid crystal state was shown to be extremely sensitive to temperature [10].

Our results on the liquid crystal transition point are in contrast to Molecular dynamics simulations (MD) [72] which give a liquid crystal transition at roughly two times the encapsulated length than in experiments ( $\sim 70\%$ ) [10]. Although there has been some controversy recently regarding the treatment of entropy in some MD studies [73, 74, 75], we believe that the main reason is in the under-representation of the anisotropy of the DNA interstrand interactions. MD studies usually use a coarse-grained model of a bead spring with added "dihedral angles" to mimic the effects of a persistence length [76]. Here each bead effects screened electrostatic interactions on other beads based on the rational notion that a line of charged point particles will produce the same interaction as a charged rod. If we examine one "solid" rod of length  $\sim L_p$  we find that in a bead spring model it would have  $\sim 100$  point particles of charge unity. Such a construction has a pronounced cylindrical symmetry. With DNA there also exists a delocalized cloud of counterions screening the bare interactions [77] which is included in the experimentally determined interactions [64]. Thus, we opted for coarse graining the DNA into rods with anisotropic interactions.

Although many theoretical models [37, 55] and numerical simulations [72, 78, 79] predict that an exclusion hole should always exist, irrespectively of the packing fraction and solvent conditions, experiments indicate that this is the case only in partially filled



**Figure 2.11:** Critical length of DNA for phase transition from isotropic to liquid crystal as a function of DNA persistence length. Results are given for three different salt concentrations and with or without any bending contribution to the liquid crystal free energy. Capsid size is  $R_c = 30$  nm. For other parameters see preceding section.

capsids at low temperatures or with added condensing agents [10]. Quenching of mostly empty capsids from room temperature to  $\sim 4^\circ$  C causes a transition from a homogeneous isotropic state to an inverse spool with a depletion of DNA density in the core of the capsid [10]. The assumption of a homogeneous density of particles is valid in the limit of  $L_p \ll R_c$  where there is no preference for the DNA to leave the central part of the capsid due to costly bends in the contour. In analogy to the Odijk two phase picture for dense packings [80, 81], we propose that as the bending contribution is increased the center of the capsid will behave as an exclusion hole for the DNA. This happens when the electrostatic repulsion energy is not large enough to maximize the distance between the strands. DNA in the center of the capsid has a tendency to stretch outwards, thus decreasing the effective volume taken by the DNA and increasing the density of interacting DNA rods. In our interpretation of the experiments a decrease in temperature can cause an increase in the persistence length, thus changing the ratio between bending and electrostatic excluded volume energies on behalf of the former. Such a transition from a homogeneous state will take place when the bending energy of random packing of DNA loops given by Eq. (2.3.23) is able to overcome the excluded volume interaction that is spreading the DNA

throughout the whole capsid volume:

$$\frac{3L_p}{2R_c^2\xi} \left[ 1 - (1 - \xi L)^{\frac{1}{3}} \right] \sim v \frac{3}{4\pi R_c^3} \left( \frac{L}{L_p} \right)^2. \quad (2.5.10)$$

We can think of it as the condition that the excluded volume repulsion overpowers the bending contribution at higher densities – forcing the DNA to fill all the volume at the cost of increased bending. Using parameters from this section, we obtain a transition at around  $\sim 30\%$  packing, albeit highly sensitive to the value of the persistence length. From experimental measurements of the DNA persistence length we find that a change of temperature of  $\sim 20^\circ$  causes a change of  $\sim 15\%$  in the persistence length, which could be enough to explain the before-mentioned partial onset of exclusion hole formation. This would indicate that the state of packaged DNA at *in vivo* temperatures is just outside the bending regime.

## 2.6 Viral ejection *in vivo*

Ejection of dsDNA from bacteriophages is, in spite of half a century of active research, still poorly understood [18, 82]. The process of ejection is of major importance in the viral infection process. The ejection begins with the bacteriophage attaching to the bacteria and the removal of a protein "cork" (tail portal covering) which stops the densely packed DNA from exiting [54]. The DNA is packed to extreme densities and exerts a pressure of 25 – 100 atm on the bacteriophage capsid [9, 20]. Models developed and tested *in vitro* (see [18, 82] and references therein) predict that the ejecting force resulting from even such a dense packing is insufficient to completely transfer the DNA into the cell interior. Estimates of cellular pressure in bacteria range from 0.1 – 25 atm [21, 83, 84]. Although cells have smaller turgor pressures than fully packed bacteriophages, the ejecting force (and pressure) in the capsid drops almost exponentially as it empties [22] indicating that ejection should stall before the genome is fully ejected.

Models of DNA ejection based on the continuum theory by Ubbink and Odijk [16, 42] have been applied to explain the ejection when both the ejected and encapsidated DNA is condensed [37]. In the model by Tzlil *et al.* [37] the capsid DNA, being confined, has a non-optimal shape which generates the force ejecting it from the capsid all until the ejection force is balanced by an external *counterforce*. If this *in vitro* model is applied to *in vivo* cases it predicts that ejection stalls when the cellular turgor pressure is larger than about 0.5 atm. Typical cellular pressures are estimated to be 0.3–5 atm for Gram-negative [21, 83, 84] and 15 – 25 atm for Gram-positive bacteria [84]. This would indicate that ejection cannot proceed without some help.

Typical *in vitro* experiments do not provide a clean separation between two (three)

compartments, one in the virus (immersed in the extracellular fluid), and the other in the cell [17], which is a key feature of the ejection process we aim to include in the model presented in this section. The theoretical models applied in *in vitro*, which assume condensed DNA inside the capsid, may thus have a limited applicability in the *in vivo* conditions.

Condensing agents are found in some capsids but this is not a general situation and has been related to reduced infectivity [85]. Many viral capsids are permeable to small ions so we expect intra-capsid solvent conditions analogous to those in their exterior. Because typical extracellular fluids have no condensing agents in sufficient concentrations, DNA ejection into bacteria is expected to proceed from a non-condensed state.

A recent single molecule Hershey-Chase experiment [23] hints that *in vivo* ejection is controlled not by the amount of DNA left inside the capsid but by the amount of DNA ejected into the cell. This means that once the pressure built-up in the bacteriophage is spent on the DNA ejection, a cellular mechanism must take over. This has been documented *in vivo* in bacteriophages T5 [86] and T7 [87]. A variety of mechanisms have been proposed for completing the ejection: nano-motors or enzymes that ratchet in the stalled part of the DNA [87, 88], a solvent flow through the semi-permeable capsid and into the cell simultaneously flushing the DNA through the tail [89], osmotic pressure from proteins remaining in the capsid [90] and diffusion of DNA with assisted pulling from proteins in the cell [91]. It appears that none of these models give a definite answer while experiments suggest a coexistence of many different mechanisms.

We will calculate the pressures and forces driving the last parts of viral ejection using thermodynamic models of the DNA inside the capsid (in an isotropic state) and condensed DNA inside the cell (Sec. 2.1), and compare them to bacterial internal pressure (turgor pressure). The phage DNA is modelled as one long strand of total length  $L_0$  and persistence length  $L_p \approx 50$  nm [54] able to move between two compartments with different thermodynamic conditions – the virus and the cell. It is thus partitioned in two pieces, one of length  $L$  inside the virus head and tail, and the other of length  $L_0 - L$  in the cell. Our premises are that DNA is in a non-condensed state in the capsid and in a condensed state in the cell.

## 2.6.1 DNA inside the cell

DNA is condensed in the cell due to the presence of multivalent cations *during* ejection, but not necessarily afterwards. Condensing DNA inside the cell does not necessarily inhibit its function. A high reaction rate, and thus a high DNA transcription rate, was demonstrated with condensed DNA [92]. Multivalent ions and osmolytes condense the DNA. The shape of the condensate is a result of a competition between DNA-DNA

attractions mediated by the cations and unfavourable effects of DNA bending (See Sec. 2.1).

The DNA inside the cell is treated as being condensed in a "hexagonally packed continuum" within the U-O model of Sec. 2.1. The packaging inside the cell is modelled similarly as by Tzlil *et al.* [37] as a DNA condensate in the Ubbink-Odijk model [16, 42]. The volume  $V$  of the condensed DNA is proportional to the contour length  $L$  of the DNA  $V = A_0(L_0 - L)$  with  $A_0$  the area per unit length of DNA. The DNA condensate has a free energy  $F_{tor} = -\gamma V + \sigma S + E_{bend}$  as presented before. The total free energy  $F_{tor}$  is minimal when the condensate has the shape of a torus [37], but there are other possibilities like rods which may become relevant if the DNA stiffness is greatly reduced [30]. A different shape would only change the surface and bending energy terms to some degree and provide a small correction to our conclusions.

The DNA in the condensate is assumed to be hexagonally packed with the area per unit length of DNA  $A_0 = \pi d_0^2 / \sqrt{12}$  – the packing constant for hexagonally packed cylinders [37]. Here  $d_0 \approx 2.8$  nm is the experimentally determined closest separation between DNA strands with added condensing agents [40]. We note that Tzlil *et al.* [37] model the surface free energy contribution by assuming the loss of half of DNA nearest neighbours while we assume the loss of a third of the nearest neighbours as derived by Ubbink and Odijk [42]. This constitutes a minor correction to the surface free energy term, but may become important if used to determine model parameters from toroid shapes.

The approximation adopted by Tzlil *et al.* assumes the toroid to have the shape of a thin torus as in Sec. 2.1. We will adopt this approximation due to its simplicity, but also provide a full minimization of the total free energy for various DNA lengths. We calculate the free energy of the condensate based on the variational approach result for the major radius  $R$  given by Eq. (2.1.14) so that the minimal free energy of a toroidal condensate [37] is

$$F_{tor} = -\epsilon_0(L - L_0) \left[ 1 - b \left( \frac{A_0^2 L_p}{d_0^4 \epsilon_0} \right)^{\frac{1}{5}} (L - L_0)^{-\frac{2}{5}} \right], \quad (2.6.1)$$

where  $\epsilon_0 = A_0 \gamma$  is the condensation free energy per unit length in an ideally packed hexagonal lattice and  $b = 1.308$  a constant. This  $\epsilon_0$  was determined by Tzlil *et al.* [37] by fitting the toroid major axis to experimental data for  $\lambda$ -DNA in a solution of polylysine which is also present in some bacteria [30]. As they chose a different surface term than the original model [16] and us, the condensation free energy per unit length they obtain  $\gamma_T$  is smaller than the one obtained here  $\gamma_{UO}$ . The shapes of toroids in the U-O model are determined by a non-dimensional parameter [16]

$$\alpha = \frac{\sigma V^{\frac{1}{3}} d_0^2}{L_p} \quad (2.6.2)$$

which depends on the choice of surface energy  $\sigma$ . By using a different choice for the surface energy the fits to toroid shapes would correspond to different values of the surface free energy  $\sigma$  than in the original model. Because  $\sigma$  is derived from  $\gamma$ , we have that the Ubbink-Odijk model, adopted here also by us, has  $\gamma_{UO} = \frac{3}{2}\gamma_T = 0.15 k_B T/\text{nm}^3$  ( $\sim 6$  atm) - this difference arises solely from different treatment of the missing neighbours at the surface of the toroid (one third missing in [16] -  $\gamma_{UO} \sim 6$  atm vs. one half missing in [37] -  $\gamma_T \sim 4$  atm). An alternative approach to fitting is to obtain the condensing energy per unit length from osmotic force measurements [40], e.g. a 20 mM solution of Cobalt Hexamine corresponds to  $0.024 k_B T/\text{nm}^3$ . The comparison of the two values obtained ( $0.15 k_B T/\text{nm}^3$  vs.  $0.024 k_B T/\text{nm}^3$ ) shows that the thermodynamics of the condensed DNA importantly depends on the condensing agent.

When most of the viral DNA is in the condensate,  $L \ll L_0$ , the surface and bending terms in Eq. (2.6.1) are negligible [37]. This is because the largest contribution to the free energy of tori comes from the bulk term in  $F_{tor}$  as can be seen from

$$1.308 \left( \frac{A_0^2 L_p}{d_0^4 \epsilon_0} \right)^{\frac{1}{5}} L_0^{-\frac{2}{5}} \ll 1. \quad (2.6.3)$$

The contributions of the surface and bending terms in the free energy are thus much smaller from the bulk contribution, as can also be seen in Fig. 2.3.

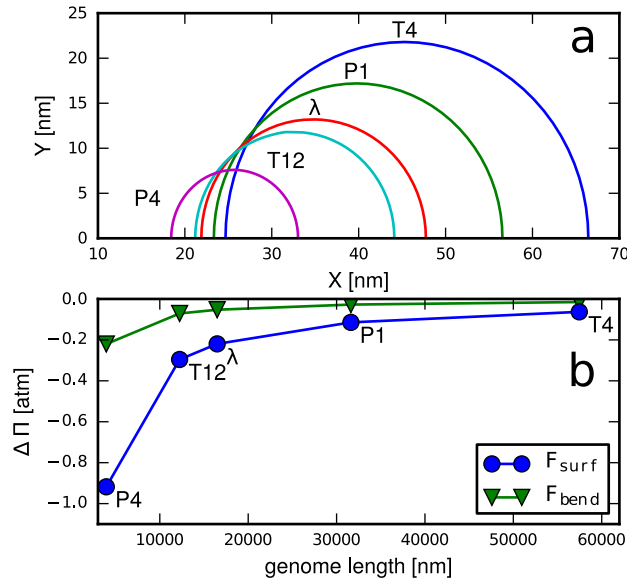
To confirm the wider applicability of this statement we also perform a full minimization for toroids with different lengths of DNA, corresponding to different phage genomes. The minimization of free energy yielded the optimal shape together with the corresponding volume free energy  $F_{vol} = \gamma V$ , surface free energy  $F_{surf} = \sigma S$  and the bending energy  $F_{bend}$  as in Sec. 2.1. The optimal condensate shapes are shown in Fig. 2.12 with the corresponding energy contributions.

## 2.6.2 DNA in the capsid

Assuming repulsive DNA-DNA interactions in the capsid (non-condensed DNA), the total force on the DNA will tend to eject it from the capsid. We will only consider the last stages of ejection, because we want to determine whether it can finish successfully. The free energy inside the capsid is the free energy of a confined semi-flexible polymer with isotropic excluded volume interactions and bending energy:

$$\frac{F_{capsid}}{k_B T} = L^2/s_0 + \frac{1}{2} \frac{L_p L}{R_c^2} \quad (2.6.4)$$

where we take the simplest bending contribution from the DNA barely filling the interior capsid surface. The isotropic excluded volume can be estimated by using Eq. (2.4.4) [13]. We obtain that the excluded volume of one DNA persistence length segment is



**Figure 2.12:** a) Optimal DNA toroid shapes for several different lengths of DNA corresponding to well known phages[54] (See text for details). Only the upper halves of the cross sections are shown. b) The calculated correction to turgor pressure  $\Delta\Pi$  (see main text) for the DNA condensate in the cell as a function of genome length arising from the surface  $F_{surf}$  and bending  $F_{bend}$  free energies. Figure from Ref. [34].

$v \approx V_c L_p^2 / s_0$  with  $s_0 \approx 73 \cdot 10^3 \text{ nm}^2$  a constant depending on the interactions and the capsid size. We take the spherical confinement radius  $R_c = 30 \text{ nm}$  as a representative value for bacteriophage capsid size. This amounts to  $v \approx V_c / 80$  ( $v \approx V_c / 30$ ) for  $L_p = 30 \text{ nm}$  ( $L_p = 50 \text{ nm}$ ) in a 100 mM salt solution.

### 2.6.3 Tug of war

We now study the balance of forces, the *tug of war*, near the end of ejection so that we can determine the maximal cellular turgor pressure that can be overcome. The crowded cellular interior exerts a turgor pressure  $\Pi$  on the volume  $V$  of any foreign material to banish it from the cell [93] while the effect of condensation draws in more DNA to the condensate. The balance of free energy *in the cell* is thus  $F_{cell}(L) = (\Pi - \gamma)A_0(L_0 - L)$  which is always negative if  $\Pi < \gamma$ . The DNA will tend to enter the cell *in spite* of the turgor pressure, due to the favorable condensation conditions. Our estimates for  $\gamma$  are between 4 and 6 atm depending on the choice of model parameters. We take  $\gamma \approx 4 \text{ atm}$  for the turgor pressure that condensing agents in a cell could overcome *on their own*. Additional contributions,  $\Delta\Pi$ , come from unfavourable free energy contributions for DNA in the capsid  $F_{cap}$  and corrections to surface and bending energy terms in the condensate (Fig. 2.13).

In order to better characterize the last stage of ejection, we will study the chemical potential for DNA inside the capsid so that we can relate it to the turgor pressure inside the

cell. The additional chemical potential for the DNA in the cell when the turgor pressure is increased by  $\Delta\Pi$  is  $\Delta\Pi A_0$ , and this should be matched by the chemical potential in the capsid to avoid the stalling of the ejection. When the two chemical potentials are equal

$$\mu = \frac{\partial F_{cell}}{\partial L} = \Delta\Pi A_0 = \frac{\partial F_{cap}}{\partial L} \Big|_{L^*} \quad (2.6.5)$$

the ejection will stall at some length  $L^*$ . From this we determine the maximal *additional*  $\Delta\Pi$  that can be overcome by the virus because the DNA is ejected from the capsid. We now study effects due to the confinement in the capsid which was the cause of the driving force in the early stage of ejection.

Some bacteriophages have tails of considerable length  $t$  so we examine if they influence the ejection process. The entropic penalty for confining a semi-flexible polymer in a tube of diameter  $w$  [94] is

$$\frac{F_{tail}}{k_B T} \approx \frac{t}{\lambda} \ln \frac{L_p}{\lambda} \quad (2.6.6)$$

where  $\lambda = w^{2/3} L_p^{1/3}$  is the Odijk deflection length. For a tail with  $w \approx 2.75$  nm [54] the effective chemical potential from the tail is  $F_{tail}/t \approx 0.27 k_B T/\text{nm}$ . This is enough to oppose an additional 1.5 atm of turgor pressure in the cell and is independent on the length of the tail. However, the effect of the tail onsets only when the last DNA base pair exits the capsid and enters the tail (i.e. when  $L = t$ ). This suggests a barrier in the chemical potential that needs to be overcome for total ejection when  $4 \text{ atm} < \Pi < 5.5 \text{ atm}$  (as will be shown later). In Nature, phage tails have lengths in the range of 10 – 800 nm [8], so the fact that the penalty for confining DNA is independent on the tail length raises questions regarding evolutionary reasons for long tails. We can assume that in addition to ensuring that the DNA end is ready for ejection, and not lost in the packed DNA, it provides some benefit to the phage.

DNA is a charged polyelectrolyte with strong repulsive electrostatic interactions between any two points on its contour. Interactions between nearby parts of the contour act to give it its large persistence length comparable to the radius of the nearly spherical capsid  $L_p \sim R_c$ . When the DNA touches the capsid, any increase of length  $L$  will force the DNA to bend in order to conform to the shape of the capsid. The bending energy in such a situation can be approximated by that of a loop of DNA with radius  $R_c$ ,  $k_B T L L_p / 2 R_c^2$ . A comparison with the previously neglected bending energy of the condensate (see eq. 2.6.1) reveals that they are matched for  $R_c \approx 50$  nm according to the thin torus model [37]. Smaller capsids could enhance the chemical potential, e.g. in the case of  $\lambda$  phage with  $R_c \approx 30$  nm [54], the change of  $\Delta\mu \approx 0.03 k_B T/\text{nm}$  is enough to compensate for  $\approx 0.15$  atm of turgor pressure. Any direct interactions between the DNA and capsid appear to be negligible - viral ejection experiments show no evidence of attractive forces [95] and because dsDNA bacteriophage capsids have no considerable charge [24] only weak van der

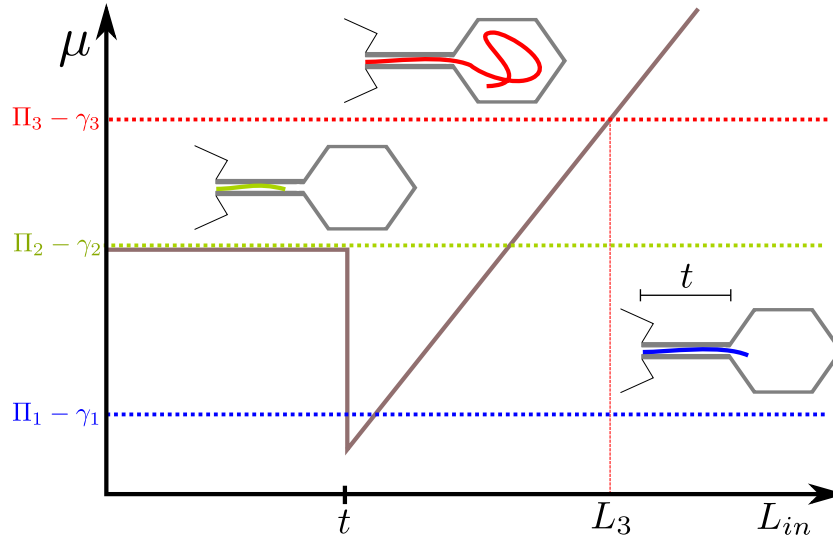


Waals interactions are possible.

When the length of the DNA in the capsid is large enough,  $L \gg L_p$ , the DNA chain statistics resemble that of a random walk of  $n_p = L/L_p$  persistence length segments [61]. This approximation is valid for steric interactions in bulk as long as  $L_p \leq R_c$  [12]. The interaction energy between different parts of the DNA strand in confinement may be estimated on the basis of the excluded volume  $v$  between two segments [12]. The corresponding Flory free energy of interaction in the capsid of volume  $V_c$  is  $F_v \approx k_B T n_p^2 v / V_c$ . This contribution vanishes as  $L \rightarrow 0$  so it can not help the ejection in its latest stages. The excluded volume between two DNA segments can be approximated as that between two charged rods. This interaction is intrinsically anisotropic, but at low packing fractions (near the end of ejection) there is no order and we can average this over all possible mutual angles between two cylinders. We obtain  $v = L_p^2 D C_0$  where  $D$  is the DNA diameter ( $D \approx 2.5$  nm) and  $C_0$  is a numeric constant. The excluded volume interactions will contribute to the total free energy as  $k_B T C_0 D L^2 / V_c$  but only in the regime when there are at least several persistence length segments inside the capsid. For  $R_c \approx 50$  nm in 100 mM monovalent salt,  $v/V_c \approx 1/30$  resulting in the effective chemical potential being an increasing function of length  $\partial F_v / \partial L \approx 2L/L_p 0.013 k_B T / \text{nm}$ . If, say, 10 persistence lengths of DNA are in the capsid the repulsive force is sufficient to oppose  $\sim 0.8$  atm of turgor pressure.

We represent the general dependence of the chemical potential for DNA inside the virus in Fig. 2.13. When the cellular turgor pressure  $\Pi$  is larger than the effective condensing pressure  $\gamma$  in the cell, the net driving pressure  $\Pi - \gamma > 0$  on the viral DNA will tend to repel it from the cell. When the net repulsive cellular pressure is smaller than the tail confinement penalty  $\mu_{tail} = F_{tail}/t$  (corresponding to  $\approx 1.5$  atm) the DNA will be stuck in the virus - the tail of length  $t$  will be completely filled, and some length  $L - t$  will reside inside the capsid. The stalling length is a result of all the repulsive interactions in the capsid cancelling out with the net driving pressure in the cell. Note, however, that if the whole DNA from the capsid (of length  $L - t$ ) enters the tail, the additional asymmetry in the free energy of the two thermodynamic reservoirs onsets. This is due to the confinement penalty of the DNA in the tail. With a partially filled tail, the derivatives of the free energy per unit length (the chemical potential) in the virus (the DNA length increases in the virus) and in the cell (the DNA length increases in the cell) are not the same. The thermodynamical balance is thus broken, and the thermodynamical gradient necessary for the ejection is restored. Therefore, there exists a potential barrier which the DNA needs to overcome for its capsid-side end to enter the tail and be swiftly ejected.

We now estimate whether the thermal fluctuations may overcome the free energy barrier. Because the capsid DNA is not condensed it is coupled to a solvent heat bath at temperature  $T$ . From the equipartition theorem, the encapsidated DNA will have



**Figure 2.13: Chemical potential for DNA of length  $L_{in}$  inside the virus.** Two different regions are marked, one for  $L < t$  when the DNA end is inside the tail of length  $t$ , and the other when the DNA end is inside the capsid. A sketch of chemical potentials in the cell corresponding to different behaviours are indicated.  $\Pi_3 - \gamma_3$  represents the case when the turgor pressure is so large that it causes the ejection of DNA to stall at length  $L_3$ .  $\Pi_2 - \gamma_2$  represents the case when the turgor pressure matches the chemical potential inside the tail, leading to DNA ejection.  $\Pi_1 - \gamma_1$  is similar to the previous case, but when the turgor pressure is much lower than the chemical potential inside the DNA, where the major contribution to ejection comes when the DNA end enters the tail.

$\sim \frac{1}{2}k_B T$  thermal energy per degree of freedom. A semi-flexible polymer of length  $L$  can be partitioned into a random walk of  $n_p \approx L/L_p$  steps with each step of length  $L_p$  having two degrees of freedom (two angles) and the origin being at the tail entrance. The resulting DNA thermal energy is  $\sim 1k_B T/L_p$ , or  $\sim 0.02k_B T/\text{nm}$  which corresponds to a fluctuation in the maximal turgor pressure of  $\sim 0.1$  atm. We can argue that the ejection can happen in a finite time if the barrier corresponds to up to, say, three standard deviations  $\sim 0.3$  atm.

We conclude that the mechanism of viral ejection from some dsDNA bacteriophages into Gram-positive bacteria could be explained as a competition between the resisting turgor pressure and a free energy gain from condensation of the ejected part of the DNA. From the experimental data for the condensation energy of DNA, we estimate that a turgor pressure in excess of 4 atm can be overcome by unassisted ejection in line with recent molecular dynamics simulations [96]. This value is somewhat reduced for smaller phages, as given in Fig. 2.12. Our model does not exclude additional ejecting mechanisms like the osmotic pressure from proteins remaining in the capsid [90] and pulling from proteins in the cell [91] and such mechanisms can help to overcome even larger turgor pressures than obtained here.



## CHAPTER 3

# Nucleic acids and condensing proteins in confinement

A part of the work presented in this chapter has been previously published in:

[97] A. J. Perez-Berna, S. Marion, F. J. Chichon, J. J. Fernandez, D. C. Winkler, J. L. Carrascosa, A. C. Steven, A. Šiber, and C. San Martin, “Distribution of DNA-condensing protein complexes in the adenovirus core,” *Nucleic Acids Res.* **43**, 4274–4283 (2015).

Some viruses have capsids filled only with nucleic acids, while it is more rare to find viruses that have proteins aiding in the condensation of genetic material. Without condensation the cargo of genetic material packed in the capsid can exert pressures sufficient to cause capsid bursting. Condensing agents like proteins facilitate genome packing. Proteins located inside viral capsids are known to have various roles in the viral infection process like scaffolding during assembly [98] and helping maturation [5, 99]. Some viral core proteins have roles in binding DNA, such as baculovirus [100], adenovirus [101], mimivirus [102] and poxviruses [103]. A special case is the polyomavirus type Simian Virus 40 (SV40) which "borrows" cellular histones to pack its DNA into a minichromosome [104]. Adenoviruses and SV40 have been shown to have DNA directly condensed by these nucleoproteins in unperturbed capsids [26, 97]. The details of how DNA and these condensing proteins interact and organize are still unknown as standard experimental techniques can not provide insight due to an apparent lack of viral core symmetry. The basic assumption for decades has been that the viral core has an ordered structure with the same symmetry as the capsid. However, icosahedrally averaged Cry-EM of adenovirus capsids indicates a more or less flat density profile [5, 105, 106], lacking any DNA shelling or ordering like other structurally related viruses [19, 107, 108]. Similarly to adenoviruses, the polyomavirus SV40 cores have a flat density profile seen both in Cryo-EM [109] and SAXS [26], with no pronounced DNA ordering.

The study of genomes inside viral capsids helps understand the viral infection process and the host cell response [6], but also the basic principles for targeted cargo delivery

with viruses or other nanocages [110]. Of special importance is the application of viruses that contain condensing proteins, like adenovirus, in gene therapy [111, 112]. Although adenovirus applications are in clinical trials, the packaging mechanism is still a mystery [5]. To help understand the packing mechanism in viruses, we will study the basic principles of packing a polymer (DNA) and proteins inside viral capsids. First, we will study the internal organization of adenoviruses using the available experimental data on the core organization [97]. With statistical indicators we will characterize the core organization and construct a simple model of condensing proteins to explain the apparent lack of core structure. Later on, we will make a full model with both the polymer (DNA) and condensing agents inside confinement and relate it to experimental findings in adenoviruses and polyomaviruses.

## 3.1 DNA and core protein organization in adenoviruses

Adenoviruses are among the largest non-enveloped icosahedral viruses with a core diameter of  $\approx 65$  nm (see Fig. 1.1) [106]. They are common infectious agent in humans and other vertebrates, being responsible for a wide array of diseases: from mild respiratory infections (the common cold) to life threatening conditions like pneumonia [6]. In adenoviruses the DNA is constrained inside the viral capsid by DNA-condensing proteins [101], which reduce the pressure on the capsid and aid stability [97]. Most of the evidence on internal organization was based on studying disrupted cores, in which clusters and fibres are seen. These clusters and fibres seemed related to the chromosome structure [113] thus implying some kind of order. Two models based on studying disrupted capsids have been proposed for the organization of the adenovirus core. One of these assumed an ordered structure of proteins and DNA forming a cluster of 12 protein "spheres" with icosahedral symmetry [114, 115, 116], while the other a chromatin like structure [117, 118, 119]. Small angle neutron and X-ray scattering indicates structures with a size  $\sim 3$  nm [120]. More recent cryo-EM of undisturbed viral cores does not show any symmetry or order in the viral core [5, 105, 106]. It may be that the original appearance of fibres and clusters with disrupted cores is an artefact of sample preparation [121], or that crowding inside the cores changes the interparticle interactions in an unknown way.

There are no structural data for any of the adenovirus core proteins, but it is assumed that the protein clusters consist of three different core proteins (called V, VII and  $\mu$ ) [5], out of which two are proposed to contribute via universal mechanisms [122] bridging two DNA strands (protein  $\mu$ ) or wrapping DNA (protein VII) [123]. Our working hypothesis, based on analysis of experimental data is that these proteins exist as clusters with a size  $\sim 3$  nm as determined in opened cores. Due to their resemblance to chromosomes, we call

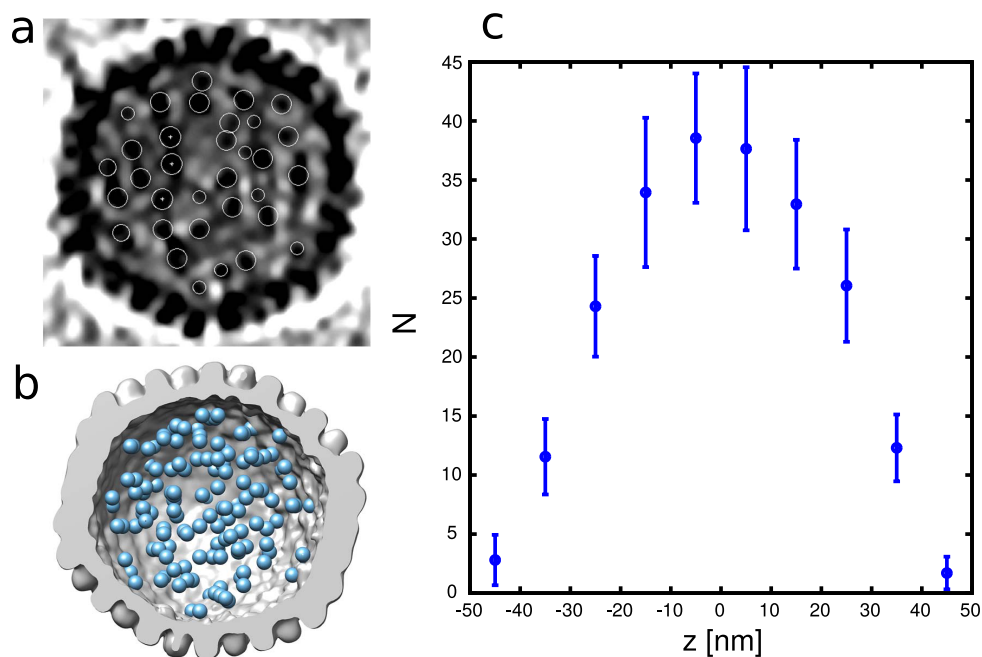
them *adenosomes*. Chromosomes use highly basic histone proteins that wrap the DNA, forming thus protein-DNA complexes called *nucleosomes* which allow the linear dimensions of the DNA to be reduced several orders of magnitude [113]. A similar mechanism of packing exists in polyomavirus SV40 which has  $\sim 20$  nucleosomes in its core, which are found in a "molten droplet" shape caused by confinement effects [26]. The interaction of histones with DNA is well documented, unlike the case of adenosomes.

Adenovirus cores have about 200 protein particles (adenosomes) which are assumed to bind or wrap DNA. We will first discuss experimental data on adenovirus core protein positions. We will attempt to explain the organization of adenovirus core particles using an effective adenosome model in order to give insight on the apparent lack of organization of the core and make estimates on the internal capsid pressure.

### 3.1.1 Cryo-EM of Adenovirus cores

Cryo-EM maps were obtained in the group of Carmen San Martin on single undisturbed adenovirus particles [97]. The interior of the cores was identified as having higher density "points" embedded in a weaker density background (Fig. 3.1a). These points of higher density were interpreted as consistent with previous data indicating a beaded pattern in the DNA-protein core complex when extracted from the capsid [118, 119], and not with the model consisting of 12 large spheres. We hypothesized that each high density region corresponded to an adenosome – similar to the chromatin picture proposed earlier [119]. The position of each adenosome center was manually determined by visual inspection of the individual virus maps in 3D (Fig. 3.1a,b). We found no preferred direction or orientation for the adenosome cluster (Fig. 3.1c). Weaker densities between adenosomes did not seem to follow a definite pattern. Statistical analysis was performed on a total of 20 virion cores. The adenosome selection procedure yielded between 190 and 280 positions per viral particle corresponding to the center of regions of high density within the core, with the mean across all analysed particles of  $N = 230 \pm 30$ , also consistent with previous observations on disrupted cores [118].

Visual inspection of the cryo-EM maps indicates lack of pronounced order or symmetry in the core. To confirm this we first checked if there was any asymmetry in the distribution of adenosome positions. The adenosome coordinates were represented in a coordinate system with the z-axis directed along the line connecting two opposite capsid icosahedral vertices [5], and the origin of coordinate system was positioned at the center of mass for each adenosome cluster (i.e. the set of adenosome positions in each viral particle). Projections of the data on the alignment axis and the perpendicular plane did not show any sign of preferred direction and no sign of five-fold symmetry was found [97].



**Figure 3.1: Cryo-EM of adenovirus core particles analysis.** a) Central 2D section of a single virus particle. The highest density regions are shown in black. Densities identified as adenosomes are encircled in white. Notice that circular outlines in the presented 2D slice have different radii because they are cross-sections of spherical markers used for picking the adenosome centers in 3D, b) Surface rendering showing the adenovirus capsid cut open, obtained from averaging 20 individual virus tomograms after aligning with respect to an icosahedral reference (gray). The core density has been computationally removed. Cyan spheres (4.5 nm in diameter) indicate the positions of adenosomes for the viral particle shown in panel a, c) The histogram of the averaged probability for finding adenosome particles in a horizontal slice, i.e. when all the particle positions are projected on the line connecting two opposite capsid icosahedral vertices[5]. For each viral particle, the origin of the coordinate system was set at the center of mass of the adenosome cluster. Figure adapted from Ref. [97].

### 3.1.2 Statistical analysis of core particle positions

In order to determine if the distribution of adenosomes is random or not, we statistically analyse adenosome positions inside the core "cluster" (Fig. 3.2a-c). First, we constructed radial distribution function  $R_0(d)$  (RDF) for the adenosomes in all viruses. The RDF  $R_0(d)$  represents the probability of finding two particles at a center-to-center distance  $d$ , calculated for all particle pairs in the adenosome cluster. The distributions of inter-particle distances were calculated from a dataset consisting of pairwise distances for all pairs of adenosomes within each viral particle, then averaged across the 20 different viral particle maps. The maximal interadenosome distance  $d$  is about 70 nm, the internal diameter of the capsid. Due to the finite size of the viral core, the number of adenosome pairs becomes smaller with larger distances leading to  $R_0(d)$  to reduce to 0 at the maximal distance of 70 nm. This complicates comparison with bulk systems where with proper normalization  $R_0(d)$  saturates to 1 for sufficiently large  $d$ .

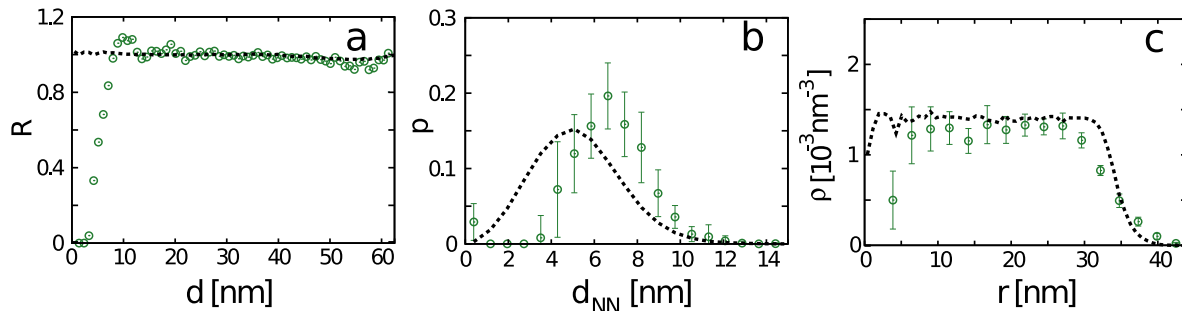
To obtain a quantity characteristic only of the interactions between the adenosomes and without the influence of finite size effects, we need to appropriately scale the distributions. The resulting normalized RDF  $R(d)$  should saturate to unity at large interparticle distances and can then be compared with its well known counterparts in bulk materials [124]. The scaling requires a calculation of the cluster shape factor,  $f(d)$ , which gives less statistical weight to particle pairs with high mutual separation to compensate for the finite sample size. For a spherical cluster of radius  $R_c$  [125] the cluster shape factor is given by

$$f(d) = \left(1 - \frac{d}{2R_c}\right)^2 \left(1 + \frac{d}{4R_c}\right); \quad d < 2R_c. \quad (3.1.1)$$

so that the normalized RDF  $R(d)$  can be obtained from the RDF  $R_0(d)$  using  $R(d) = R_0(d)/f(d)$ . Scaling of experimental data with the shape factor for a sphere produced the desired behavior of RDFs at large distances. When a numerically obtained scaling factor  $f(r)$  for an icosahedron was used there were no noticeable differences. We may thus treat the adenosome cluster shape as a sphere. As a "bonus", the shape factor also enables us to determine the effective radius of the adenosome cluster, since the requirement that the RDF reduces to unity for large interparticle distances fixes the appropriate value of  $R_c$ . Small changes in  $R_c$  cause large deviations in RDF as  $r$  approaches  $2R_c$ , which enables us to pinpoint  $R_c$  with some precision. We thus obtained  $R_c = 35 \pm 2$  nm consistent with experimental data on internal radii of capsids [106]. Note also that the experimental data exhibits pronounced noise as  $d$  approaches  $2R_c$  as there are very small number of pairs available there and the sampling is poor – for this reason the RDFs in Fig. 3.2 are shown only up to about 60 nm as the RDFs for  $d$  larger than 60 nm are dominated by noise augmented by  $f(d)$ .

RDFs indicate an effective repulsive nature of the adenosome-adenosome interaction at





**Figure 3.2: Statistical indicators of the adenosome position dataset.** The three statistical indicators (radial distribution function, nearest neighbour distance distribution and density distribution) constructed from experimental data on adenosome positions are represented by green circles. The icosahedron random points reference is represented by a dashed line. Figure adapted from Ref. [97].

small distances, featuring a characteristic depletion for distances smaller than 8 nm where no pairs are observed. Were the positions of particles random, it would give a constant value of 1 in  $R(d)$  as seen in the icosahedron reference calculation in Fig. 3.2a) obtained from 230 randomly distributed dots in an icosahedron of mid-radius of 35 nm. The slow decay of probability as  $d$  approaches zero indicates a very soft interaction potential [125]. The RDF reduces from about 1 to about 0.5 in a radial interval of 3 nm (from  $d \sim 8$  nm to  $d \sim 5$  nm), and reaches 0.1 at  $d = 3.5$  nm below which there is practically no probability of finding a particle pair. This leads us to define a *hard* particle diameter of  $\sim 4$  nm (effectively impenetrable), and a *soft* diameter of about  $\sim 10$  nm, where the RDFs start to decay. In their study of X-ray scattering on HAdV-2, Devaux *et al.* [120] reported a scattering maximum corresponding to 2.9 nm which could be interpreted as the hard core of our soft adenosomes. One should also note a small-amplitude, yet persistent peak at 10 nm, suggesting a very weak degree of adenosome first neighbour positional correlation as expected for a "fluid-like" state. Beads of 9.5 nm diameter connected by variable lengths of dsDNA had been observed in disrupted cores [119]. This is comparable with the soft diameter estimation obtained here.

We can use the previously obtained adenosome parameters to estimate the degree of crowding inside the capsid. Using the experimental hard particle diameter, the adenosomes have a volume fraction of  $\phi^{hard} = 0.043$ , while the soft particle diameter predicts a volume fraction of  $\phi_{ad}^{soft} = 0.67$  (quite close to the density limit of 0.64 for random packing of monodisperse hard spheres [126]). For comparison, the lower limit for the volume fraction of DNA inside the capsid can be obtained by treating the DNA as a fibre of length  $12\mu\text{m}$  and base-pair width of 2.5nm thus obtaining  $\phi_{DNA} \sim 0.33$ . We conclude that there is little available free volume. The presence of a crowder can significantly influence the effective interactions [27, 127], rendering any comparison with the measurements done on diluted capsid interiors doubtful.

As a second indicator of the core organization we analysed the distribution  $p(d_{NN})$  of nearest neighbour distances,  $d_{NN}$  for all the adenosomes in a virus, normalized so that the total probability is unity. The nearest neighbour distribution gives some additional insight on the RDF for small inter particle distances. As Fig. 3.2b shows, when compared against the random icosahedron reference calculation (effectively an *ideal gas* of adenosome particles), one observes that the mean value of the adenosome distribution,  $\bar{d}_{NN}$  is displaced (outward) by  $\sim 2$  nm from that expected for a random distribution. The position of the maximum in the random distribution for nearest neighbours scales as  $\bar{d}_{NN}^{rand} \sim \sqrt[3]{2/(3N)}R_c$ . The shift of the distribution is a consequence of the repulsive nature of the inter-adenosome interaction which pushes them further apart than would be the case for a random distribution, leading also to a narrower probability distribution of the nearest neighbour distances.

Finally, as the third indicator, we analysed the density of adenosome positions as a function of the distance from the cluster center of mass (Fig. 3.2c). A decrease of density is observed in regions with large radial distances producing a depletion layer known in the polymer field theory [2, 12]. A self consistent calculation for a polymer in the semi-dilute regime, confined inside a spherical cavity of radius  $R_c$ , gives  $\rho(r) \sim \tanh [(R_c - r)/\xi]^2$  with  $\xi$  the correlation length. This profile seems to cover the basic shape of the distribution, however RDF shows no evidence of bonds. In our case the depletion layer is partially a geometric effect with a similar shape to the field theory result, due to reduction of available volume in the icosahedron as the radius increases from that of the inscribed sphere to that corresponding to the circumscribed sphere. This would lead to a  $\xi \sim 1$  nm, much smaller than we would expect. Still, a comparison with the density profile obtained for random points inside an icosahedron revealed additional unexplained depletion effects (Fig. 3.2c). Also, there is a lack of any shelling near the capsid surface which is usually seen with confined polymers [67] and spherical particles [128].

The statistical analysis has revealed several representative features in the experimental dataset. Firstly, a soft repulsive interaction between effective adenosome particles is seen in RDF and nearest neighbour distributions. Secondly, a flat profile in the density of particles which slowly decays to zero at the capsid surface. We will now try to reproduce the internal organization of adenosomes with models which account for different possible ways in which the DNA can interact with adenosomes.

### 3.1.3 Modelling the core proteins

The system we are dealing with consists of condensing proteins interacting with the DNA, confined in the capsid, immersed in a salt solution. Because of a lack of information on the structure of the adenosome particles, we opted for simple models which reproduce basic features of the experimental data while giving some basic insight into the underlying

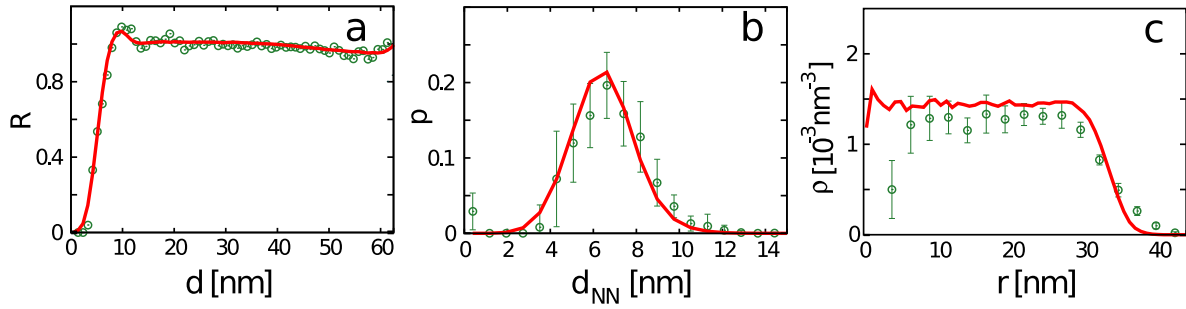
physics. The large quantity of condensed DNA crowds the capsid and for all effective purposes it would be difficult to identify if different parts of DNA are close on the DNA contour. Thus, we treat the DNA as an effective medium which may, in specific models, directionally connect the adenosomes influencing the spatial conformation. This motivates treatment of particles in the simulations as quasiparticles – condensing proteins dressed with parts of the DNA surrounding them. Yet, even with such a simplified picture, we may construct very different coarse-grained models, representing different relevant physical effects involved.

We used molecular dynamics (MD) to simulate the interior structure of the viral capsids. MD was performed in LAMMPS Molecular Dynamics Simulator [129] using a Langevin thermostat [130] with 230 interacting particles in confinement at room temperature  $k_B T = 1$ . Sampling of the MD data was done after a sufficiently long equilibration / thermalization run – 200 simulation snapshots were taken during  $2 \cdot 10^6$  Verlet time-steps and statistically averaged over 50 different runs with random starting conditions. To identify effects related solely to the specific nature of the geometry, in all the simulations presented we have preserved the icosahedral symmetry of the confinement. The confining icosahedron interacts with the adenosome particles via a superposition of a soft repulsive Morse potential and an repulsive potential of the Weeks-Chandler-Andersen type [131]. The range of the hard potential was chosen to be extremely small so as to properly mimic a hard wall, while the soft potential parameters were found by fitting the numerical data to the experimental RDF, maintaining the hard wall fixed. Other specifics of the simulation are detailed in the corresponding sections.

### **Adenosomes as a fluid with short-ranged soft repulsive interactions**

The simplest model represents the adenosomes as a fluid with soft repulsive interactions, confined in a capsid with which it also interacts repulsively. This means that the details of the DNA packing are completely smeared, i.e. the DNA only renormalizes the inter-adenosome interactions. This does not mean that a possible pronounced association of the condensing proteins and the DNA is not accounted for by the model. The model can indeed account for such effects, but only in the short range sense; the protein-DNA association can be included in the effective potential, but no topological constraints related to the finite length of DNA or to its elasticity survive in the coarse-grained representation.

In the soft repulsive fluid model, interactions between adenosomes are represented by a potential function  $v_r(d)$ , where  $d$  is the separation between adenosomes. As a sufficiently simple model for  $v_r(d)$ , we chose the shifted Morse potential and determined the parameters of the potential that best fit the experimental data using molecular dynamics simulations.



**Figure 3.3: Statistical indicators of the adenosome position dataset, and their comparison with expected values for the fluid model.** The three statistical indicators (radial distribution function, nearest neighbour distance distribution, and density distribution) constructed from experimental data on adenosome positions are represented by green circles. The simulations performed with the fluid model are represented by full lines. Figure adapted from Ref. [97].

The potential (shifted Morse) is given by

$$v_r(d) = \begin{cases} D \left[ e^{-2\kappa(d-d_0)} - 2e^{-\kappa(d-d_0)} + 1 \right], & d < d_0; \\ 0, & d > d_0. \end{cases} \quad (3.1.2)$$

where  $d_0$  is the potential cut-off radius. Comparison of the results of the molecular dynamics simulations with the experimental data yielded the best-fit potential parameters  $D \approx 1 k_B T$ ,  $d_0 \approx 11$  nm, and  $\kappa \approx 0.08$  nm<sup>-1</sup>. The soft part of the capsid-adenosome interaction is well modelled with the same cut-off distance but with a smaller  $\kappa \approx 0.055$  nm<sup>-1</sup> (and a short-range hard-core repulsion enforcing the impenetrability of the confinement).

We have also considered other models for  $v_r(d)$ , and we find that the pronounced *softness* of the potential is its robust feature, regardless of the model used. The model of very soft, disconnected quasi-particles reproduces very satisfactorily the radial distribution function (Fig. 3.3a) and the nearest neighbor distance distribution (Fig. 3.3b) observed in experiments. The only (slight) deviation from the experimental data is seen in the density  $\rho(r)$  near the capsid wall at 30 – 40 nm (Fig. 3.3c) - there we see a slower decay of the experimental density than predicted by the simulation. This may indicate that the effective confining potential induced by the capsid has a complicated spatial dependence, rendering the regions just below the icosahedron vertices more approachable to the adenosomes than those below the icosahedron sides. Additionally, we have assumed that the effective inter-adenosome potentials are independent of their position in the capsid. As these potentials are mediated by the DNA background, it seems likely that they will be different when the two adenosomes are close to the capsid than when they are deep in the bulk of the core.

### Adenosomes as a regular array of beads on a string

If the adenosomes are similar to nucleosomes [113], then the appropriate model would consist of quasiparticles linked together by the DNA to form an effective polymer. We may model the DNA "background" in this case as a yardstick, imposing certain distances between the proteins along the chain (equilibrium bond length,  $l_0$ , constant throughout the chain). In such a model, any increase and decrease of the interprotein distance would require energy, the parabolic dependence on the change of distance being the simplest choice. So, in addition to the soft adenosome repulsion,  $v_r(d)$ , that acts between all of the adenosome pairs, in this model the adenosomes are also connected along the chain with a harmonic potential between two neighboring adenosomes. The harmonic potential used to model the DNA connecting the adenosome particles is given by

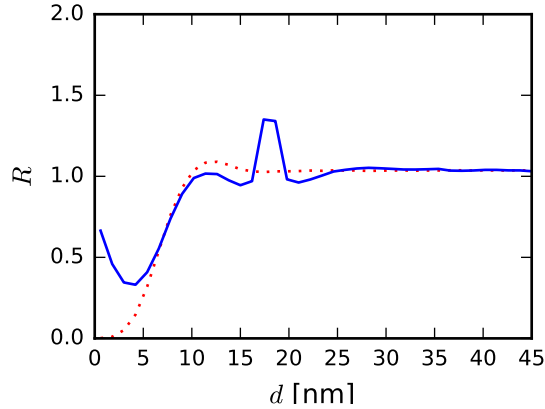
$$v_s(d) = \frac{k}{2} (d - l_0)^2, \quad (3.1.3)$$

where  $l_0$  is the equilibrium length of the spring and  $k$  is the bond spring constant. The equilibrium bond length  $l_0$  chosen in this model cannot be completely arbitrary, since there is a finite amount of DNA in the adenovirus core. We will first estimate possible values for the equilibrium bond length and the bond spring constant  $k$ .

Equilibrium bond lengths  $l_0$  can be estimated based on the ratio of DNA length to number of adenosomes. The upper limit is thus  $l_0^{max} = 12 \mu\text{m} / 230 = 52 \text{ nm}$ , but this does not include any DNA associated with (wrapped around) the condensing proteins. With effective diameter of the condensing proteins  $\sim 4.5 \text{ nm}$ , a single wrap of DNA on the protein uses up about  $2(4.5/2 + 2.5/2)\pi \approx 22 \text{ nm}$ , so that the length of DNA remaining for linking the two adenosomes is  $52 - 22 = 30 \text{ nm}$ . We chose representative bond lengths of 30 nm, 8 nm and 19 nm corresponding to: approximately one turn of DNA around the adenosome; the shortest possible DNA linkage length consistent with the experimental data; and a value in between. The value of 8 nm is consistent with the proposed minimal contact distance between two adenosomes, e.g. as seen in the dimensions of beaded strings in disrupted cores [119] while still shorter bonds significantly modify the exclusion zone in the RDFs.

Assuming a straight linker piece of DNA between two adenosomes, the energetics of its lengthening (DNA stretching) and shortening (DNA bending) can be estimated and approximately related to the bond constant of a simple harmonic bond. We assume that the persistence length of DNA is  $L_p = 50 \text{ nm}$  in physiological conditions. The two most important mechanisms for shortening of the bond length  $d - l_0$  are DNA molecule stretching and bending. The bending energy to bend a straight DNA piece of length  $l_0$  to an arc of radius  $R$  is

$$F_b = k_B T \frac{1}{2} \frac{l_p l_0}{R^2} \quad (3.1.4)$$



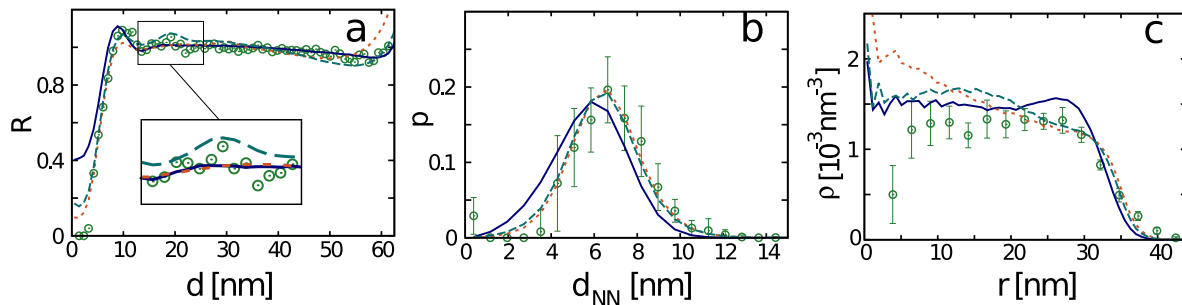
**Figure 3.4: Comparison of radial distribution function without and with a strong harmonic bond.** The radial distribution function is shown for the soft fluid model from Fig. 3.3 (red dashed line) compared to a model of beads on a string for equilibrium bond length  $l_0 = 19$  nm and bond strength  $k = 1 k_B T$ .

with  $l_p \approx 50$  nm the persistence length of DNA. This can be expanded for small changes in the curvature  $1/R$  to obtain:

$$\frac{F_b}{k_B T} = \frac{12l_p}{l_0^2}(l_0 - l) = c(l_0 - l), \quad (3.1.5)$$

where  $l$  is the linear distance between the DNA ending points in the bent state. As the potential energy dependence on the extension  $l_0 - l$  is linear (it is a quadratic function of  $l_0/R$ ) we match the two potentials  $\frac{1}{2}k(\Delta l)^2 = c(L_0 - L)$  for different plausible extensions  $\Delta l = 1, 2, 5$  nm. For  $L_p = 50$  nm and  $l_0 = 19$  nm we find  $k \sim 3, 1.6, 0.7 k_B T/\text{nm}^2$ . In the case of stretching, the bond constant was found to be  $k \sim 10 - 100 k_B T/\text{nm}^2$  after matching a molecular dynamics harmonic potential model to experimental data for DNA [76]. This gives  $k \sim 10 k_B T / \text{nm}^2$  and  $k \sim 1 k_B T / \text{nm}^2$  for stretching and bending respectively, both of which are quite stiff and produce significant correlations in the RDFs as we have checked numerically. This is easily seen if one considers that a bond of strength  $1 k_B T/\text{nm}^2$  would allow thermal energy to move the particle position only  $\sqrt{2}$  nm from equilibrium, resulting in a large correlation in RDF (Fig. 3.4). However, different, and more complicated *effective* bonds can be envisaged in a crowded and strongly confined environment and soft harmonic springs may mimic such a situation. That is why we have chosen the bond constant of  $k \approx 0.05 k_B T / \text{nm}^2$ , a smallest value that still yields noticeable disagreement with at least one experimental indicator. It is also almost two orders of magnitude smaller than the conservative estimate for DNA bending.

Results presented in Fig. 3.5 indicate that the existence of any type of bond worsens the agreement with the experimental data. Soft 8 nm bonds shift the nearest neighbour distance away from the experimental data because a short bond forces spatial nearest



**Figure 3.5: Statistical indicators of the adenosome position dataset, and their comparison with expected values for the soft fluid with springs model.** The three statistical indicators (radial distribution function, nearest neighbour distance distribution, and density distribution) constructed from experimental data on adenosome positions are represented by green circles. The simulations performed with the model of beads on a string are shown for equilibrium bond lengths  $l_0 = 8$  nm, 19 nm, and 30 nm, respectively. Figure adapted from Ref. [97].

neighbours to also be the nearest neighbours on the bead-string backbone. One also observes a stronger correlation peak in the RDF with a corresponding "anti-correlation" peak at  $\approx 15$  nm – although such a bond could be hidden in the experimental RDF it produces clear changes in the nearest neighbour distribution. The density shows a peak near the capsid surface due to an increased order imposed by confining a "polymer" [132]. If we take a larger bond equilibrium length of  $l_0 = 30$  nm we see that the RDF deviates from unity at large interparticle distances – adenosomes do not conform to a spherical cluster. This is easily explained, as a bond length comparable to the capsid diameter requires that the adenosomes explore the space beneath the icosahedron vertices to minimize their free energy. Also, there is a reduction of density around 10 nm from the capsid wall inconsistent with the experimental data (Fig. 3.5c). The model with 19 nm bond lengths shows a combination of these characteristic effects with the addition of a correlation peak in the RDF at the bond length (Fig. 3.5a, inset). Clearly, stiff bonds with large  $k$  show up as clear maxima at  $\sim l_0$  in RDFs due to distance correlations imposed by the bond length. This result that we find in our simulations (with large  $k$ 's) is contrary to what is seen in experiments (no pronounced maxima in RDF), and we conclude that the inter-adenosome bonds, if they exist, must be very easy to stretch and compress (weak) in thermal equilibrium. In that case, the correlations between the adenosome positions are thermally smeared and the linkage, although present may not be seen in the indicators. Irrespective of the equilibrium bond lengths chosen, the model of beads on a string is in worse compliance with the experimental data than the model of fluid of soft interacting particles.

### 3.1.4 Core organization as a mixture of effective particles

Based on our two basic effective models for adenosomes we conclude that there is no evidence of a strict type of ordering in their positions inside the capsid. Adenosomes exist as a fluid of soft particles without a strictly (and stiffly) defined DNA backbone. The DNA appears to act only as an effective medium for the soft inter-adenosome interactions in the crowded environment. The essential feature of adenosomes is a soft repulsive interaction they impose on neighbouring adenosomes, so that they, in lowest order approximation, behave as a fluid of soft repulsive spheres. This interaction results in a finite excluded volume, and the nearest neighbor distance distribution shifted to larger values than would be expected for random points in an icosahedron. The range of the repulsive interaction is quite large, 3 nm, significantly larger than the range of electrostatic interaction in the Debye-Huckel approximation (about 1 nm at 150 mM salt concentration). The simplest model of adenovirus core that accounts for most of the indicators is that of hard spheres (condensing proteins) living in a "soup" of DNA where the DNA mediates an effective soft repulsive interaction.

No evidence of a strict yardstick-like linkage between the adenosomes is seen in any of the experimental indicators - the data can be adequately explained by using the previous model of soft particles in confinement. However, numerical studies [132] do not strictly exclude other types of "links": a) extremely weak bonds (as those we investigated), b) a *distribution* of bond lengths where adenosomes slide along the contour, c) sliding bonds where the total length of non-bound DNA fluctuates - in dynamical terms, this signifies easier rearrangement of adenosomes in the DNA background than would be expected if they were tied together by stiff linker DNA pieces. The existence of adenosome filaments after disrupting the virion [118, 119] does imply some kind of "springs" between the adenosomes. Such interactions in disrupted cores, however, do not necessarily imply the same inside the crowded environment of the core [133].

#### **The outward pressure on the capsid**

Internal pressure has been measured or estimated inside other dsDNA viruses (bacteriophage and herpesvirus) and it appears to originate from the strong repulsion between the nucleic acid strands in close proximity [19, 134]. The soft repulsion between adenosomes gives a modest internal pressure in the adenovirus capsid. The outward pressure on the capsid from the adenosomes in the effective medium obtained from the simulations is  $0.055 \pm 0.002$  atm. It is not clear at present if such a modest outward pressure would play a role in the initial stages of adenovirus uncoating, but it should be mentioned that there are additional contributions to the pressure, not accounted for by the numerical model. Recent estimates using atomic force indentation of capsids [135] give the internal pressure



on the order of  $\sim 30$  atm [123], two orders of magnitude larger. Although the uncertainty of pressures from indentation experiments is high, it would be interesting to see if we can correct for contributions omitted in the model.

The outward pressure obtained in the best fit (soft repulsive) model is obtained on the basis of an effective particle picture. A calculation of the pressure  $p_{tot}$ , including all particles present in the capsid, would in the lowest order (of the virial expansion) consist of contributions from DNA and adenosome entropy of confinement ( $p_{DNA} + p_{ad}$ ), DNA-DNA ( $p_{DNA-DNA}$ ), adenosome-adenosome ( $p_{ad-ad}$ ) and DNA-adenosome ( $p_{DNA-ad}$ ) interactions:

$$p_{tot} \approx p_{DNA} + p_{DNA-DNA} + p_{ad} + p_{ad-ad} + p_{DNA-ad}. \quad (3.1.6)$$

As we do not know the exact composition of the core nor the bare potentials acting between the proteins and DNA, we constructed our soft repulsive model by renormalizing the bare interactions to include the effect of the crowded DNA medium. Thus, the molecular dynamics result for the pressure  $p_{MD}$  includes the adenosome entropic contribution ( $p_{ad}$ ) and *effective* adenosome-adenosome ( $\tilde{p}_{ad-ad}$ ) interactions:

$$p_{MD} \approx p_{ad} + \tilde{p}_{ad-ad}. \quad (3.1.7)$$

This value is a lower bound on the true pressure ( $p_{MD} \leq p_{tot}$ ). Because each adenosome has a certain amount of DNA associated with it (within the soft core radius  $d_0$ ) we assume that all the DNA-adenosome interactions are included in  $\tilde{p}_{ad-ad}$ . Also, from the association of the DNA in the soft core radius with the adenosome we assume that this DNA is screened by the adenosomes. The contributions including the DNA,  $p_{DNA-DNA} + p_{DNA-ad}$ , are to some degree included in  $p_{MD}$ . The only remaining DNA-DNA interactions unaccounted for are in the capsid volume not filled by the (soft) effective adenosomes,  $V_c(1 - \phi_{ad}^{soft}) \approx 1/3V_c$ . Assuming a homogeneous density of DNA in the capsid, we estimate that about one third of the DNA produces a nearly homogeneous background field of electrostatic interactions. The pressure of such field can be roughly estimated in the Debye-Hückel approximation for a homogeneously filled spherical capsid [9]

$$p_e = \left( \frac{Nq}{V_c} \right)^2 \frac{1}{2\kappa^2\epsilon} \quad (3.1.8)$$

where  $N$  is the number of unscreened basepairs of DNA,  $q = 2e$  the number of charges per base pair,  $1/\kappa \approx 1$  nm the Debye electrostatic screening and  $\epsilon \sim 5$  the dielectric constant of an effective medium consisting of proteins and DNA [136]. Putting in  $N = 35000/3$  we obtain  $p_e \sim 0.06$  atm, comparable to  $p_{MD}$ . We can thus increase the lower bound on the adenosome pressure to  $\sim 0.12$  atm. This is still at least one order of magnitude smaller than in the case of unscreened DNA charge interactions, observed in some bacteriophages [19], but not implausible as we have basic proteins helping the condensation of DNA

particles.

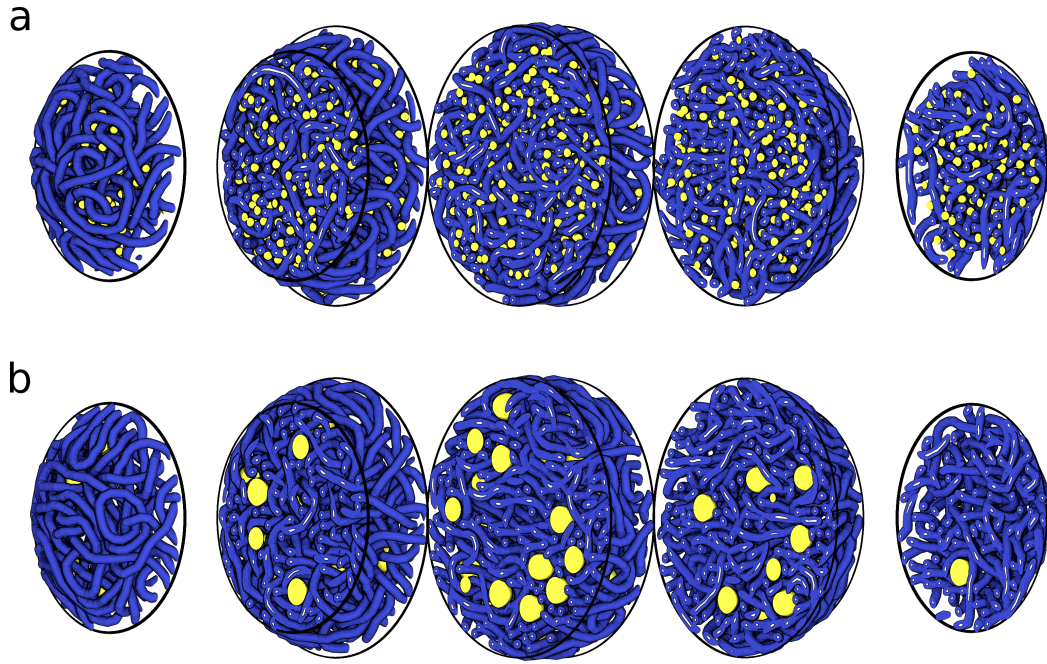
## 3.2 Mixtures of a polymer and condensing particles in confinement

Modelling virion core structures consisting of DNA and condensing proteins has so far been based on effective models. Both our work from the previous section [97] and the work done by Saper et al. on SV40 cores [26] only implicitly accounted for the presence of DNA, included by adding a tethering bond between proteins or by modifying interactions. The phase behaviour of DNA with binding proteins in dilute solutions is interesting [137], but the question remains how it is influenced by confinement or crowding. In this section, we use a model which explicitly includes both the condensing proteins and the polymer (DNA) inside a spherical confinement. We study such a mixture in the context of adenovirus and polyomavirus cores – in confinement and without confinement to mimic opened cores. The explicit treatment of DNA allows us to investigate how the condensers are connected by the DNA. It permits insight in the unordered, but not random, interiors of crowded viral capsids.

First, we will introduce the numerical model used to simulate the system. Afterwards, we will define and adopt several statistical indicators which will be used to study the effects of the parameters of the system on the internal organization and connectivity of condensing proteins and polymer. Results obtained for a DNA-like polymer will be compared to experimental data on adenoviruses [97] and polyomaviruses [26].

### 3.2.1 Molecular dynamics simulations

To sample the possible configurations of our system, we performed molecular dynamics simulations using the LAMMPS programming package [129]. In the simulations we set  $k_B T = 1$  so that the units of energy represent the thermal energy, and we use the radius of the polymer bead  $a_0 = 1$  as the basic unit of length. We study a fixed length of DNA mixed with condensing agents inside a spherical confinement. Our system comprises of two particle types: spheres representing condensing agents and polymer beads on a string, representing DNA or RNA, confined inside a sphere of radius  $R_c$  (Fig. 3.6). The number of polymer beads  $N_p$  of size  $R_p = a_0$  is defined through the volume fraction  $\phi_p$  they occupy inside confinement such that  $N_p = \phi_p R_c^3 / R_p^3 * c_p$  where  $c_p \approx 1.015$  is a correction due to neighbouring beads overlapping. Neighbouring beads are connected with a FENE spring modelling DNA-DNA bonds [76, 138]. Condensing agents (condensers), representing proteins, are implemented as spheres of radii  $R_s$  interacting with each other with repulsive Lennard-Jones interactions. Condensers interact with polymer beads by a short range,



**Figure 3.6: Simulated mixture of DNA and condensing proteins (spheres) in a spherical confinement.** Figure shows five "slices" out of a mixture in spherical confinement (black outline). a) Confined mixture of DNA with volume fraction  $\phi_p = 0.3$  corresponding to  $N_p = 8221$  polymer beads,  $\phi_s = 0.05$ , stiffness  $K = 25$ , condenser-DNA binding energy  $\epsilon = 2k_B T$  and condensing particle radii of  $R_s = 1$  for a total of  $N_s = 1350$  condensers. Condensing particles are represented as yellow spheres, and the DNA polymer as a blue coil. b) Same as in a, except the radii of condensing particles is  $R_s = 3$  for a total of  $N_s = 50$  condensers.

almost contact, non-specific attractive interaction, such that the energy gained in a "bond" is  $\epsilon$ . The attractive part of the potential lies within a layer of width  $a_0$  outside the sphere (condenser) surface. There are a total of  $N_s$  condensing spheres (condensers) related to their volume fraction  $\phi_s$  inside the confinement with  $N_s = \phi_s R_c^3 / R_s^3$ .

Molecular dynamics simulations were performed using the Langevin thermostat [130]. The corresponding equation

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = \mathbf{F}_i - \lambda_i \frac{d\mathbf{r}_i}{dt} + \sqrt{2k_B T \lambda_i} \boldsymbol{\eta}_i(t). \quad (3.2.1)$$

is solved for all particles  $i$  in the system. Here  $\mathbf{F}_i$  represents the total force on the particle  $i$ ,  $m_i$  the mass of the particle,  $\lambda_i$  the friction coefficient,  $k_B$  the Boltzmann constant and  $T$  the temperature. The mass of the particles is dependent on the particle radii  $R_i$  such that  $m_i = R_i^3$  and the friction coefficient as  $\lambda_i = R_i$  according to Stokes law.  $\boldsymbol{\eta}_i(t)$  is a random time-dependent Gaussian  $\delta$ -correlated noise of unity magnitude. The simulation region is confined to a sphere of radius  $\approx R_c$  (representing the viral capsid) by inserting a repulsive potential on at a sphere of radii  $R_c + 1$  so that the probability of finding a

particle with radial coordinate  $r > R_c$  is negligible.

The polymer is modelled with a beads on a spring model [76]. Neighbouring beads in the polymer interact with the finitely extensible nonlinear potential (FENE) representing bonds of the form:

$$U_{FE}(d) = -\frac{1}{2}K_{FE}r_0^2 \ln\left(1 - \frac{d^2}{r_0^2}\right) \quad (3.2.2)$$

where  $d$  is the distance between two neighbouring beads,  $K_{FE} = 30 k_B T/a_0^2$  and  $r_0 = 3.0a_0$ . The stiffness of the DNA (bending rigidity) is represented by a potential depending on the angle  $\theta$  between three neighbouring beads in the polymer. The potential is given by

$$U_b = K_b(1 + \cos\theta) \quad (3.2.3)$$

where  $K_b = 25a_0$  corresponding to a persistence length of  $L_p = 2a_0K_b$  or in our nondimensional units  $L_p = 2K_b$ .

All interactions between particles of the same type (polymer-polymer and sphere-sphere) and the confinement are repulsive only Lennard-Jones with the potential energy

$$U(r) = 4\epsilon_{LJ} \left[ \left(\frac{b}{r}\right)^{12} - \left(\frac{b}{r}\right)^6 \right] + \epsilon_{LJ} \quad \text{if } r < 2^{1/6}b \quad (3.2.4)$$

where  $b$  is a constant equal to a sum of the diameter of the interacting particle  $b = 2R_i$ , or for the case of confinement-particle interactions to  $b = R_i + a_0$ . The cut-off for the potential is such so that the resulting force is 0 at the cut-off. For all particle-particle cases  $\epsilon_{LJ} = 1$ , while for confinement-particle interactions  $\epsilon_{LJ} = 10$  (in units of  $k_B T$ ) is used ensuring that the effective confinement has a radii of  $R_c$ .

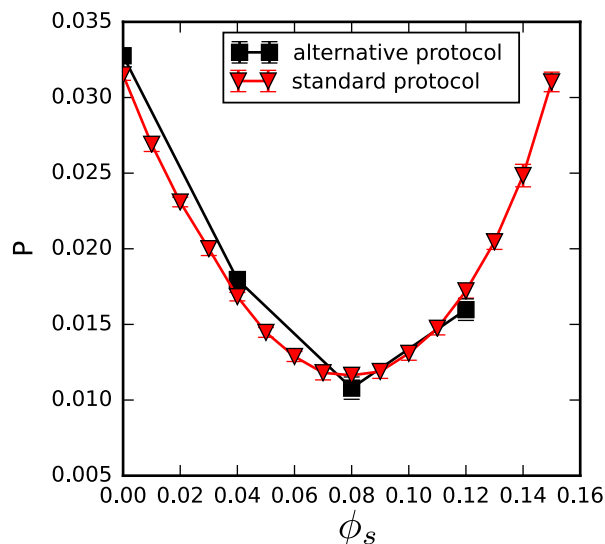
Interaction between condensers (spheres) and polymer beads is of the Lennard-Jones type

$$U_{s-p}(r) = 4\epsilon \left[ \left(\frac{b}{r}\right)^{12} - \left(\frac{b}{r}\right)^6 \right] + \epsilon \quad \text{if } r < 2.8b, \quad (3.2.5)$$

with  $b = R_s + a_0$ . This interaction results in a non-specific binding with the maximum bond energy of  $\epsilon$ . The spatial extent of the attractive potential allows only one layer of polymer beads to interact (attractively) with the condensing sphere.

In order to set up the initial state of the system, we first randomly place condensers inside along with a random walk representing the DNA in confinement. Interactions between all particles are initialized to act as a soft repulsive force with a barrier for penetration (and crossing) of a height of  $100 k_B T$  and harmonic bonds between beads. The total energy of the system was minimized by iteratively adjusting atom coordinates using the provided procedure [129]. Afterwards the system was imbued with true interactions and the total energy minimized again. Then the system was equilibrated for at least

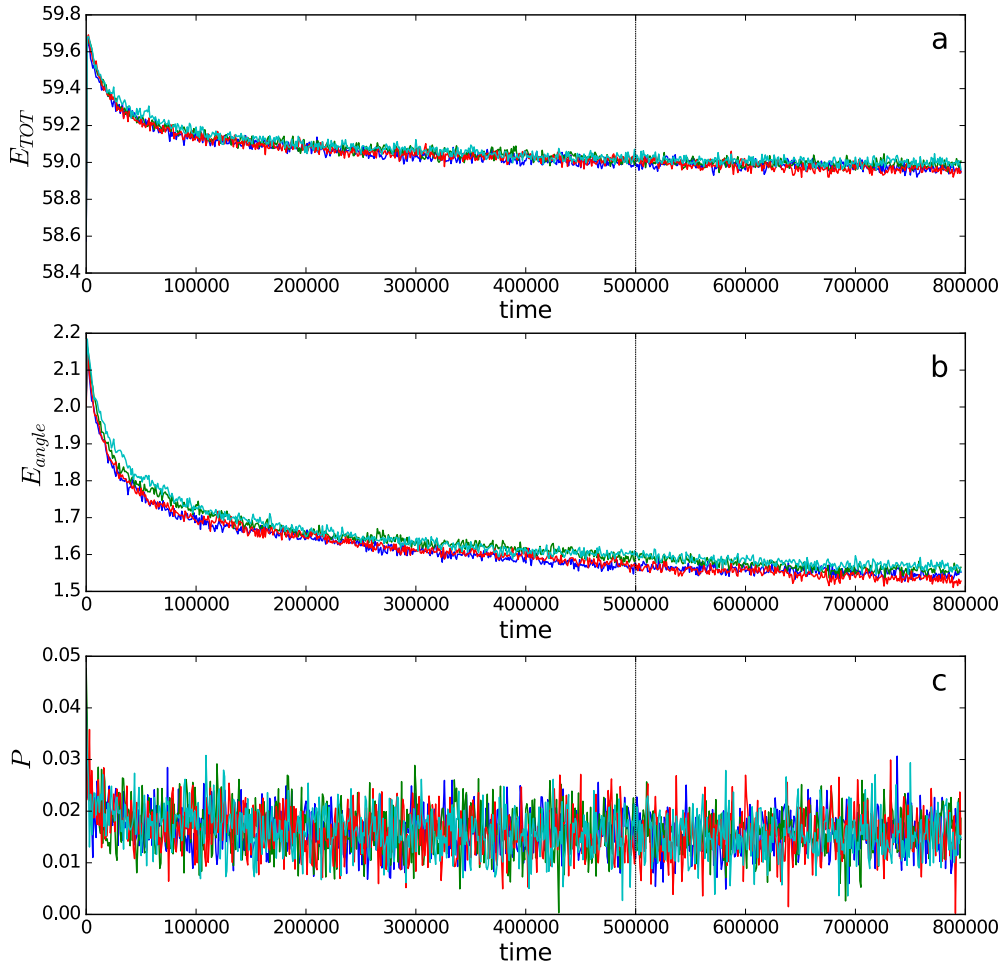
500 000 time steps using a Verlet time integrator with timestep  $dt = 0.05$ . After the system internal energy was equilibrated (Fig. 3.8), the system configuration was sampled every 10 000 timesteps for 300 000 timesteps. An alternative method for generating the starting configuration was tried in order to minimize the polymer knotting or entanglement. The system was generated in a five times larger confinement, and then the confinement radius was slowly reduced to  $R_c$  – but no significant difference was found in the pressures (Fig. 3.7). When studying systems with the confinement removed, the configurations were obtained using the preceding protocol without without any confinement in a bounding box of sufficient size with periodic boundary conditions. After at least 1200000 equilibration steps to allow the released cores to thermalise, the system configuration was sampled every 10 000 timesteps for 300 000 timesteps.



**Figure 3.7: Comparison of standard protocol used here with an alternate aimed to reduce entanglement and knotting.** System parameters were  $\phi_p = 0.3$ ,  $K_b = 20$ ,  $R_s = 1$  and  $\epsilon = 2$  with  $R_c = 30$ .

### Model units and choice of parameters

All subsequent graphs are shown in non-dimensional units:  $\epsilon$  in units of  $k_B T$ , distance ( $K_b$ ,  $R_s$ ,  $r$ ,  $d$ ) in units of  $a_0$ , and pressure in units of  $k_B T/\text{nm}^3$ . This allows one to compare the results for different polymers. The correspondence between our model and (double stranded) DNA is obtained by taking  $a_0 \approx 1$  nm leading to  $\sim 4$  base pairs per polymer bead [76], and  $K = 25$  leading to  $L_p \approx 50$  nm valid for DNA in physiological conditions. For (single stranded) RNA one needs to take  $a_0 = 0.5$  nm, leading to  $\sim 2$  nucleotides per polymer bead with  $K_b = 0$ . The pressures given here can be converted to atmospheres (atm) by multiplying by  $\approx 240k_B T/a_0^3$ . To make comparisons to real systems, we will focus on a capsid radius of  $R_c = 30a_0$  matching the internal mid radii of adenoviruses



**Figure 3.8: Representative equilibration graph for Langevin dynamics.** All data is represented as a function of time with time=0 corresponding to the start of thermalization, time = 500 000 corresponding to the end of thermalization (vertical dashed line), while the time range from 500 000 to 800 000 was used for sampling of configurations. Panel a shows the dependence of the total energy of the system, panel b the energy of polymer bending, and panel c the pressure exerted on the confinement. System parameters were  $\phi_p = 0.3$ ,  $\phi_s = 0.05$ ,  $K_b = 25$  and  $\epsilon = 2$  with  $R_c = 30$ .

( $\sim 32.5$  nm) [106]. We note that polyomavirus SV40 has a comparable capsid radius of 18 nm [139]. We take a DNA volume fraction of  $\phi_p = 0.3$  and condensing agent volume fraction of  $\phi_s = 0.05$  as a representative sample of adenovirus core organization based on considerations in the previous section [97]. Note that we do not include long range electrostatic repulsion, which significantly influences the magnitude for the pressure on viral capsids [9]. Similar coarse grained modelling of SV40 virial cores show that steric interactions are sufficient to explain the internal structure [26]. In our modelling, we see no significant difference in the studied indicators if we include such interactions at a large computational cost as it increases the numerical cut-off radius. This is because electrostatic interactions decay rapidly so that the interaction tail that remains after the Lennard Jones interaction dies out is negligible compared to the thermal energy for

determining the configuration, but it influences the total pressure.

## 3.2.2 Statistical indicators

We will use two statistical indicators to indicate organization of the core: namely the particle probability density  $\rho$  and normalized radial distribution function  $R$ . Additionally, two indicators of polymer-condenser interconnectivity will be used: the wrapping number  $w$  and the connectivity  $\gamma$ . All  $\rho$  and  $R$  given here were averaged first over configurations corresponding to different snapshots in time, and then averaged for different randomised simulation starting conditions. The same was done for  $w$  and  $\gamma$ , with exception that we use the probability distributions of these values  $p(w)$  and  $p(\gamma)$  for each configuration and then average those. We will now proceed to define each of these indicators.

### Particle probability density

The radially symmetric (angle-averaged) probability density  $\rho(r)$  is calculated by determining the probability of finding a particle, at a position in the interval  $\langle r, r + \Delta r \rangle$  from the confinement center. The resulting probability distributions are normalized such that  $\int_0^\infty \rho(r) 4\pi r^2 dr = 1$ . We opt to present the probability density, and not the particle density, as it allows direct comparisons between different condenser sizes, and thus a different particle number in the system.

### Radial distribution function

In order to determine the correlations between positions of condensing proteins, we use the normalized (reduced) radial distribution function  $R(d)$  [125]. The normalized radial distribution function is obtained by taking the radial distribution function  $R_0(d)$  and normalizing to account for the finite size and shape of the "sample" in question. The radial distribution function  $R_0(d)$  is defined as the average number of particles  $n(d)$  found at a distance  $d$  from any particle in the system

$$R_0(d) = \frac{1}{4\pi r^2 N \rho_0} \langle n(d) \rangle \quad (3.2.6)$$

where  $N$  is the number of particles,  $\rho_0 = N/V$  the average particle density found at distance  $d$  from the particle, averaged over all particle pairs in the configuration. The particle density has in itself the volume  $V$  which does not necessarily correspond to the whole volume of the confinement  $V_c$ . In some cases, condensers might not access the whole interior, due to being bound to the polymer and preferring to be away from the repulsive confinement. The function  $R_0(r)$  obtained on a finite sample, in our case a spherical cluster of particles, decays to 0 for  $r = 2R_c$ , due to the largest distance between two particles in

such a cluster being two times its radius. But, in order to be able to compare correlations to bulk samples, or between two samples of different size, one needs to renormalise the size (and shape) effects. The normalized radial distribution function is then defined as  $R(d) = R_0(d)/f(d)$  where  $f(d)$  is the shape factor. The shape factor  $f(d)$  has the property of  $f(0) = 1$  and decays to zero at sufficiently large  $d$ . Shape factors can be obtained in analytical form for most basic shapes with homogeneous density, but in our case although the shape is spherical, we do not have a homogeneous particle density. The shape factor is obtained by normalizing the radial density-density distribution function [125]:

$$f(d) = \frac{1}{v_0} \int \rho(\mathbf{x})\rho(\mathbf{x} + \mathbf{d})d^3x. \quad (3.2.7)$$

$f(d)$  represents the probability of finding two units of density  $\rho(r)$  at a mutual separation of  $d$ .  $v_0$  is a normalization constant ensuring  $f(0) = 1$ .

Our normalized radial distribution functions  $R(d)$  are obtained by first calculating the shape factor from the corresponding particle density function  $\rho(r)$ . We then use the sum rule property of the shape factor [125]:

$$\int_0^\infty 2\pi r^2 f(r)dr = V \quad (3.2.8)$$

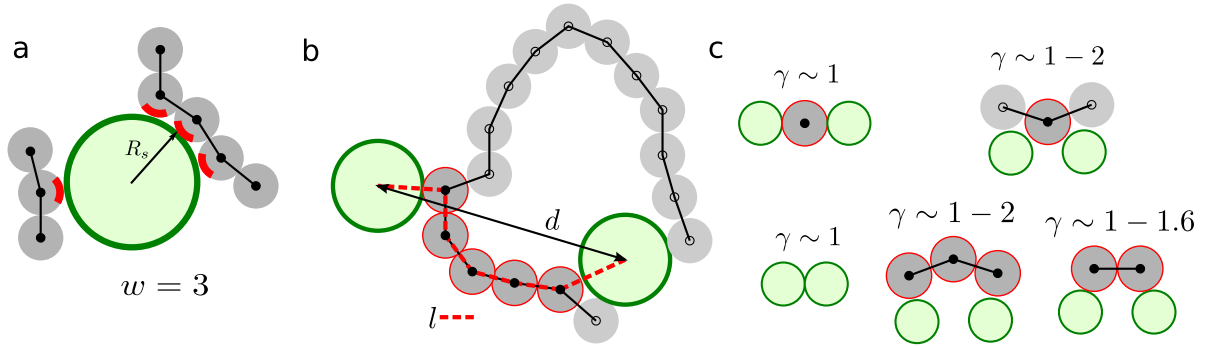
giving us the volume of the sample. This sum rule allows us to recheck the effective radius of confinement in our simulation runs, and we find that the confinement radius is always within 1% of the specified value. The radial distribution function  $R_0(d)$  is then calculated with the obtained true particle volume  $V$  for each configuration, and then averaged over all configurations obtained in simulations. After that, we use the shape factor to obtain the normalized radial distribution function  $R(d) = R_0(d)/f(d)$ .

### Polymer-condenser wrapping and connectivity

In order to study how DNA polymer beads interact with condensers we need to define what constitutes a bead bound to a condenser. A polymer bead at  $\mathbf{r}_b$  and condenser at  $\mathbf{r}_s$  are considered bound if their center to center distance  $d = |\mathbf{r}_b - \mathbf{r}_s|$  is such that the bead-condenser interaction is attractive, i.e.  $(R_s + a_0) - a_0/2 < d < (R_s + a_0) + a_0/2$ . This corresponds to the bead being located in the attractive part of the condenser-DNA interaction potential.

The wrapping number  $w$  of a particular condenser is defined as the longest length of DNA, in number of beads, that is bound to it (see Fig. 3.9a). For each snapshot, we make a probability distribution function  $p(w)$  that any condenser in the current configuration has the longest continuous length of DNA consisting of  $w$  beads. The total probability of such a function is normalized to unity  $\sum_{j=0}^\infty p(w = j) = 1$ . The value  $p(w = 0)$





**Figure 3.9: Sketch of wrapping and connectivity indicator construction.** Polymer beads are shown as grey filled circles, with a black circle in the middle, and with bonds between neighbouring beads as black lines connecting their centers. Green circles without markings in the center represent condensers. a) Two polymers touching a condenser are shown. Red arc on polymer beads show which beads are bound to the condenser. One segment with wrapping 1 and one segment with wrapping 3 are shown, leading to  $w = 3$  for this bead. b) Method of calculating  $\gamma$  is shown. The shortest path connecting the two condensers is marked with darker polymer beads with filled centers. The red dashed line represents the total path between two condensers, while  $d$  marks the shortest distance. c) Representative example of  $\gamma$  values obtained from geometrical constraints.

thus corresponds to the probability that a condenser has no polymer beads bound to it, while  $p(w)$  corresponds to the probability that a condenser has the longest number of continuously wound beads equal to  $w$ . Note,  $p(w)$  does not give any information of the number and distribution of shorter bound segments  $< w$ , or if there are multiple segments of length  $w$ . As such,  $w$  is an indicator of the maximal achieved wrapping length on the condenser.

The connectivity  $\gamma_i = \min_i\{l_i\}/d_i$  is defined as the inverse ratio between the nearest (point to point) distance  $d_i = |\mathbf{r}_a - \mathbf{r}_b|$ , between any two condensers  $a$  and  $b$ , and the shortest DNA contour distance  $\min_i\{l_i\}$ , from a set of possible contour distances  $\{l_i\}$ , if we follow a contour of DNA starting at bead  $j_a$  bound to condenser  $a$  and leading to bead  $j_b$  bound to condenser  $b$  (See Fig. 3.9b):

$$l_i = 2R_s + \sum_{k=\min(j_a, j_b)}^{\max(j_a, j_b)-1} |\mathbf{r}_k - \mathbf{r}_{k+1}|. \quad (3.2.9)$$

As it is possible to have two paths along the contour connecting two condensers, we use  $(j_a, j_b)$  which give the shortest path along the contour  $\min_i\{l_i\}$ . With this definition, two condensers directly touching have  $\gamma = \frac{2R_s+0}{2R_s} = 1$ , as well as two condensers connected by a straight line of polymer beads. Thus,  $\gamma$  indicates if the polymer connecting two condensers is direct ( $\gamma \sim 1$ ), or goes in a sideways or wobbly manner ( $\gamma > 1$ ). For each distinct pair  $i$  of condensers, we find its  $\gamma_i$  values and make a probability distribution  $p(\gamma)$  for each configuration. The resulting probability configuration is then averaged over

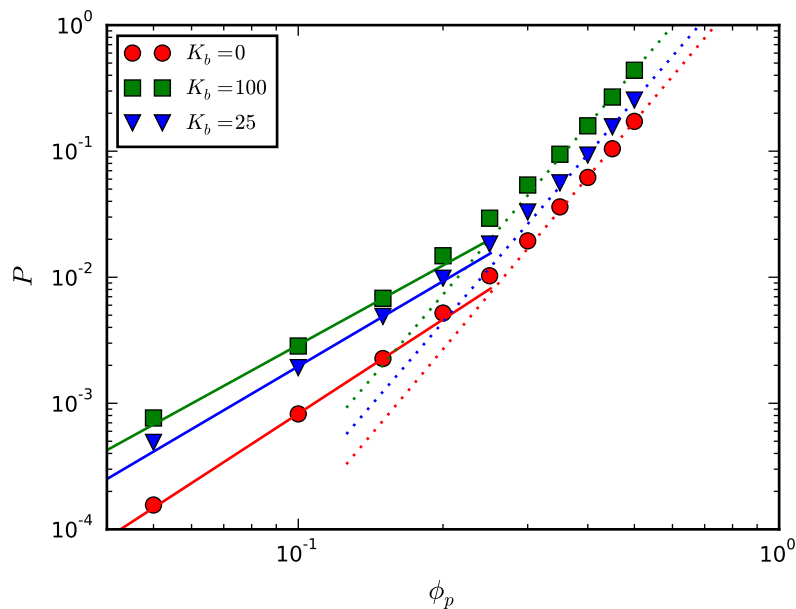
multiple snapshots done over different timesteps and with different initial conditions.

### 3.2.3 Pressures of packing

An important indicator of viral stability is the pressure exerted on the capsid which depends on the amount of packaged DNA [140]. The resulting stress on the capsid is different for various capsid structures [141] and is known to cause capsids to burst [142]. In order to tie our results close to experimentally relevant indicators, we will discuss how the pressure on the capsid  $P$  changes with different parameters before continuing to discuss the internal organization of packaged polymer and condensers. The pressures we obtain do not include a significant quantitative contribution from long range electrostatic interactions. Electrostatic interactions can not compete with steric interactions at high densities due to their fast decay [26], such as those studied here, but are a necessity for predicting viral pressures [9].

Note, all points on pressure graphs in this subsection are an average of at least 12 different initial conditions with 30 different snapshots in time per initial condition.

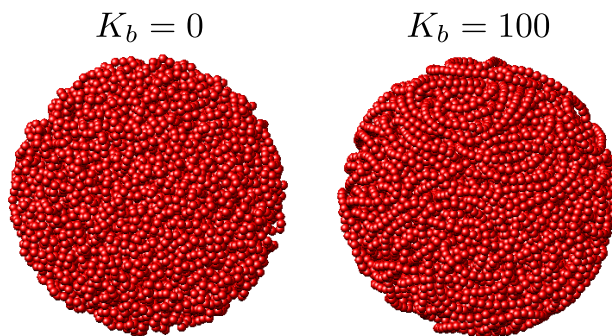
#### Polymer only packing



**Figure 3.10: Pressure from a confined polymer at different volume fractions.**

Pressure as a function of polymer volume fraction for three stiffness constants  $K_b$ . Full lines show a fit to the data at low volume fractions  $\phi_p \leq 0.15$  with a power law  $P \sim \phi^n$  exponent  $n$  of  $2.25 \pm 0.05$ ,  $2.01 \pm 0.06$  and  $2.0 \pm 0.1$  for  $K_b = 0, 25, 100$ , respectively. Dashed lines show a fit to the data at high volume fractions  $\phi_p \geq 0.3$  with a power law exponent of  $4.5 \pm 0.1$ ,  $4.4 \pm 0.1$  and  $4.5 \pm 0.1$  for  $K_b = 0, 25, 100$ , respectively. The radius of confinement was  $R_c = 30$ .

Before studying the influence of condensers on the polymer, we examine how pure polymer behaves inside a confining sphere. We study how the pressure  $P$  exerted on the confinement depends on the volume fraction  $\phi_p$  of confined polymer for different stiffness constants  $K_b$  (Fig. 3.10). Our simulations indicate two regimes, at low and high packing with a crossover at  $\phi_p \sim 0.15$ . At low concentrations, the pressure scales as  $P \sim \phi_p^n$  with  $n$  between 2.0 and 2.25, while at high densities,  $n \approx 4.5$ .



**Figure 3.11: Figure of a confined polymer with different bending stiffness  $K_b$ .** Figure represents one snapshot of a polymer in confinement (the confinement is not visible) at  $\phi_p = 0.2$  and  $R_c = 30$  for different  $K_b$ , as indicated.

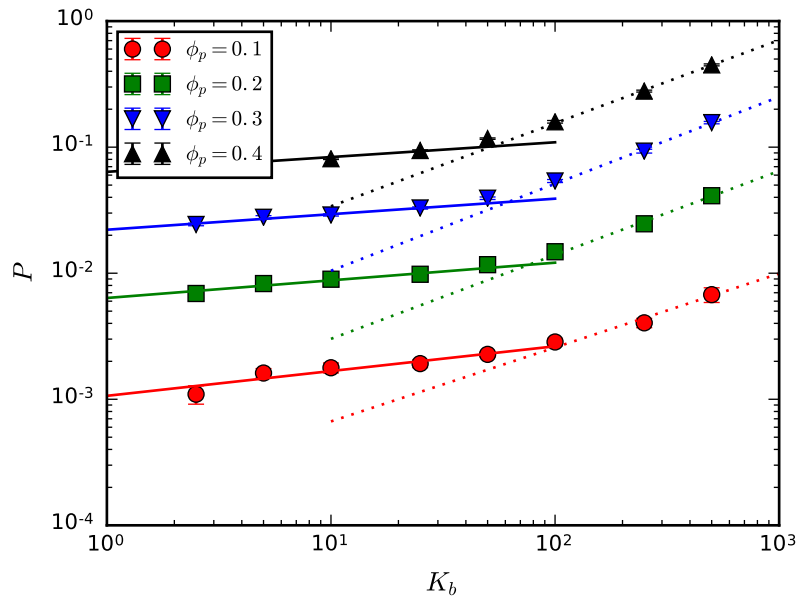
If we define an order parameter  $S$ , such that  $S = \langle P_2(\cos\theta) \rangle$  as in nematic liquid crystals (Sec. 2.5), but with taking into account the radial component of the vector, we obtain  $S = 0$  in all cases confirming a visual representation that the system is globally disordered, albeit there is a degree of local ordering seen with larger  $K_b$  (Fig. 3.11). A transition into a ordered (nematic) liquid crystal has been seen in similar simulations [72], albeit in that study the polymer representing DNA self-interacts with long range repulsive interactions. Additionally, the confinement is gradually filled through the virus tail. It is this gradual filling combined with long range interactions that seems to induce an ordered packing similar to experiments with filling spheres with stiff wire [56]. We also assume that the addition of condensing particles will inhibit local ordering of the polymer due to condensers acting as impurities. This view is consistent with Cryo-EM images of adenoviruses [5] which show no core ordering.

Polymer scaling theory based on the approach by Flory [2] provides scaling behaviours for several polymer regimes. At sufficient densities a flexible polymer is assumed to be in a semi-dilute regime with a scaling analogous to bulk solutions [12]. This semi-dilute regime is characterized by the polymer scaling as if it was surrounded with other (impenetrable) polymers forming an effective cavity. The free energy of a polymer in the semi-dilute regime scales as  $F \sim N_p \phi_p^{1/(3\nu-1)}$ , with  $\nu = 0.588$  the Flory exponent [2, 11]. From  $P = -\partial F / \partial V_c$  we obtain that the pressure scales as

$$P \sim \phi^{3\nu/(3\nu-1)} \quad (3.2.10)$$

or  $P \sim \phi^{2.25}$  (the des Cloizeaux law [2]). Cacciuto and Luijten [11] numerically confirmed that the scaling for polymers confined to a cavity is the same as in bulk semi-dilute solutions. Using Eq. (3.2.10) they obtain from their simulation  $\nu \approx 0.594$  for a polymer consisting of hard sphere beads without bending (their confinement size would be  $R_c = 40$  in our units). If we take our values for  $K_b = 0.0$ , corresponding to a model of RNA, we obtain a value of  $\nu = 0.6$ , also consistent with the Flory exponent. We attribute the small difference in exponents due to different models for polymer bead interactions – softer Lennard-Jones interactions we use give a different excluded volume. For higher  $K_b = 25, 100$  we see a reduction of the exponent to a scaling of roughly  $P \sim \phi_p^2$ . This exponent is the one obtained in the (mean-field) approximation at semi-dilute densities [2, 12], for cases when the excluded volume is dominant, and corresponds to the Flory free energy Eq. (2.4.8) overpowering the bending penalty for confinement.

At high density ( $\phi_p > 0.15$ ), the semi-dilute scaling picture breaks down [11]. For flexible polymers, the correlations between monomers become too large, while for semi-flexible polymers the first virial expansion in density (proportional to the excluded volume) is lacking of higher order terms. Cacciuto and Luijten [11] show that from  $\phi_p = 0.15$  up to the maximum value they test  $\phi_p = 0.3$  the system enters a concentrated regime, and the pressure from numerical simulations scales roughly as  $P \sim \phi_p^3$ . Theoretical mean



**Figure 3.12: Pressure from a confined polymer of different persistence length.** Pressure as a function of polymer stiffness  $K_b = L_p/2$  for different polymer volume fractions  $\phi_p$ . Full lines show a fit to the data at  $K_b \leq 25$  with a power law  $P \sim K_b^n$  exponent  $n$  of  $0.1 \pm 0.2$  for all  $\phi_p$ . Dashed lines show a fit to the data at  $K_b \geq 100$  with power law exponents  $n$  of  $0.6 \pm 0.1$  ( $\phi_p = 0.1$ ),  $0.67 \pm 0.06$  ( $\phi_p = 0.2$ ),  $0.69 \pm 0.05$  ( $\phi_p = 0.3$ ) and  $0.66 \pm 0.03$  ( $\phi_p = 0.4$ ). The radius of confinement was  $R_c = 30$ .

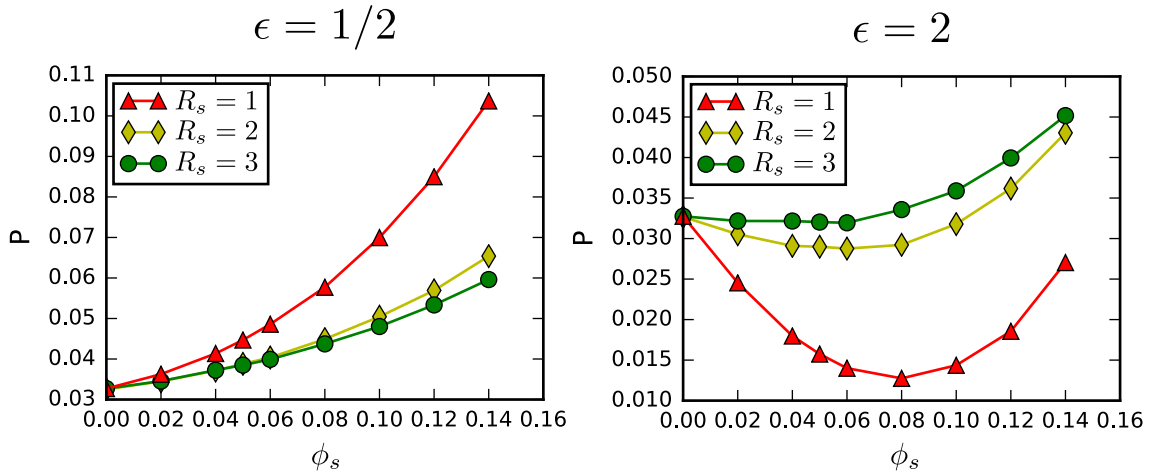
field calculations indicate a scaling of  $P \sim \phi_p^3$  at intermediate densities before entering the concentrated regime [12]. Our data indicate that after  $\phi_p = 0.3$  the system enters a high density regime with a well defined scaling of  $P \sim \phi_p^{4.5 \pm 0.1}$  suggesting a concentrated polymer regime. A degree of caution is needed in this interpretation as the exponent at high densities could be dependent on the choice of the polymer bead interaction potential. It appears that previous works [11, 143, 144] did not show investigate scaling regime comprehensively.

We can also see how the pressure  $P$  depends on the values of the stiffness constant  $K_b$ . From the previous chapter we know that the DNA should have a transition from a bending dominated to a flexible regime at  $L_p = 2K_b \sim R_c$ . Fig. 3.12 shows such a scaling behaviour happening for  $L_p$  between  $R_c$  and  $2R_c$ , as expected [12]. The scaling behaviour in the flexible regime ( $K_b \leq R_c$ ) shows no strong dependence of the pressure on the value of the persistence length  $P \sim L_p^{0.1 \pm 0.2}$ , albeit the pressure does grow weakly with  $L_p$ .  $2L_p$  is the new monomer size of the effective polymer [2]. In the bending dominated regime ( $K_b \geq 100$ ) we see a power law behaviour  $P \sim L_p^{0.64 \pm 0.03}$  inconsistent with both bending models of onion given by Eq. (2.3.23) and inverse spool by Eq. (2.3.21) loop packing which predict  $P \sim L_p$ . There is no direct experimental test of this scaling dependence as far as we know of but the loop models are able to predict scaling at constant  $L_p$  for wire packing in spheres [56]. Some simulations and mean field models also predict  $L_p$  dependence in the bending dominated regime [57], albeit at low densities.

With molecular dynamics simulations we obtain scaling exponents for the pressure of confined polymers. We confirm the existence of a flexible and bending dominated regime as bending stiffness is changed, with a transition at  $2K_b = L_p \sim R_c$ . At different concentrations, we find first the semi-dilute for flexible polymers and the mean-field (Flory) semi-dilute for rigid polymers. At higher concentrations the pressure enters the concentrated regime with the same scaling independent of bending stiffness. Interestingly, in the bending dominated regime we obtain a scaling of the pressure on  $L_p$  inconsistent with any of the theoretical models discussed here (Sec. 2.3).

## Condenser and polymer packing

We now study how different condensers influence the pressure of a confined mixture of condensers with the polymer. Figure 3.13 shows the pressure exerted by the mixture of polymer (DNA) and condensers on the capsid as a function of the volume of condensing particles for different condensing particle radii  $R_s$ . When only a polymer is confined ( $\phi_s = 0$ ) the pressure exerted on the capsid is a growing function of polymer volume fraction  $\phi_p$  (Fig. 3.10). If condensing proteins are added with a condensing energy larger than  $k_B T$ , the addition of condensing materials first reduces the pressure until a minimum is reached for a certain  $\phi_s$ , after which the pressure rises again. At higher volume fractions



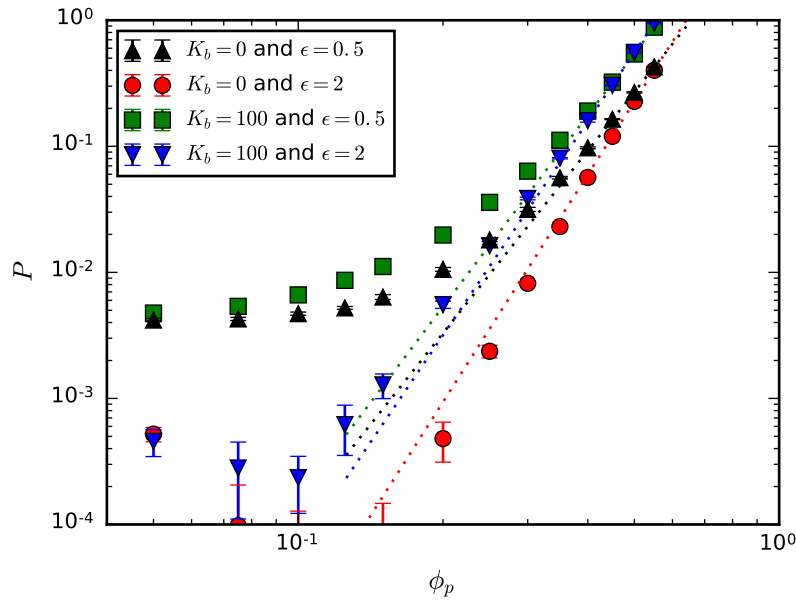
**Figure 3.13: Pressure on confinement for different condenser binding energies.** Pressure as a function of condenser volume fraction for three condenser radii  $R_s$  with condenser binding energy  $\epsilon$  as indicated. Both panels have  $\phi_p = 0.3$ ,  $R_c = 30.0$ ,  $K = 25$ .

of condensers, steric repulsion becomes the dominant force and the pressure increases irrespective of the size or binding strength of the condensers. With weak binding, an increase of the volume fraction of condensing proteins always increases the pressure.

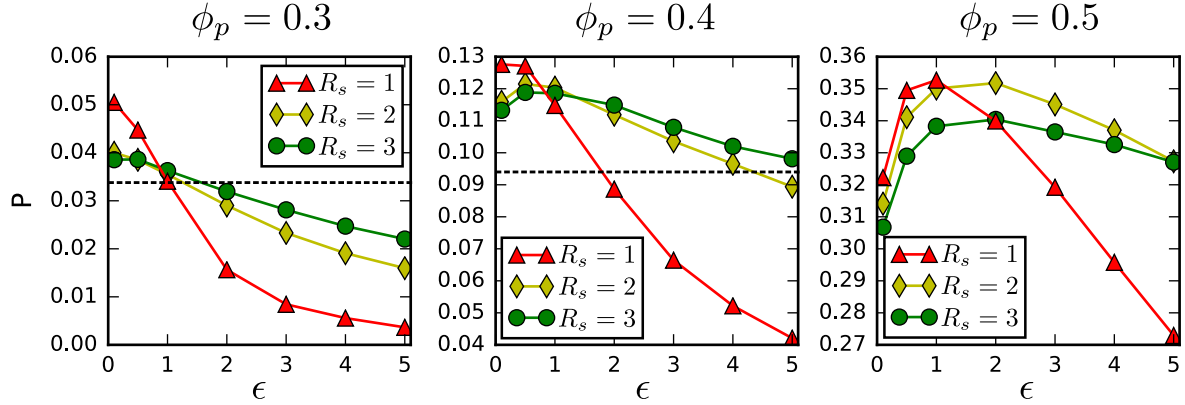
There exists an optimal volume fraction of condensers when the capsid pressure is maximally reduced. With smaller condensers ( $R_s = 1$ ) the addition of condensing agents contributes to lowering the pressure with an optimal concentration (here  $\phi_s = 0.08$ ) where the pressure exerted on the capsid is reduced by more than 50%. For larger condensers, the lowering of capsid pressure is suppressed, leading ultimately to an increase of pressure with added condensers (at  $R_s = 3$ ). The observed behaviour of pressure comes from the total number of binding sites available for DNA-condenser interactions. If the size of the condensing particle  $R_s$  is increased, while the volume fraction is kept the same, the total binding energy that can be realized is decreased. The maximal binding sites that can be realized per condenser is proportional to the surface area  $\sim \epsilon R_s^2$ , while the number of condenser particles  $N_s$  is proportional to  $\sim \phi_s R_s^{-3}$ . This means that the free energy that can be realized from binding goes as  $\sim \epsilon \phi_s R_s^{-1}$ . Thus, if the volume fraction is the same, smaller condensers are able to better connect the DNA and stabilize it because they are capable of permeating the crowded DNA structure better to realize more bonds.

With mixtures, more protein content  $\phi_p$  leads to an increase of pressure (Fig. 3.14). Independent of the binding strength, at large  $\phi_p$ , the pressure has a value dependent only on the bending stiffness  $K_b$ . When crowding becomes dominant, a fixed number of condensers can not contribute to reducing the pressure. The asymptotic value at high  $\phi_p$ , defined by  $K_b$  is approached with a different power law for each  $\epsilon$ .

A comparison of pressure for different condenser-DNA binding strengths at  $\phi_p = 0.3, 0.4$  shows two binding regimes (see Fig. 3.15), one when the binding energy  $\epsilon$  is smaller than



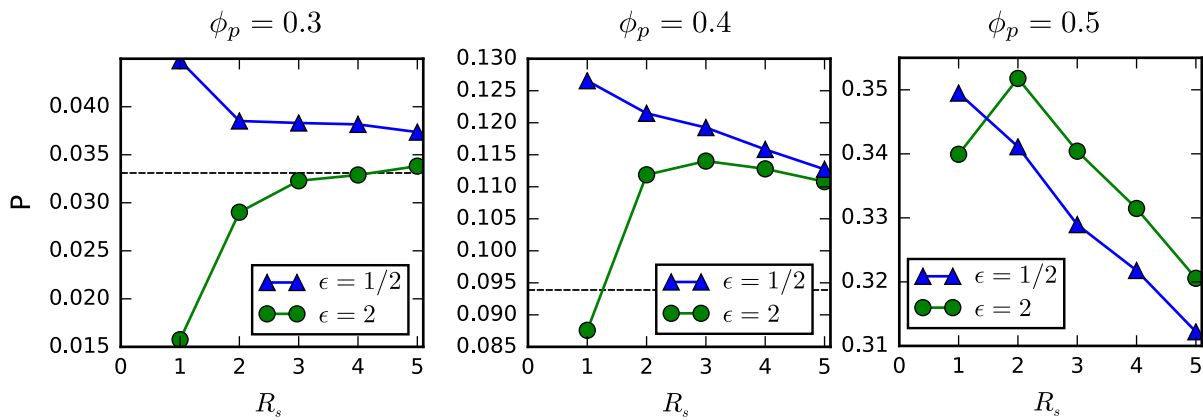
**Figure 3.14: Pressure from condensers at different polymer volume fractions.** Pressure as a function of polymer  $\phi_p$  for different  $K_b$  and  $\epsilon$  at  $\phi_s = 0.05$  with  $R_c = 30.0$ . Lines represent a power law fit  $P \sim \phi^n$  for  $\phi \geq 0.4$  with power law coefficients in the bending regime:  $5.04 \pm 0.08$  ( $K_b = 100$  and  $\epsilon = 0.5$ ) and  $5.63 \pm 0.04$  ( $K_b = 100$  and  $\epsilon = 2$ ), and in the flexible regime:  $4.81 \pm 0.08$  ( $K_b = 0$  and  $\epsilon = 0.5$ ) and  $5.99 \pm 0.03$  ( $K_b = 0$  and  $\epsilon = 2$ ).



**Figure 3.15: Pressure on confinement for different condenser binding energies at three different polymer volume fractions.** Pressure as a function of interaction energy for three condenser radii  $R_s$  at polymer fraction  $\phi_p = 0.3, 0.4, 0.6$ . All panels have  $\phi_s = 0.05$ ,  $R_c = 30.0$ ,  $K = 25$ . Horizontal dashed line represents the pressure without any condensing proteins, except for  $\phi_p = 0.5$  where the pressure without condensers is 0.254.

the thermal energy (weak binding) and one when  $\epsilon$  is larger than the thermal energy (strong binding). The major difference is that for strong binding ( $\epsilon > k_B T$ ) smaller condensers are better in reducing pressure, while for weak binding ( $\epsilon < k_B T$ ) larger condensers are better. In the strong binding regime a single bound bead is thermally stable, while in the weak binding regime more than one is needed for stable binding. In the weak binding regime, we

find that an increase in condenser size reduces the pressure on the capsid. However, in this regime it is still more favourable to have *no* condensing proteins inside the capsid, as any increase of number of condensers increases the pressure. As each condenser contributes to the total free energy less than  $k_B T$  via binding, other contributions become relevant – the entropic contribution to the pressure dominates over the binding part. At higher polymer volume fractions (Fig. 3.15 for  $\phi_p = 0.4$ ), and with the same condensing protein volume fraction, we see the same two regimes (strong and weak binding) but they no longer correspond to a net decrease of pressure. Only at higher binding energies, depending on the size of condenser, does the addition of condensers reduce the pressure. Crowding becomes the dominant contribution, requiring stronger condenser binding to negate it. At even higher volume fractions (Fig. 3.15 for  $\phi_p = 0.5$ ), there exists a maximum in the pressure at about  $\epsilon \sim k_B T$ . Regarding a reduction of pressure, it becomes optimal to either have no binding or extremely strong binding. It seems that bonds comparable to the thermal energy somehow stiffen the structure, enhancing the pressure. Biologically, it seems that only with strong binding with lower crowding do condensing proteins contribute to reducing the pressure exerted on the capsid. Larger weak binding condensers do not inhibit the internal pressure significantly, so their presence is "tolerable" if they are required to facilitate other aspects of the infection process.

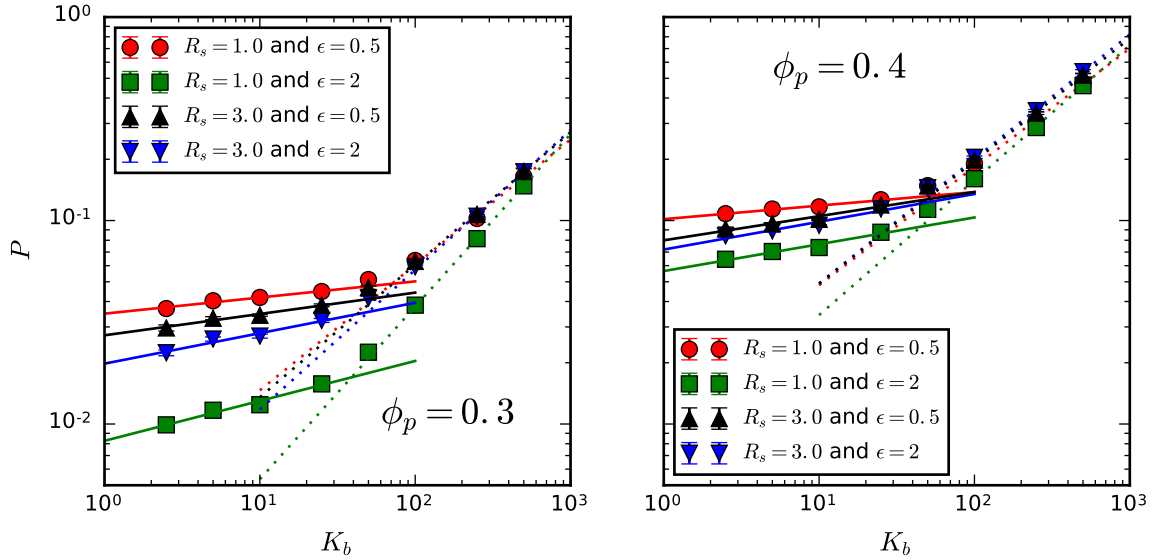


**Figure 3.16: Pressure on confinement for different condenser sizes.** Pressure as a function of condenser size for two condenser binding energies at three different  $\phi_p = 0.3, 0.4, 0.5$ . Dashed lines represent the pressure without any condensers. All panels have  $\phi_s = 0.05$ ,  $R_c = 30.0$ ,  $K = 25$ .

Figure 3.16 shows how the pressure changes with the size  $R_s$  of condensers. Both at  $\phi_p = 0.3$  and  $\phi_p = 0.4$  we see no crossover between pressure lines for different binding energies. With strong binding, we see that the pressure saturates with an increase of  $R_s$ , even resulting in an increase of pressure at around  $R_s = 4$  with respect to the case without condensers. For weak binding, we see a decrease of pressure with an increase of  $R_s$ . A similar trend is seen for  $\phi_p = 0.4$ , albeit only strongly bound  $R_s = 1$  decrease the



total pressure. A decrease in pressure with  $R_s$  in the weak binding can be explained as being influenced by a decrease of entropic pressure, which is proportional to the number of particles. On the other hand, in the strong binding regime the pressure saturates at high  $R_s$ , likely as the total energy of binding,  $\sim \epsilon \phi_s R_s^{-1}$ , is reduced at high  $R_s$ . At  $\phi_p = 0.5$  larger and strong binding condensers increase the pressure with respect to weak binders.



**Figure 3.17: Pressure from a confined polymer with condensing proteins for different polymer stiffnesses.** Pressure as a function of polymer stiffness  $K_b = L_p/2$  for two different volume fractions of  $\phi_p = 0.3$  (left panel) and  $\phi_p = 0.4$  (right panel) at  $\phi_s = 0.05$  with  $R_c = 30.0$ . Full lines show power law fits with a scaling exponent  $0.1 \pm 0.1$ . Dashed lines show a power law fit with exponents: a) for  $\phi_p = 0.3$  we have exponents  $0.62 \pm 0.05$  ( $R_s = 1.0$  and  $\epsilon = 0.5$ ),  $0.85 \pm 0.01$  ( $R_s = 1.0$  and  $\epsilon = 2$ ),  $0.62 \pm 0.05$  ( $R_s = 3.0$  and  $\epsilon = 0.5$ ) and  $0.69 \pm 0.03$  ( $R_s = 3.0$  and  $\epsilon = 2$ ), b)  $\phi_p = 0.4$  with exponents  $0.58 \pm 0.03$  ( $R_s = 1.0$  and  $\epsilon = 0.5$ ),  $0.66 \pm 0.02$  ( $R_s = 1.0$  and  $\epsilon = 2$ ),  $0.60 \pm 0.01$  ( $R_s = 3.0$  and  $\epsilon = 0.5$ ) and  $0.61 \pm 0.02$  ( $R_s = 3.0$  and  $\epsilon = 2$ ).

The addition of condensing proteins will also influence the dependence of the pressure on the polymer stiffness. A comparison of scaling with different parameters (Fig. 3.17) indicates that the bending dominated regime and the flexible polymer regime persist in spite of added condensers (Fig. 3.12). In all cases, a higher bending stiffness increases the pressure and makes binding ultimately inefficient in reducing the pressure. In the flexible regime ( $2K_b < R_c$ ) there is no well defined scaling on the persistence length  $L_p$ , just a monotonous rise in the pressure. In the bending regime at lower crowding ( $\phi_p = 0.3$ ) we can argue that there are two possibilities. Weak binding and large condensers produce a scaling  $P \sim K_b^n$  with the exponent  $n \sim 0.6$  similar to the scaling with only polymer (Fig. 3.12). If the condensers have a strong binding energy ( $\epsilon > 1k_B T$ ), the scaling relationship is changed. In order to approach the same high  $K_b$  pressure value, the scaling on  $K_b$  in the bending regime approaches an exponent of 1. At higher crowding ( $\phi_p = 0.4$ ), the pressure reduction with strong and small condensers is less relevant, and the resulting asymptotic

scaling exponent is also smaller. Thus, strong condensers are more influenced by a stiff polymer backbone, initiating a faster return to a pure (repulsive) pressure value.

We conclude that the presence of condensing agents with a polymer in confinement influences the total pressure exerted on the confinement. There exists an optimal condenser volume fraction which minimizes the pressure, but only with strong binding and sufficiently small condensers able to realize many bonds with the polymer. At higher polymer densities, the crowding effect overcomes the reduction of pressure from condensers. We find two regimes in both volume fraction of packaged cargo and stiffness of the polymer – an increase in concentration brings a change from a semi-dilute to a concentrated regime, and an increase in bending stiffness changes the system from a flexible polymer to the bending dominated regime.

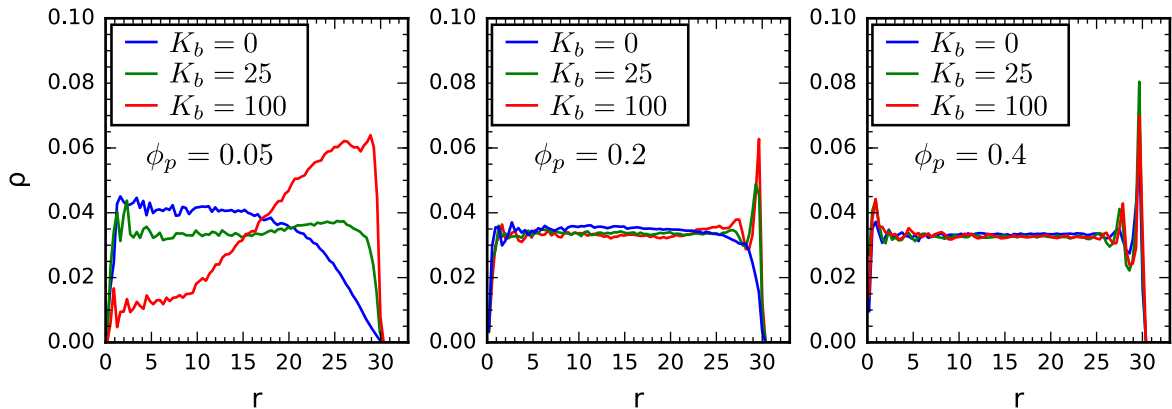
### 3.2.4 Internal organization

We will now study how different parameters influence the internal organization of confined mixtures of condensing proteins and polymer. Experimentally it is possible to probe both the density and radial distribution function for condensing proteins using either electron microscopy [97, 121] or small angle X-ray scattering [26]. Our aim is to study how parameters of the model influence these experimentally accessible indicators. Additionally, our inclusion of an explicit polymer allows us to study how crowded and confined mixtures of polymer and condensers are interconnected. Based on the analysis of the pressures we will concentrate on confinements matching the typical size of viral capsids, with strong and weak binding, and both flexible and stiff polymers. Afterwards, the implications of these findings will be discussed in the context of virology and artificial nano-cargo delivery.

All statistical indicators presented here have been averaged for at least 48 random initial states of the system, with 30 snapshots in time per each initial state.

#### **Polymer only packing**

Packing of polymers inside spherical confinement is a well studied problem [11, 143, 144], examined also in the context of DNA packing and ejection in bacteriophages [72, 79, 145, 146], allows us to benchmark our model before continuing studying mixtures with condensing particles. Fig. 3.18 shows how the probability density for polymer beads depends on a choice of stiffness constants at different polymer volume fractions  $\phi_p$ . At low densities ( $\phi_p = 0.05$ ) the polymer stiffness has a profound influence on its density profile. A flexible polymer ( $K_b = 0$ ) fills the interior and has a depletion layer at the walls of the capsid, the same as obtained in polymer field theory, with a density profile  $\rho(r) \sim \tanh^2[(R_c - r)/\xi]$  for  $r < R_c$  where  $\xi$  is the mean field correlation length dependent



**Figure 3.18:** Probability density for different stiffness constants  $K_b$  and polymer volume fractions  $\phi_p$  for  $R_c = 30.0$ .

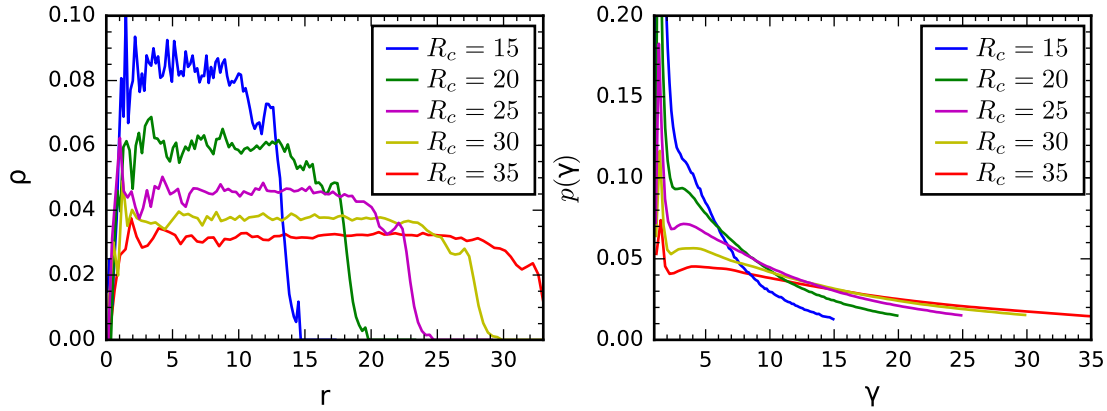
on the volume fraction  $\phi_p$  [12]. On the other hand, a stiff polymer ( $K_b = 100$ ) in the bending regime has an exclusion hole in the center of the confinement, the bending stiffness forcing it to cover the interior of the confinement surface to minimize its bending energy. The case with  $K_b = 25$ , corresponding to the stiffness of DNA when  $a_0 = 1$  nm, lies in between these two regimes, with the exclusion zone starting to form. Although DNA lies outside the bending dominated regime, its density profile shows similarities to profiles of higher bending stiffness.

At intermediate densities ( $\phi_p = 0.2$ ), just beyond the semi-dilute regime, we see that the polymer fills the whole volume of the confinement. A flexible polymer ( $K_b = 0$ ) still has a depletion layer near the confinement, but it is barely visible. Stiffer polymers are located with a high probability in well defined layers near the surface of the confinement with a higher stiffness giving a higher probability. This layering is a direct consequence of interactions with the confining surface, and is a known effect both in confined polymers [67] and spheres [128]. At higher densities, the difference between polymers of different stiffness becomes less pronounced, as the lack of free volume and the topology of the chain determines the packing density.

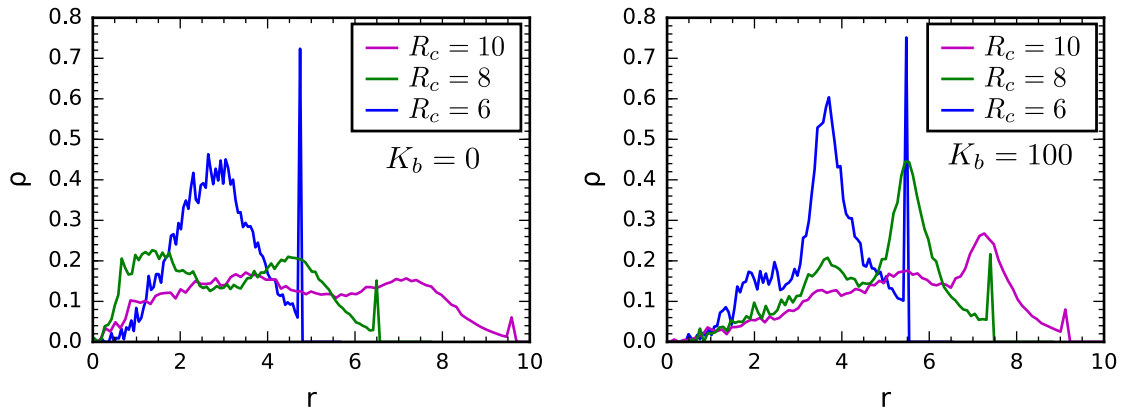
### Confinement size effects

The size of confinement can influence the internal organization of the system. Our primary interest lies with large confinement radii which we can compare to viruses. For example, adenoviruses filled with DNA would correspond to  $R_c = 30$  nm with  $a_0 = 1$  nm. But, smaller confinements, e.g. representing protein cages, with  $R_c \sim 8$  nm are also interesting for applications in nano-cargo delivery [110]. We first test large confinement radii, which would correspond to both adenoviruses and polyomaviruses.

Fig. 3.19 shows the dependence of the particle probability density  $\rho$  and connectivity



**Figure 3.19: Structure variation for different confinement radii  $R_c$ .** Particle probability density  $\rho$  and connectivity  $p(\gamma)$  are shown for  $\phi_p = 0.3$ ,  $\phi_s = 0.05$ ,  $\epsilon = 2$ ,  $K_b = 25$ .

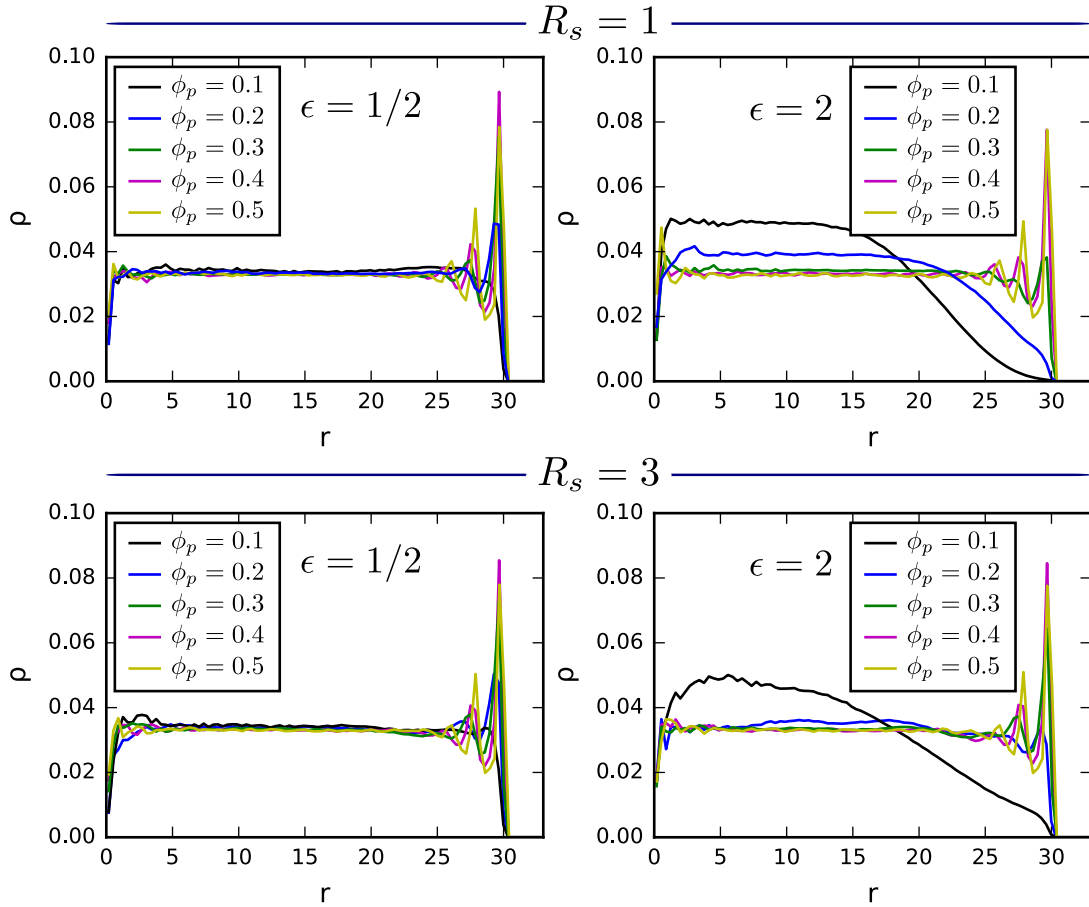


**Figure 3.20: Probability density for different small confinement radii  $R_c$  for a flexible ( $K_b = 0$ ) and stiff polymer ( $K_b = 100$ ).** Particle probability density  $\rho$  and connectivity  $p(\gamma)$  are shown for  $\phi_p = 0.3$ ,  $\phi_s = 0.05$  and  $\epsilon = 2$ .

$\gamma$  for different confinement radii comparable to viral capsids. The density  $\rho$  shows the same general profile in all cases, but  $\gamma$  shows a difference between different  $R_c$ . A larger probability for finding two condensers directly connected, i.e. a small  $\gamma$ , is seen in small confinements. This is expected as the complexity of paths is smaller with less DNA connecting the system. When the confinement becomes comparable to the size of the particles in the simulated mixture (Fig. 3.20) one begins to see changes in the density due to finite size effects, but also subtle changes in the radial distribution function. As our primary interest lies in viral capsids, we will not discuss this case further.

### Internal organization at varying polymer density

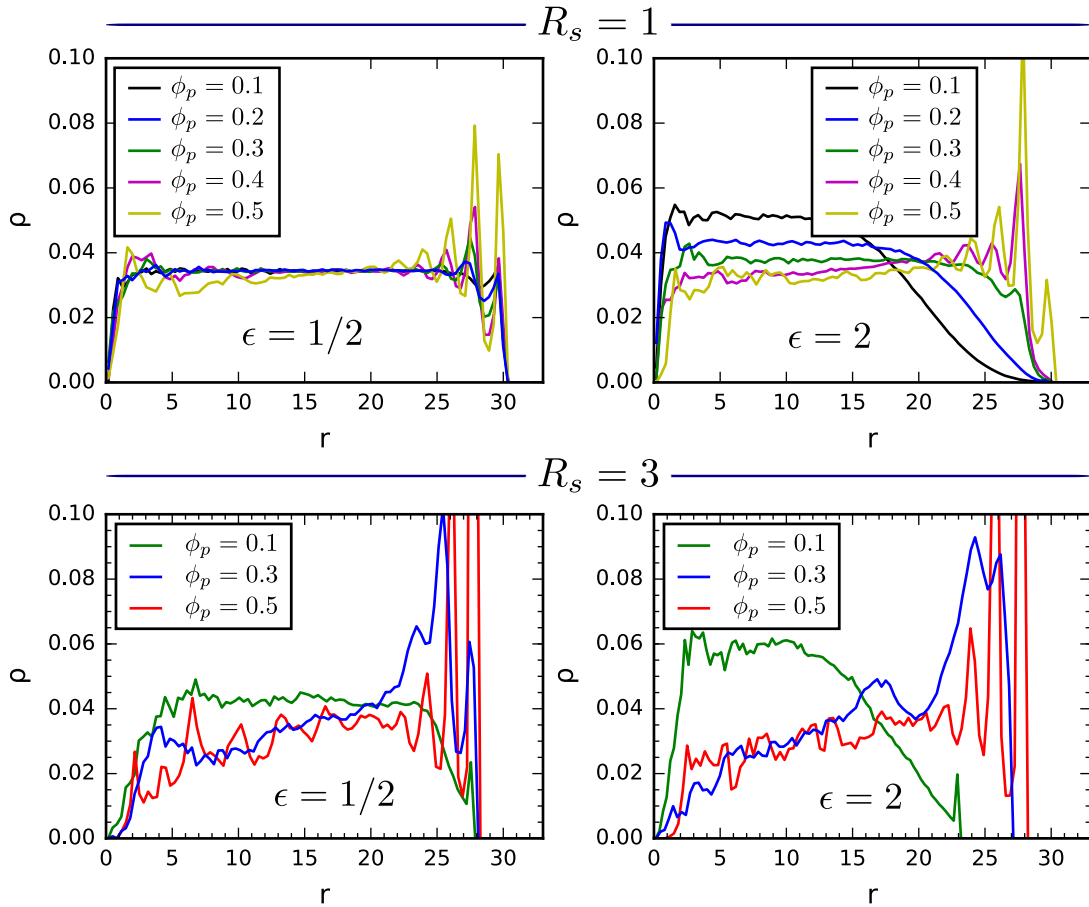
Figure 3.21 shows the polymer density profile for various polymer volume fractions with the stiffness corresponding to that of DNA ( $K_b = 25$ ) with added condensing proteins ( $\phi_s = 0.05$ ). The general trend with all condenser sizes is that the presence of condensers



**Figure 3.21: Polymer density profile with added condensers for different polymer volume fractions.** Data is shown for stiffness  $K_b = 25$  (DNA) with condenser volume fraction  $\phi_s = 0.05$  confined in a sphere of radius  $R_c = 30$  for different combinations of condenser radius  $R_s = 1, 3$  and condenser binding strength  $\epsilon = 0.5, 2$ .

does not influence the density in a major way if the condensers are weakly binding. When condensers bind strongly ( $\epsilon = 2k_B T$ ), they produce a more compact structure at low densities, albeit still filling most of the available free volume. Larger condensers at low densities produce a more spread out polymer density. In all cases layering in the area near the confinement surface is observed at sufficiently high densities.

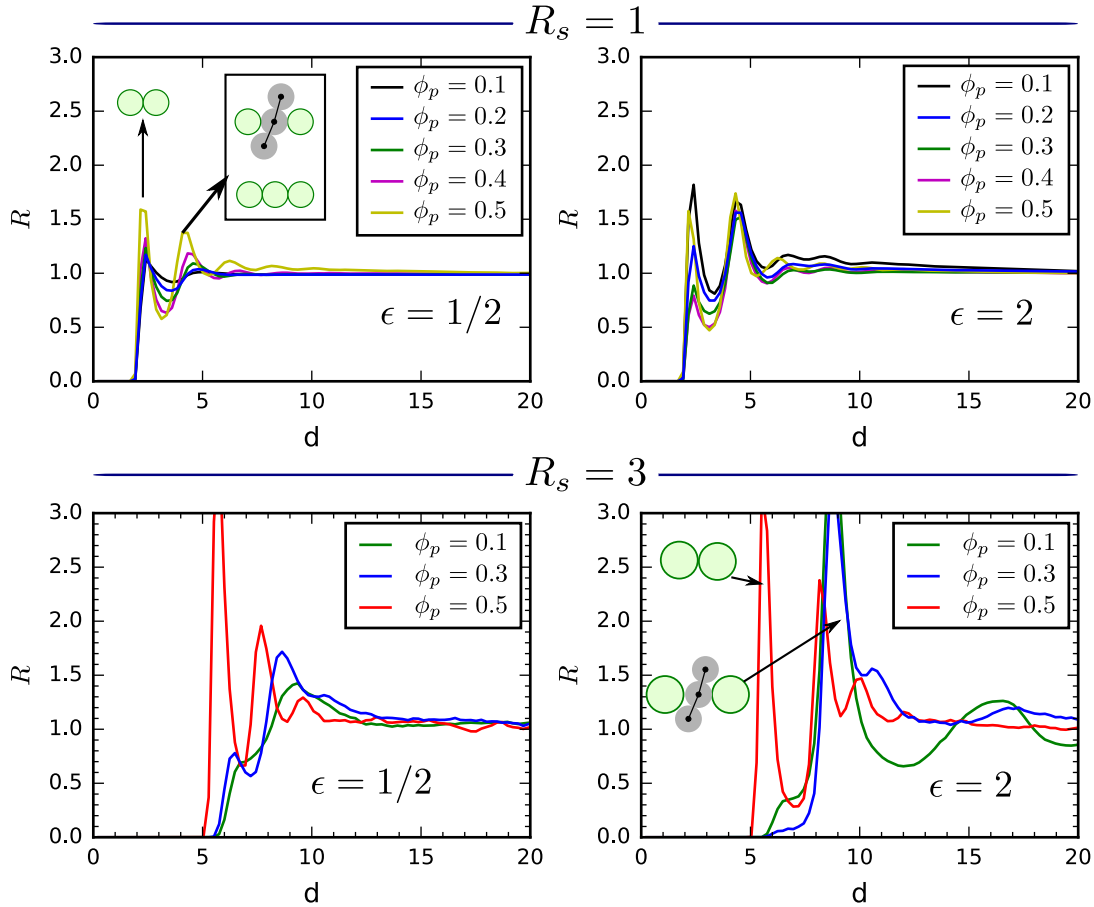
The density distribution of condensers (Fig. 3.22) exhibits similar behaviour as the polymer density. When the binding is weak and the condensers are small, the condenser densities follow similar trends as the polymer densities. Strong binding shows a different profile with the cluster of condensers and polymer becoming more compact without any significant ordering near the confinement until  $\phi_p = 0.45$ . A reduction in the pressure (Fig. 3.14) is correlated with more compact structures. The first layer of condenser density at the confinement surface is suppressed in comparison to the second (interior) layer, especially with strong binding. We argue this is a combination of two effects: a) the polymer is pushed outwards from the center of the confinement due to its bending



**Figure 3.22: Condenser density profile for different polymer volume fractions.** Data is shown for stiffness  $K_b = 25$  (DNA) with condenser volume fraction  $\phi_s = 0.05$  confined in a sphere of radius  $R_c = 30$  for different combinations of condenser radius  $R_s = 1, 3$  and condenser binding strength  $\epsilon = 0.5, 2$ .

rigidity, and b) if condensing particles constitute the outermost layer, they have a reduced number of polymer bead neighbours thus reducing the opportunity for binding. With larger condensers ( $R_s = 3$ ) configurations show depletion of probability near the center of the confinement.

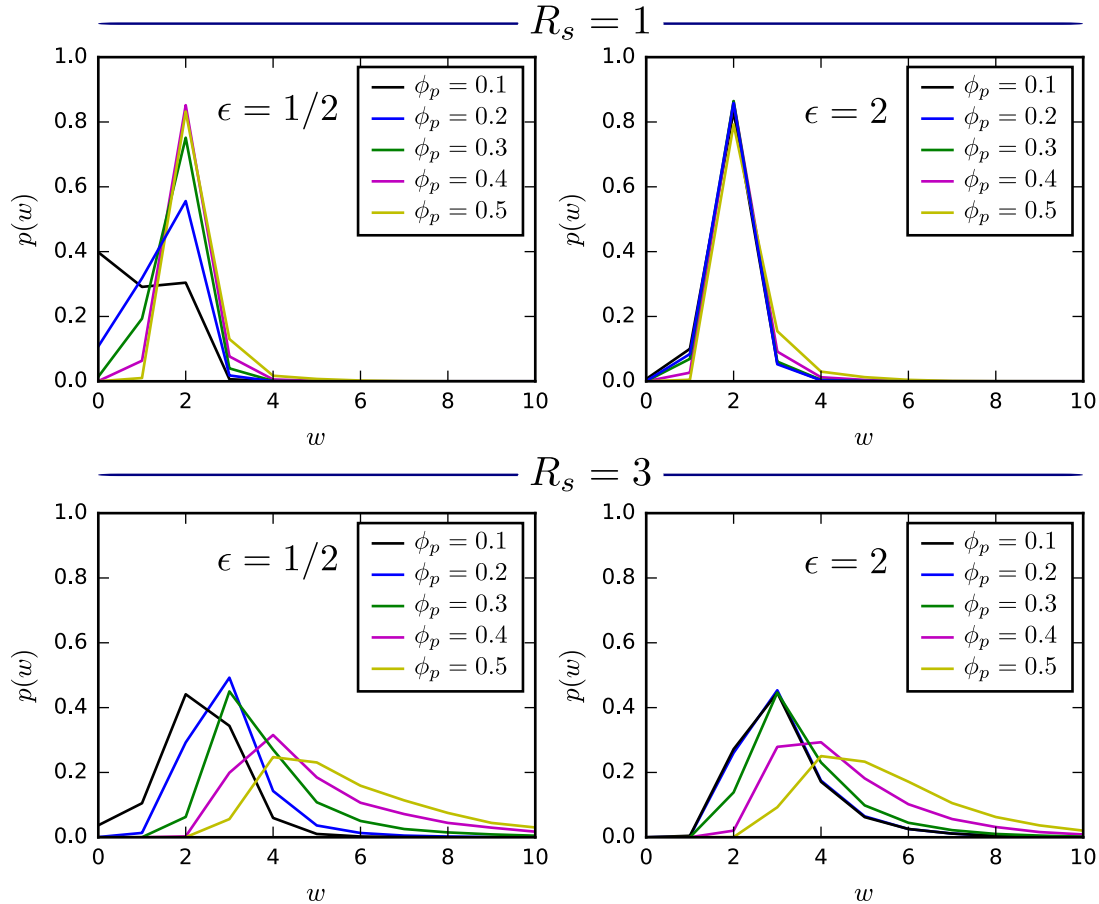
The radial distribution function  $R(d)$  for condensers (Fig. 3.23) indicates that the correlation of the particle positions is highly dependant on the size and binding strength of condensing proteins at all densities. Small condensers ( $R_s = 1$ ) show that a high density leads to a better correlation of first neighbour condensers. In the case of strong binding, both low and high polymer volume fractions show a high correlation for the first neighbour (first peak). But, in between concentrations have a reduction in the probability of finding a condenser at a direct contact with another one (at  $d = 2R_s \approx 2$ ), where a minimum is seen for  $\phi_p = 0.3$  and  $\phi_p = 0.4$ . In this case the second peak, attributed to one polymer bead in between two condensers (at  $d = 2R_s + a_0 \approx 3$ ), becomes dominant. Larger condensers exhibit a reduction in the probability of direct contact of two condensers irrespectively of



**Figure 3.23: Radial distribution function of condensers for different polymer volume fractions.** Data is shown for stiffness  $K_b = 25$  (DNA) with condenser volume fraction  $\phi_s = 0.05$  confined in a sphere of radius  $R_c = 30$  for different combinations of condenser radius  $R_s = 1, 3$  and condenser binding strength  $\epsilon = 0.5, 2$ .

the binding strength. It seems that the presence of a polymer forces two condensers to be more distant, as it is energetically more favourable for them to be bound. Interestingly, with larger condensers ( $R_s = 3$ ) this effect leads to a almost complete elimination of the direct contact, which is present even when more than one bond, as in the case with  $\epsilon = 1/2$ , is needed for binding to be thermally stable. At high volume fractions  $\phi_p = 0.5$  all the peak positions are moved to lower values, as the particles are at a sufficient density to warrant closer contacts.

The wrapping number  $w$ , which corresponds to the longest length of a continuous polymer segment whose beads are bound to the same condenser, is shown in Fig. 3.24 for different polymer volume fractions. Small and weak binding condensers at low concentrations are unbound with significant probability. At high densities, we see saturation at  $w = 2$ , the same value seen for all small and strongly bound condensers. The limiting factor is in this case the size of the condenser. Larger condensers seem to efficiently wrap the polymer, and as such can bind the polymer even with weak binding. Additionally,



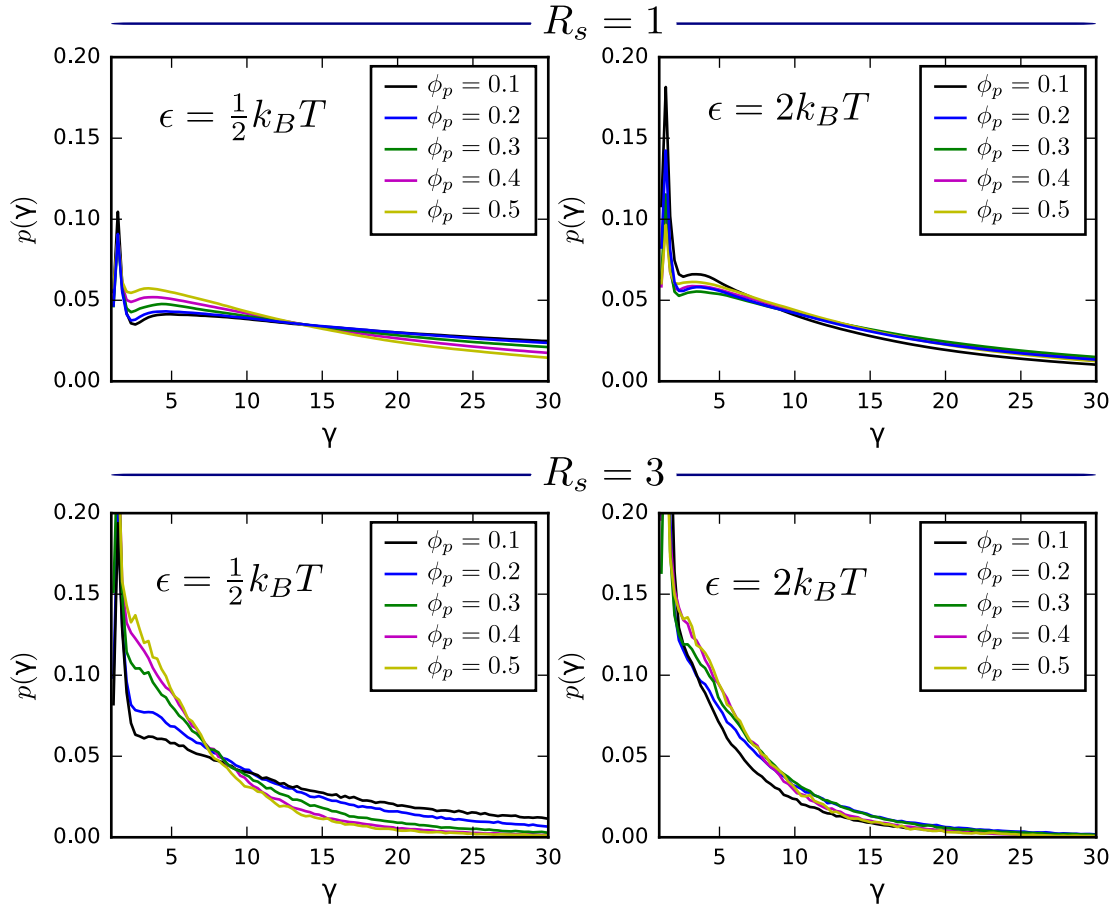
**Figure 3.24: Wrapping for different polymer volume fractions.** Data is shown for stiffness  $K_b = 25$  (DNA) with condenser volume fraction  $\phi_s = 0.05$  confined in a sphere of radius  $R_c = 30$  for different combinations of condenser radius  $R_s = 1, 3$  and condenser binding strength  $\epsilon = 0.5, 2$ .

the distribution  $p(w)$  is much wider, and increases with an increase of polymer volume fraction. The strength of the binding is also not a major factor, as it seems that it is the confinement that forces the polymer to more tightly interact with condensers.

For small condensers with  $R_s = 1$  we see that the most probable winding number is  $w = 2$  for both binding regimes, meaning that most of the condensers will bind maximally two connected DNA beads. An increase of condenser radius enables a larger  $w$  on average, as the DNA can have longer segments bound to the condenser as both the surface is larger and the required polymer bending smaller. The obtained growth is proportional to  $R_s$  and surprisingly shows no major dependence on the interaction strength, again indicating that the dominant effect comes from confinement, and not interaction strength in this regime. From the wrapping indicator we conclude that when the size of the condensing particles is small,  $R_s \sim 1$ , they predominantly act as bridger molecules, while larger condensers ( $R_s > 3$ ) approach the behaviour expected for wrapping particles [122].

We have found that the condensers bind small segments of polymer on them, but

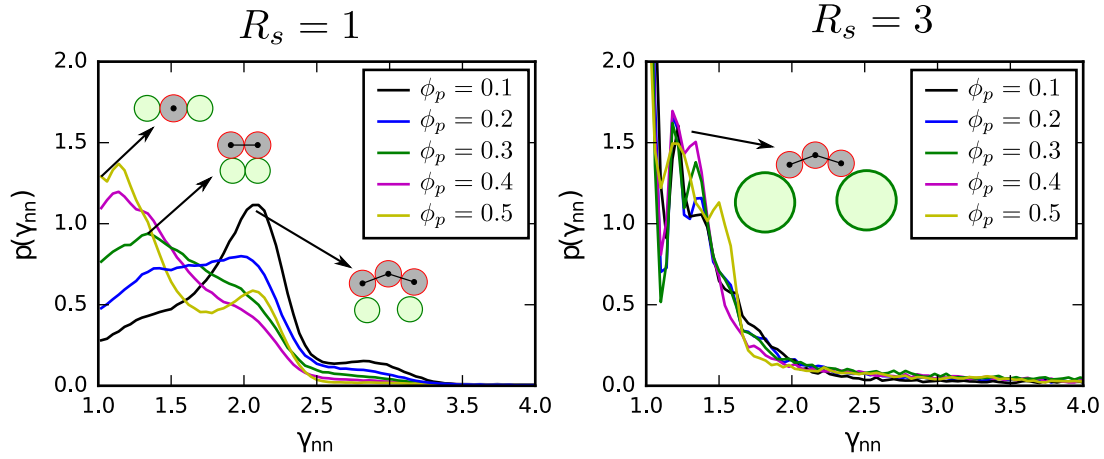




**Figure 3.25: Connectivity indicator for different polymer volume fractions.** Data is shown for stiffness  $K_b = 25$  (DNA) with condenser volume fraction  $\phi_s = 0.05$  confined in a sphere of radius  $R_c = 30$  for different combinations of condenser radius  $R_s = 1, 3$  and condenser binding strength  $\epsilon = 0.5, 2$ .

that does not provide information on how or if different beads are interconnected by the polymer backbone. We can study how condenser pairs are connected via the polymer backbone to see if the connecting DNA follows the shortest path between them. For each pair of condensers we construct the connectivity indicator  $\gamma = \min_i \{l\} / d$  which is the ratio between the smallest distance along a contour of DNA connecting these two proteins, and the closest (point to point) distance  $d$  between them. Figure 3.25 shows how the probability distribution for  $p(\gamma)$  changes for different polymer volume fractions. We see that  $R_s = 1$  condensers have a flat probability distribution for  $\gamma$  indicating that any two condensers are mostly connected indirectly. There is a small maximum at roughly  $\gamma \sim 1.5$  that seems to be related to spatially nearest neighbours directly interconnected with the contour. A higher density slightly enhances more direct contacts. If we increase the condenser radius, we see that smaller  $\gamma$  become increasingly more probable, meaning they are more directly connected. As the cross section of a condenser goes like  $\sim \pi R_s^2$  a "random walk" (in a crowded and confined environment) has a much higher probability of stumbling across a nearby condenser. A stronger interaction facilitates more direct connections, for

then the probability for small  $\gamma$  is larger, while maintaining the same position of the peak in  $p(\gamma)$  is maintained. The presence of condensers with stronger binding seems to force the DNA to have a more directed path between any two condensers.



**Figure 3.26: Connectivity indicator between nearest neighbours for different polymer volume fractions.** Data is shown for stiffness  $K_b = 25$  (DNA) and condenser binding strength  $\epsilon = 2$  with condenser volume fraction  $\phi_s = 0.05$  confined in a sphere of radius  $R_c = 30$  for different condenser radiuses  $R_s = 1, 3$ .

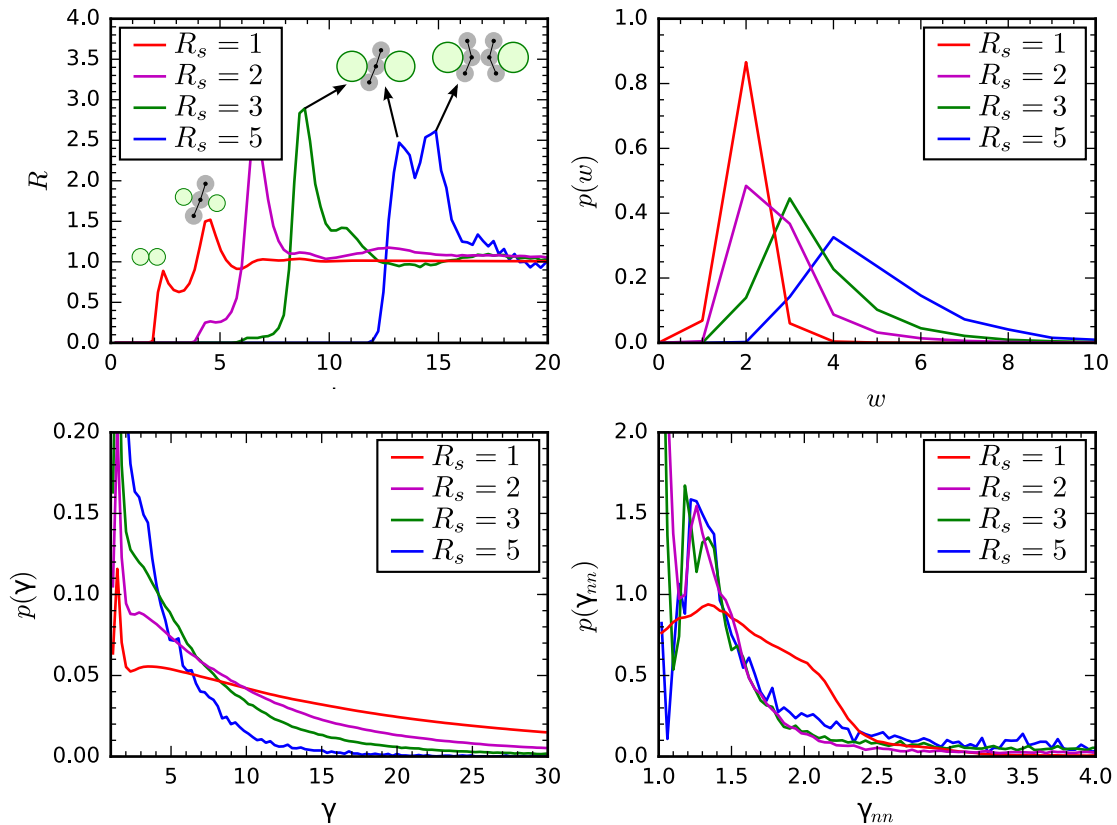
We can also use  $\gamma$ , the ratio between the DNA length connecting two condensers and their mutual distance, to study how relaxed the interconnectivity is in the neighbourhood of condensers. We can define  $\gamma_{NN} = \min_i \{l^{NN}\} / d_{NN}$  as the ratio of the shortest contour length connecting two spatially nearest neighbours and their mutual distance. Figure 3.26 shows the variation of  $\gamma_{NN}$  with  $\phi_p$  for different condenser sizes in the strong binding regime. For  $R_s = 1$  there are three major contributions to  $p(\gamma)$ : a) at  $\gamma_{NN} \sim 1$  from direct contacts through one polymer bead located between two nearest neighbours, b) at  $\gamma_{NN} \sim 1.25$  coming from the polymer touching (grazing) two neighbouring beads touching at a mutual distance of  $d = 2R_s$ , and c) at  $\gamma_{NN} \sim 2$  most probably coming from the polymer grazing two neighbouring beads at a mutual distance of  $d = 2R_s + a_0$  (second peak in Fig. 3.23). At low volume fractions  $\phi_p$  and small condensers ( $R_s = 1$ ) the system is relaxed and nearest neighbours are contacted indirectly. As  $\phi_p$  is increased, and although the volume fraction of condensers is present, we get more directly connected condensers as the crowding has increased. Larger condensers possess a slightly more pronounced probability of finding direct contacts between them, but have shorter indirect contacts centred at  $\gamma_{NN} \sim 1.25$ . This is because larger condensers have a higher probability of direct contacts due to their larger size thus reducing the total number of possible topological paths in respect to smaller condensers.

We conclude that higher densities of polymer enhance binding until we reach the concentrated regime when the crowding inhibits efficient binding. Small condensers are more able to permeate the structure and distribute themselves than larger condensers.

Larger condensers are unable to be directly in contact, as they are limited by the polymer they have bound, or wrapped, on themselves. Crowding large condensers with more polymer enhances this wrapping.

### Formation of polymer-condenser quasi-particles

Larger condensing particles indicated (Fig. 3.23) a tendency for the mutual distance between closest condensers to increase. We identify two major contributions in  $R(d)$ , one from direct contacts between two condensers at  $d = 2R_s$ , and another from a polymer located in between two closest condensers at  $d = 2R_s + a_0$ . In the strong binding regime we see a transition from a mixture of direct and DNA-mediated contacts to only DNA-mediated condenser contacts for large condensers. The direct contact between condensers is slightly less probable in respect to a condenser-bead-condenser contact for  $R_s = 1$ . With larger condensing particles, the direct contact between condensers becomes suppressed, and all interactions are mediated by a polymer layer around the condensers. This becomes more relevant for larger ( $R_s = 3$ ) condensers with strong binding, so we test how do the

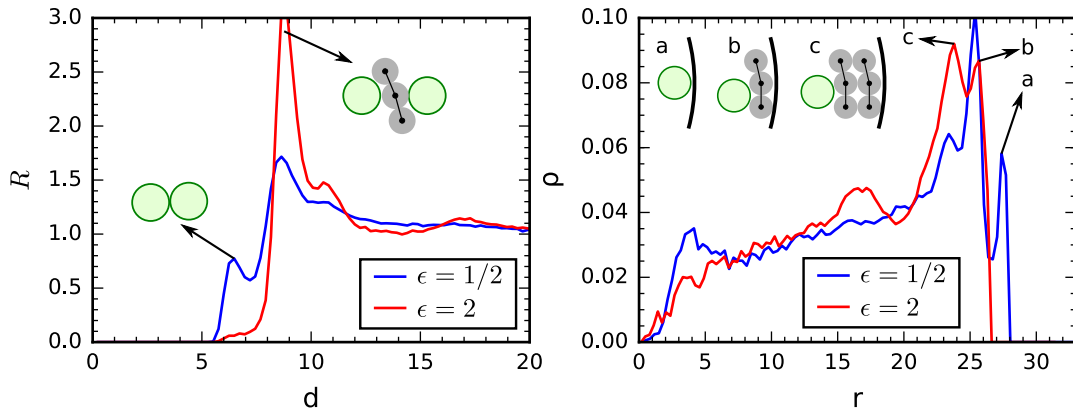


**Figure 3.27: Structural change for different radii  $R_s$  of condensing proteins.**

Graphs show normalized radial distribution functions  $R$ , wrapping  $w$  and the probability distribution for connectivity for all particle pairs  $p(\gamma)$  and for just spatially nearest neighbours  $p(\gamma_{NN})$ . Parameters were  $R_c = 30.0$ ,  $\phi_p = 0.3$ ,  $\phi_s = 0.05$  and  $\epsilon = 2$ .

indicators change with condensing proteins of different radii  $R_s$  (Fig. 3.27). An increase of condenser radius to  $R_s = 3$  shows that direct contacts between condensers are highly improbable – there is always at least one polymer bead in between two condensers. When the radius is further increased to  $R_s = 5$ , we see two close peaks at  $d = 2R_s + a_0$  and  $d = 2R_s + 2a_0$  with equal statistical weight. The increase of radius has allowed the condensers to cover themselves each with their own polymer "layer". The system transitions from a state where condensers can interact directly, to a state where each condenser is covered by at least a layer of polymer which mediates all interactions, a form of quasi-particle like the one one used in the previous section in our effective model of adenosomes [97]. This formation of quasi-particles is followed by an increase in the wrapping indicator  $w$ .

An increase of condenser size also enhances wrapping and makes the core better interconnected, as  $p(\gamma)$  show a higher probability of direct paths between any two condensers. From the wrapping indicator we conclude that when the size of the condensing particles is small ( $R_s \sim 1$ ), they predominantly act as bridging molecules, while larger condensers ( $R_s > 3$ ) approach the behaviour expected for wrapping particles [122]. The connectivity of nearest neighbours  $\gamma_{NN}$  indicates a transition between a more sparse relationship at  $R_s = 1$  to a more direct connectivity for larger  $R_s$ . These changes in connectivity can be attributed to an increase in the cross section of the condensers.

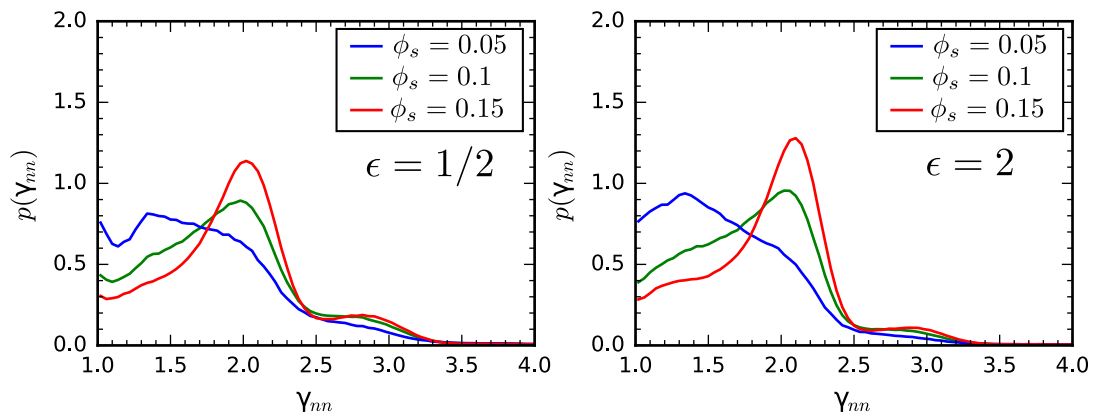


**Figure 3.28: Layering comparison for two binding strengths.** Graphs show normalized radial distribution functions  $R$  and density probability for condenser particles  $\rho(r)$ . Parameters were  $R_c = 30.0$ ,  $\phi_p = 0.3$ ,  $\phi_s = 0.05$ ,  $R_s = 3$  and  $K_b = 25$ .

The formation of adenosomes covered with a "halo" of polymer beads can also be seen in the density (Fig. 3.28). After the system transitions from a mixture of direct and indirect (polymer mediated) contacts to only indirect, we see a change in the density. First, the outermost layer of condensers in direct contact with the confinement disappears. Secondly, two peaks evident in the density  $\rho$  at a distance  $R_s + a_0$  and  $R_s + 2a_0$  from the edge of the confinement become comparable for  $\epsilon = 2$  – the first one from a quasi-particle

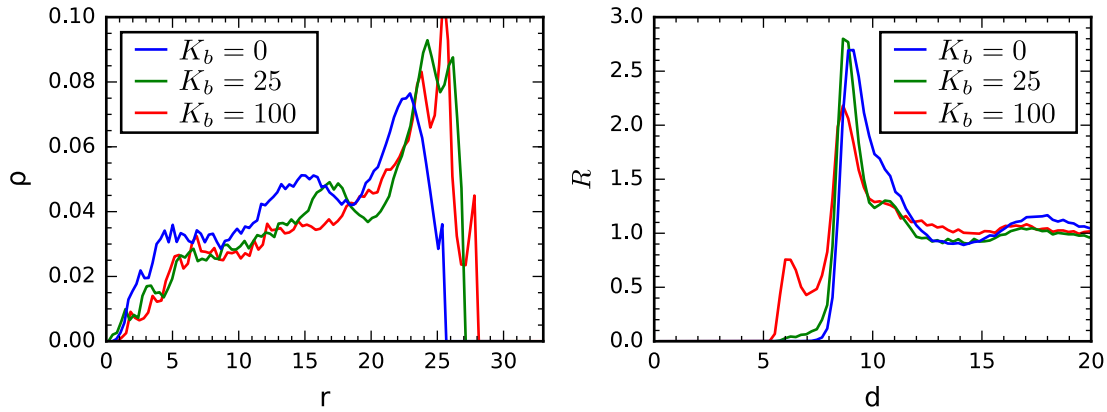
directly contacting the surface, the second from an additional layer of polymer between the quasi-particle and the confinement wall. Additionally, a peak at roughly  $2R_s$  from the confinement surface begins to form. Thus, we see quasi-particles in both the radial distribution function  $R$  and in the density  $\rho$ . Surprisingly, this is followed by negligible differences in the interconnectivity of condensers, only direct links  $p(\gamma_{NN} = 1)$  becoming slightly more probable, and also a negligible difference in  $p(w)$  between the weak and strong binding case. So although wrapping is connected to formation of the quasi-particle, it is not sufficient.

### Condenser volume fraction

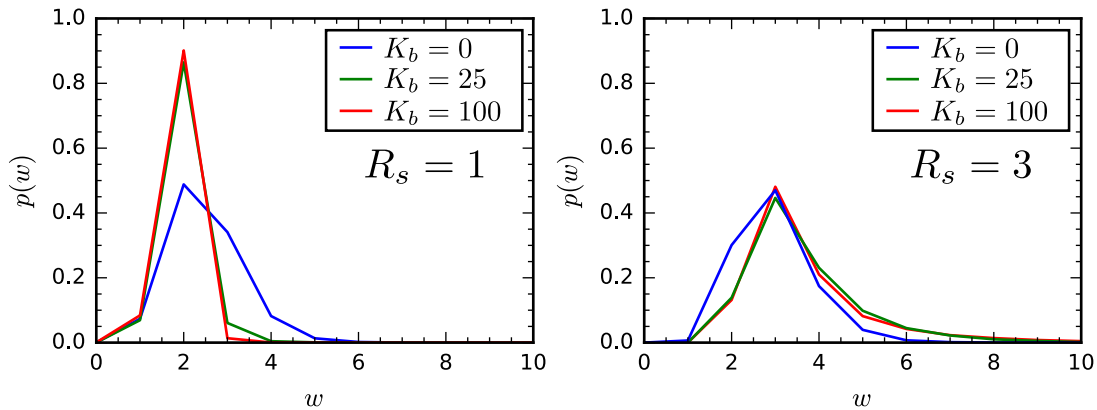


**Figure 3.29: Nearest neighbour connectivity for different condenser volume fractions  $\phi_s$  at  $R_s = 1$ .** Data is shown for stiffness  $K_b = 25$  (DNA) with polymer volume fraction  $\phi_p = 0.3$  confined in a sphere of radius  $R_c = 30$  for condenser binding strengths  $\epsilon = 0.5, 2$ .

Beforehand we have fixed the condenser volume fraction  $\phi_s = 0.05$  at the value that is both expected inside adenovirus cores, and the minimum in pressure obtained for  $R_s = 1$  and  $\epsilon = 2$  condensers. An increased amount of condensers enhances the direct contact between condensers and suppresses any layering. The connectivity  $\gamma$  remains the same on the scale of the whole confined system, but the nearest neighbour connectivity  $\gamma_{NN}$  is changed for small condensers ( $R_s = 1$ ). Figure 3.29 shows how an increase of condenser volume fraction changes the local connectivity of condensers. Surprisingly, an increase of condenser volume fraction makes the nearest neighbour links less direct, although the probability to find another condenser at the minimal distance  $R(d \approx 2R_s)$  is increased. It seems that the increased crowding complicates the paths the polymer must take between neighbouring condensers, although the density of condensers is larger.



**Figure 3.30: Density and radial distribution function for different bending stiffness.**  $\rho$  and  $R$  are shown for different bending stiffness parameters  $K_b$ . Parameters were  $R_c = 30$ ,  $\epsilon = 2$ ,  $R_s = 3$ ,  $\phi_s = 0.05$  and  $\phi_p = 0.3$ .



**Figure 3.31: Wrapping for different bending stiffness.**  $p(w)$  is shown for two different condenser radii  $R_s$  for different bending stiffness parameters  $K_b$ . Parameters were  $R_c = 30$ ,  $\epsilon = 2$ ,  $\phi_s = 0.05$  and  $\phi_p = 0.3$ .

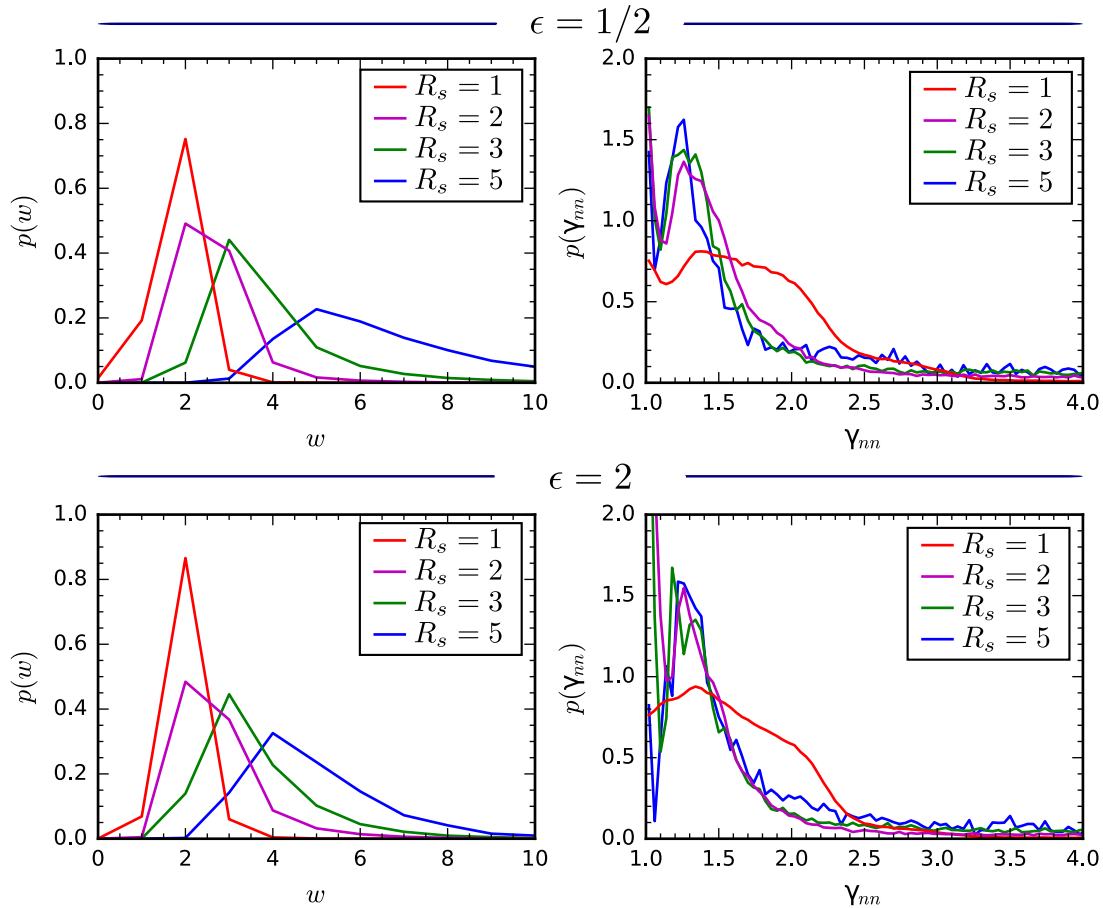
### Comparing various bending regimes

Figure 3.30 shows how the bending stiffness  $K_b$  for flexible ( $K_b = 0$ ), stiff ( $K_b = 100$ ) and DNA-like ( $K_b = 25$ ) polymers influence the internal organization in confinement. Stiff polymers are found to reduce the layering of polymers on condensers, while flexible polymers enhance the quasi-particle picture from before. As the stiffness is reduced, the density  $\rho$  exhibits a transition from two peaks for condensers and polymers near the surface of the confinement, into one merged peak for the quasi-particle. Similarly in  $R$  we see a widening of the first peak for the flexible polymer, corroborating the quasi-particle picture. Interestingly, wrapping (Fig. 3.31) indicates that flexible polymers have a larger winding number  $w$  for small condensers, and smaller  $w$  for larger condensers, with respect to stiff polymers. Stiffer polymers, once they make contact with a condenser, have a

higher probability of continuing on the same path, grazing the surface of the condenser. Such paths allow the stiff polymers to be slightly diverted to follow the shape of large condensers more closely, in order to achieve more binding opportunities with the same condenser. Flexible polymers have almost no correlation between subsequent steps in their "random walk", and are expected to more easily, temporarily, leave the neighbourhood of a condenser.

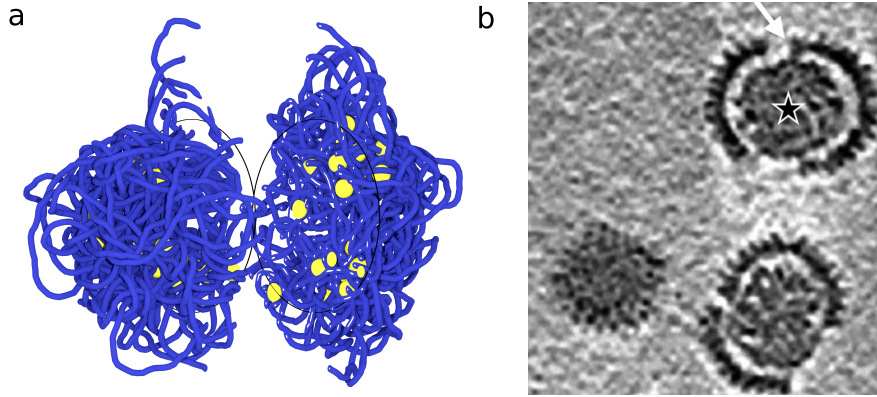
### 3.2.5 Opening the capsid

To understand what is the degree that confinement influences the structure, we study how removing the confinement influences the mixture. We first show an example of confined wrapping  $w$  and nearest neighbour connectivity  $\gamma_{NN}$  for parameters close to the ones we believe match the conditions in adenoviruses best (Fig. 3.32).



**Figure 3.32: Connectivity of DNA and condensing proteins with confinement.** Probability distributions for the wrapping number  $w$  and  $\gamma_{NN} = \min_i \{l_i^{NN}\} / d_{NN}$  averaged for all condensing particle pairs which are nearest neighbours are shown in the weak ( $\epsilon = 1/2$ ) and strong ( $\epsilon = 2$ ) binding regime for different values of the condenser radius  $R_s$ . Parameters were  $R_c = 30$ ,  $\phi_p = 0.3$ ,  $\phi_s = 0.05$  and stiffness  $K = 25$ .

After the capsid is removed and a sufficient time has passed for the system to equilibrate,



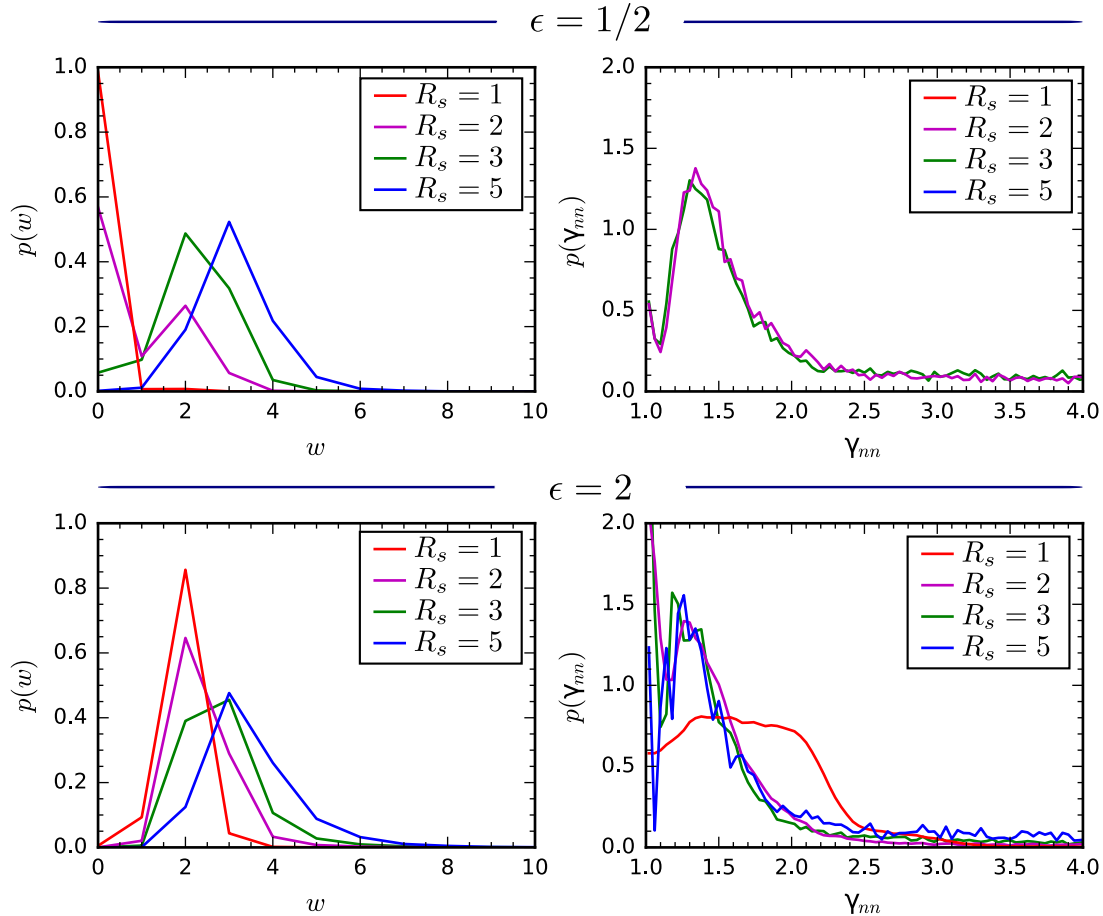
**Figure 3.33:** a) DNA and condensing particle cluster after removing confinement. Same as Fig. 3.6 except the capsid confinement has been removed. b) Cryo-EM of opened adenoviruses showing a compact core and remnants of the capsid. Adapted from Ref. [121].

we see that the structure has relaxed but retains the general shape of a cluster (Fig. 3.33a). In the case of small weakly binding condensers ( $\epsilon = 1/2k_B T$ ) almost all of the condensing particles leave the cluster.

To gain insight on the internal connectivity we study wrapping ( $w$ ) and connectivity ( $\gamma$  and  $\gamma_{NN}$ ) for DNA condenser mixtures without spherical confinement as shown in Fig. 3.34. Wrapping  $w$  shows the most striking difference from previous results in confinement. In all cases, the tendency for wrapping has reduced. In the case of strong binding, the dependence of  $w$  on size has become weaker, indicating that the entropic penalty for the DNA to remain near the bead remains stronger than the energy gained from binding. In the weak binding regime, we see that most of the smaller condensers do not have any DNA bound to them (seen as  $p(w) = 0$ ), as they have escaped the cluster. Larger condensers for  $R_s = 5$ , and  $R_s = 3$ , show the same wrapping profile for both weak and strong binding regime. When the size of the condenser becomes more comparable to the persistence length of DNA, the binding is able to "bend" the DNA to touch the surface.

The contour lengths of DNA connecting condensers, as represented in  $\gamma_{NN}$  (Fig. 3.34), indicate a more sparsely connected, if connected at all, structure for weak binding without confinement. Smaller condensers do not participate in forming any network and leave the structure, while larger condensers start to exhibit the general profiles seen with confinement for which the probability for direct links (small  $\gamma_{NN}$ ) becomes larger than for indirect links. Strong binding indicates a similar structure as in confinement. The DNA links for nearest neighbours shows the most striking difference for small and weak binding condensers where there are no direct links between condensers, only indirect ones where DNA subsequently touches two condensers during its random walk ( $\gamma_{NN} \sim 1.4$ ). In contrast, strongly bound condensers exhibit the same general trend as in the confined case, except the structure is more relaxed, with indirect links becoming more probable. This confirms the visual conclusion that strongly binding condensers retain the same general connectivity as in





**Figure 3.34: Connectivity of DNA and condensing proteins with confinement removed.** Probability distributions for the wrapping number  $w$  and  $\gamma_{NN} = \min_i \{l_i^{NN}\} / d_{NN}$  averaged for all condensing particle pairs which are nearest neighbours are shown in the weak ( $\epsilon = 1/2$ ) and strong ( $\epsilon = 2$ ) binding regime for different values of the condenser radius  $R_s$ . All other parameters are the same as in Fig. 3.32 but with the spherical confinement removed.

confinement, while weak condensers cause a loss of connectivity except for larger condensers where several bonds can be achieved simultaneously to "anchor" the condenser inside the structure. If we take  $\gamma$  for all connections in the system, we find minor differences from the confined case implying that the global connectivity remains the same.

### 3.2.6 Implications for viral packing

By using molecular dynamics simulations of mixtures of explicit DNA and condensing proteins we investigated the organization of material inside viral capsids. The approach we presented enables us to expand on the effective model of core quasi-particles (adenosomes) and show that we can explain some basic experimental features: the depletion layer near the capsid surface and that core particles interact as effective quasi-particles.

The parameters used for the condenser and polymer model correspond to a simplified model applicable to adenoviruses if we take a DNA volume fraction of  $\phi_p = 0.3$  and

condensing agent volume fraction of  $\phi_s = 0.05$  [97]. Density profiles for  $R_s \sim 1$  in the strong binding regime (Fig. 3.22) follow the experimental profiles showing a flat density profile inside and a gradual decay of the probability for finding condensers close to the capsid surface – a depletion layer of particle density. The total density as it would be seen in electron microscopy, would be a weighted sum of DNA and protein densities. The DNA in this regime shows little layering (Fig. 3.21), as does the condenser density distribution ( $R_s = 1$  in Fig. 3.22), and has the same general profile seen in core protein density for condensing particles in adenoviruses ( $R_s = 1$  in Fig. 3.2) [97] and chromatin in SV40 [26]. The existence of surface layering is a well known feature of models with only spherical particles [128] and only DNA polymers [67] and originates in the confinement wall inducing "ordering". This particle layering near the capsid surface is not necessarily seen only in density, but also happens in the orientation of elongated condensing proteins near the capsid walls as seen in coarse grained models of polyomavirus SV40 [26]. Thus, care is needed when interpreting the origin of density fluctuations from electron microscopy images, as their origin can come from any combination of the preceding effects.

Comparing the radial distribution functions obtained from simulations (Fig. 3.23) to those obtained for condensing proteins in adenoviruses [97], we see we can reproduce two basic behaviours. First, there exists a regime where condensing proteins do not interact directly but through a DNA medium. If there was tethering between two condensing particles, it would present as a high probability at finding the two particles at some mutual distance. No such evidence for tethering is seen. Combined with the density  $\rho(r)$  depletion layer seen also in adenoviruses, it seems that our model can successfully cover the basic internal structure in adenoviruses. This is in spite of the simplicity of our model, and seems to be a feature of the crowded and confined environment. We can also compare to the core organization of histone proteins inside SV40 viral capsids [26]. Histones can be modelled as oblate particles which wind DNA around them [113], so our model would constitute an oversimplification. Yet, Cryo-EM of SV40 cores also indicates lack of ordered core organization [109] which is also collaborated by SAXS studies corroborated with a coarse grained model of histones in an effective medium of DNA confined in a viral capsid with attractive interactions [26]. The coarse grained model by Saper et al. [26] treats the SV40 capsid as being filled with oblate ellipsoids with steric repulsion, non-specific binding and tethered by a harmonic spring to simulate the backbone of DNA connecting the histones. They find that the core structure exists in a disordered state with densities that correspond to the experimental evidence. Although the basic geometry is different because of the shape of condensing particles, we can compare general trends for the density as chromatin is estimated to have a binding energy of  $\sim 6 k_B T$  [113]. Our density profiles in the strong binding regime for  $R_s \sim 1$  (Fig. 3.22) best approximate the density seen, corresponding also to a model of oblate ellipsoidal particles with nonspecific long range attraction, tethered with a harmonic spring approximating DNA. Alternate models used

by Saper et al. [26] for a tighter/closer tethering have the same basic structure as we see in the case of larger condensing particles. Although similarities are seen, our results indicate that the addition of a polymer can provide different explanations for the same behaviour as does tethering combined with a different particle shape. In the case of SV40 it is well known that chromatin does indeed wrap DNA around it, giving plausibility for the tethering model. To consistently prove that tethering is not seen *in vivo* due to crowding a more complex model for the condensing proteins is needed as is the inclusion of possibly attractive interactions with the capsid [24, 26].

Removal of confinement (Fig. 3.33a) for parameters close to those expected in adenoviruses has shown a relaxed but still compact structure. Although it was thought that after opening of adenovirus capsids, the core resembles a fibre structure [118, 119], recent experiments have shown that this is an artefact from sample preparation [121]. Cryo-EM shows that most adenovirus cores after the removal of the capsids remain compact (Fig. 3.33b) [121], validating parallels between the model behaviour and adenoviruses. Similarly, in SV40 it is known that the state of its minichromosome depends on the experimental conditions [147] as the view that chromosomes *in vivo* exist as ordered structures is being challenged [148, 149]. The finding that cores remain compact is not surprising, but it is interesting to study the degree of structural change.

The scarcity of experimental evidence on the packing of nucleic acids with condensing proteins has motivated us to make direct comparisons with the only two systems with direct experimental evidence of internal structures: adenovirus [97] and polyomavirus [26, 104]. Papovaviruses from the family of polyomaviruses also have genomes condensed with histones forming a minichromosome [104]. There is no evidence of order in the viral core, even in the areas close to the inner capsid surface [150]. On the other side, BK virus from the polyomavirus family shows a small degree of core ordering (shelling) near the capsid walls [151], which could be caused by higher density, or a change of interactions as supported by both our model (caused by crowding) and Saper et al. (caused by different interaction screening) [26]. There exist other possible candidates with core proteins that could be capable of acting as condensers: baculovirus [100], mimivirus [102] and poxviruses [103]. The same lack of capsid internal ordering is seen in the giant mimiviruses [102] and the vaccinia poxvirus [152]. Vaccinia exhibits an increase in total density near the capsid surface [152] – reminiscent of our increase in the probability for finding larger condensers at the confinement surface in respect to the center. Clearly, there is ample opportunity for testing of our results.

Obtained density profiles of protein content inside viral capsids, and radial distribution functions for these proteins, provide a picture of a disordered structure. Density profiles are able to explain the lack of symmetry and order seen in viral capsids of adenoviruses [5, 97] and polyomaviruses [26, 151]. The correlations between condensing protein positions

shows that DNA can be considered an effective medium, mediating condenser-condenser interactions. We find no direct DNA tethering between condensers, the connections between them being mostly random for smaller condensing particles, and directed for larger condensers. The internal connectivity between condensers mediated by DNA is found to be indirect and with no evidence of direct winding of large lengths of DNA. For the polymer, the interior of the capsid is a crowded environment where the lack of possible configurations promotes winding of DNA around condensers. Confinement and crowding seem to influence the way condensers with non-specific binding interact with the polymer – we see that different sizes of condensers promote linker molecules to approach the behaviour expected from wrapping particles [122]. The resulting crowding facilitates more direct contacts between the DNA polymer and the condensing particles. We conclude that confined mixtures of DNA and condensing proteins are importantly influenced by confinement and crowding, and not only the exact interactions, which may, in certain regimes, be secondary.



## CHAPTER 4

# Conclusions

By studying different phases of pure DNA in its compacted form, we were able to characterize the DNA condensed inside a cell and DNA confined inside viral capsids. The state of confined DNA was studied in different regimes which depend on the DNA stiffness, confinement size and electrostatic screening. We showed how emptying of such a capsid could proceed thus paving the ground for further understanding of the interplay of forces during ejection of viral dsDNA from bacteriophages into cells. We examined *in vivo* ejection of non-condensed DNA from tailed bacteriophages into bacteria. The ejection is dominantly governed by the physical conditions in the bacteria - the confinement of the DNA in the virus capsid only slightly helps the ejection. In spite of the assumptions that passive ejection is not possible, we show that the mechanism of viral ejection from dsDNA bacteriophages into Gram-positive bacteria could be explained as a competition between the resisting turgor pressure of a cell and a free energy gain from condensation of the ejected part of the DNA. The presence of condensed DNA in the cell enables us to estimate the maximal bacterial turgor pressure against which the ejection can still be fully realized. The thus calculated pressure ( $\sim 5$  atm) shows that the ejection of DNA into Gram-negative bacteria could proceed spontaneously, i.e. without the need to invoke active mechanisms.

The packing mechanism for DNA in adenoviruses has long evaded a precise description since the viral core, including DNA and proteins, lacks icosahedral order characteristic of the virus protein coating (capsid). We analysed experimental cryo-EM images showing an apparently random distribution of unknown core particles inside the adenovirus core. These core particles, called "adenosomes", are identified as complexes of core proteins which condense parts of the genome. Our analysis of their positional relationships shows that the core lacks symmetry and strict order, yet the distribution of the condensing proteins is not entirely random. The features of the distribution can be explained by modelling the condensing proteins and the part of the DNA each of them binds as very soft spheres, interacting repulsively with each other and with the capsid. The DNA appears to act only as an effective medium for the soft interadenosome interactions. A backbone of DNA

linking the condensing proteins is not needed to explain the experimental results. We conclude that although these condensing proteins are connected by DNA in disrupted virion cores, the *in vivo* capsid is a crowded environment changing the effective interactions involved in the packing of the DNA material.

Modelling of virion core structures was approached using Langevin dynamics simulations which include both condensing proteins and an explicit DNA polymer inside spherical confinement. The DNA and polymer interact attractively via non-specific interactions. We find there exists an optimal concentration of condensing particles which helps packaging of the genome by reducing the interior pressure. Internal organization of condensing particles shows that they tend to cover themselves with the DNA polymer which provides an effective medium for interactions with other condensers, confirming the applicability of our effective model for core particle organization in adenoviruses. Crowding of the viral interior and confinement influences the conformation of the DNA and proteins, changing how DNA interconnects the condensing proteins. It facilitates more direct contacts between the DNA polymer and the condensing particles. We find no direct DNA tethering between condensers, the connections between them being mostly random for smaller condensing particles, and directed for larger condensers. The internal connectivity between condensers mediated by DNA is found to be indirect and with no evidence of direct winding of large lengths of DNA. Our model is able to explain the general internal organisation of adenovirus cores, and provide insight into packing of genetic material in similar systems.

## Outlook

Modelling mixtures of DNA and condensing proteins will become increasingly important in the domain of virology as experimental techniques are starting to unravel even unordered cores [26, 97, 153]. More complex systems can be studied within the same framework in order to test how different DNA condensing molecules (linkers, bridgers and wrappers) [122] influence the internal organisation and viral capsid pressures. Full DNA and protein models also enable comparing nano-indentation of viruses using atomic force microscopy [135] and determining how the core proteins interact with the DNA and affect the internal pressure [123] or influence capsid stability [154]. Studying the packing of DNA with proteins inside viral capsids, might provide information on how cellular crowding, through changing interactions [27], influences chromosomal packing [148, 149] or even protein folding [27].

# Prošireni sažetak

## 1 Uvod

Pakiranje nanočestica, sintetičkih polimera i genoma u ograničeni prostor je od izrazitog fundamentalnog i tehnološkog značaja. S jedne strane, to je proučavanje interakcija unutar makromolekularnog kompleksa s mnogim potpisima "života" pa je stoga važno za razumijevanje životnih procesa [9]. S druge strane, virusi su evoluirani nano-strojevi sa "svrhom" prolaska kroz staničnu membranu radi dostavljanja svog "tereta", genoma [6]. Bilo kakve izmjene tog tereta, kao što je korištenje dizajniranih makromolekula ili nanočestica, nose sa sobom značajan potencijal za primjene u osobnoj medicini (ciljano dostavljanje lijekova ili genska terapija). Ipak, temeljno razumijevanje stanja DNK u virusnom omotaču pri različitim gustoćama pakiranja, a samim time i fizike koja stoji iza pakiranja virusnog genskog materijala i potonjeg procesa infekcije [9, 10], je nepotpuno. Iako je problem pakiranja savitljivog ili polu-savitljivog polimera u ograničene prostore dobro definirane geometrije (pukotine, šupljine) bio intenzivno proučavan [2], tro-dimenzionalne šupljine tek su nedavno postale predmetom istraživanja [11]. U ovisnosti o savitljivosti polimera, stupnju ograničenja, gustoći i interakcijama (isključenog volumena) očekuje se bogati fazni dijagram [12]. Za razumijevanje faznih stanja DNK i polimorfizma njezinih konformacija u ograničenom prostoru [10] potreban je pristup koji kombinira i međusobno povezuje različite teorijske modele. Što se tiče izotropnog (neuređenog) stanja (pri malim gustoćama) dobro je poznat fazni prijelaz u tekući kristal. Do faznog prijelaza dolazi zbog svojstvene anizotropije u isključenom volumenu dugih polimera [13].

Unatoč tome, rezultati eksperimenata ukazuju na neočekivane pojave u uzorcima makroskopskih dimenzija što potiče razvoj novih pristupa u modeliranju tekućih kristala u ograničenom prostoru [14]. Faza tekućeg kristala pokriva otprilike trećinu dijagrama gustoće [10, 12] te završava u gusto kondenziranoj fazi [16, 42]. Poznato je da formiranje gustih faza DNK (kondenzirana DNK) ovisi o kombinaciji elektrostatskih i hidrofobnih međudjelovanja između različitih dijelova negativno nabijene DNK "kičme" i pozadine jedno- i više- valentnih protuiona [9]. U tom duhu, gusto pakirana DNK u virusima (npr. bakteriofagima) je relativno jednostavan slučaj prikladan za proučavanje faza ograničene DNK bez složenosti koju donosi stanični (*in vivo*) okoliš.



Jedno od glavnih otvorenih pitanja vezano uz pakiranje DNK u viruse odnosi se na mehanizme njezinog oslobađanja u stanicu. Unatoč 50 godina istraživanja, objašnjenje svih relevantnih (termodinamičkih) sila koje vode izbacivanje dvostruke uzvojnice DNK iz repatih bakteriofaga u bakterijsku stanicu, još uvijek nedostaje [18, 19]. Izbacivanje počinje kao oslobađanje DNK iz potpuno ispunjenog virusnog omotača (proteinske presvlake bakteriofaga). DNK je pakirana do ekstremnih gustoća te na omotač vrši pritisak od 25 – 100 atm [9, 20]. Modeli razvijeni i testirani *in vitro* (vidi [18] i reference unutar) predviđaju da sile izbacivanja prouzročene čak i takvim ekstremno gustim pakiranjem nisu dovoljne za potpuni prijenos DNK u unutrašnjost stanice. Iako stanice imaju manji osmotski tlak od potpuno pakiranih bakteriofaga [21], sila izbacivanja (a time i tlak u omotaču) brzo pada kako se omotač prazni [22]. Nedavno provedeni Hershey-Chase experiment na razini jedne molekule [23] nagoviješta da izbacivanje *in vivo* nije kontrolirano količinom DNK preostale u omotaču već količinom izbačenom u stanicu. To znači da nakon što se tlak nakupljen u bakteriofagu "potroši" na izbacivanje DNK, proces preuzimaju stanični mehanizmi. Predloženi su različiti mehanizmi za završetak izbacivanja, no čini se da ni jedan ne daje cjelokupnu sliku, dok eksperimenti nagovještavaju postojanje mnoštva različitih mehanizama.

Postoje razlozi za vjerovati da su neki virusni proteini evolucijski prilagođeni za omatanje nukleinskih kiselina [24], što bi moglo biti važno za pakiranje umjetnog tereta u proteinski omotač [25]. Dobar primjer su adenovirusi koji pakiraju (dvolančanu) DNK u virusni omotač s DNK-vežućim proteinima, koji potencijalno pomažu u neutralizaciji naboja i kondenzaciji [5]. Ni za jedan od ovih proteina ne postoje strukturni podaci visoke rezolucije, i točna konformacija takve mješavine polimera (DNK) i nanočestica (proteini) još nije poznata. Pošto je DNK djelomično vezana na protein ili neutralizirana zbog prisutnosti proteina, može se modelirati kao efektivni medij za DNK-vežuće proteine. Ovdje pak efekti virusnog omotača mogu prouzročiti određeni stupanj uređenja u raspodjeli tih proteina. Slični efekti su nedavno opaženi u organizaciji kromatina u virusu SV40 gdje vezanje između proteina i prostornog ograničenja mijenja stanje spakiranog DNK materijala [26].

Prenapučeni okoliš polimera/polielektrolita i nanočestica također je zanimljiv s biološke točke gledišta jer se može povezati s prenapučenom unutrašnjosti stanice. Pokazano je da prenapučenost vodi na mnoge pojave kao što su renormalizacija golih interakcija i anomalna difuzija [28]. Unatoč tome, točna priroda ovog mehanizma kao i posljedice koje bi mogla imati na naše razumijevanje zbivanja DNK nisu poznati.

## 2 Sažimanje DNK

DNK se smatra sažetom ako je njena prostorna protežitost smanjena zbog vanjskog utjecaja. U prirodnom stanju DNK, kao i svaki polimer, ima konformaciju nasumičnog šetača u prostoru. Tretiraju se dva najvažnija mehanizma sažimanja DNA: a) kondenzacija DNK u guste faze [16, 37, 42], b) trodimenzionalno prostorno zatočenje DNK u male volumene [2, 11].

### 2.1 Kondenzirana DNK

U razrijeđenim otopinama, DNK-DNK interakcije su snažno odbojne te je za kondenzaciju DNK potreban jak poticaj. Poznato je da kondenzacija DNK može imati različite uzroke [39] te da se pri ekstremnim gustoćama kondenzacije manifestira u obliku heksagonalne faze [40, 41]. Heksagonalne faze DNK mogu biti uzrokovane viševalentnim ionima poput kobalt heksamina (3+) [40], ili spojevima prisutnim *in vivo* poput polilizina [30] i spermidina (3+) [29]. Alternativno, kondenzacija može biti uzrokovana vanjskim tlakom nekog polimera, npr. polietilen glikola [40, 40]. Kondenzacija DNK demonstrirana je i u kontroliranim biološkim uvjetima (vidi [42] i [41] te reference unutar). Kondenzirana DNA u unutrašnjosti stanice se smatra posljedicom prenapućenog okruženja s prisutnim kondenzirajućim proteinima [29, 30, 31, 32].

Koristimo kontinuumski model Ubbinka i Odijka [16, 42] kako bismo karakterizirali kondenzirane faze DNK, posebice one koje nalazimo u stanicama, te ga kasnije primijenjujemo za opis izbacivanja DNK iz bakteriofaga u stanicu. Ubbink i Odijk (UO) [16, 42] model opisuje slobodnu energiju DNK kondenzata u gustoj heksagonalnoj fazi preko trodimenzionalnog oblika kondenzata. Model tretira ukupnu energiju kondenzirane DNK u tri dijela: a) energetski doprinos proporcionalan volumenu koji teži kondenzaciji, b) energetski doprinos proporcionalan površini uslijed gubitka energetski povoljnih veza, te c) elastični doprinos od savijanja snopa DNK koji se odupiru formiranju kondenzata. Elastični doprinos će težiti širenju kondenzata u oblik torusa kako bi se snop DNK što manje savijao, dok površinski doprinos teži urušavanju kondenzata u globulu (sferu). Numeričkom optimizacijom oblika DNK kondenzata u UO modelu istražuju se mogući oblici kondenzata te konačne ovisnosti o količini kondenzirane DNK u kondenzatu, kao i njenoj stabilnosti pri malim duljinama kondenzirane DNK.

### 2.2 Faze zatočene DNK

DNK koja je prostorno zatočena unutar sfere pokazuje specifično ponašanje. U ovisnosti o međusobnom djelovanju duljine ustrajnosti  $L_p$ , radijusa  $R_c$  sfernog zatočenja i ukupne duljine DNK  $L$  očekujemo nekoliko režima [12]. Pri velikim gustoćama DNK ulazi u

gustu heksagonalnu fazu [40], a ispod tih gustoća njeno ponašanje ovisi o njenoj krutosti. Kada je duljina ustrajnosti manja od radijusa zatočenja nalazimo se u fleksibilnom režimu. Pri malim koncentracijama DNK se ponaša kao fleksibilni polimer u polu-razrijeđenom režimu koji pri većim koncentracijama prolazi kroz lokalno uređenje niti DNK u režim tekućeg kristala. U fleksibilnom režimu glavni doprinos dolazi od energije savijanja pri malim koncentracijama, dok pri većim koncentracijama nadvladava doprinos od interakcija isključenog volumena. Te interakcije isključenog volumena uzrokuju lokalno uređenje u tekući kristal [13, 51], prvo kroz nematičku, a zatim kolesteričku fazu [10]. Ukoliko je radijus zatočenja manji od duljine ustrajnosti, DNK se nalazi u režimu dominiranom savijanjem (kruti polimer). Ovaj režim je najmanje proučavan budući da je DNK u fiziološkim uvjetima upravo iznad granice prijelaza u taj režim [10, 65].

## 2.3 Savijanje zatočene DNK

Slobodna energija zatočene DNK opisuje se u dva režima pomoću modela srednjeg polja. Prvo, opisuje se energija savijanja DNK pri gustoćama kada DNK u potpunosti ispunjava prostor te kada energija savijanja nije dominantni član u slobodnoj energiji. Cilj je opis doprinosa savijanja za prijelaz u tekući kristal. Energija savijanja DNK opisuje se preko lokalne zakrivljenosti u prostoru [3] za koju je pretpostavljena cilindrična simetrija. Svaka lokalna "ćelija" DNK, ovisno o globalnom parametru uređenja, ili slijedi cilindričnu simetriju ili biva neuređena. Rezultat je doprinos savijanju slobodne energije koji ovisi o parametru uređenja koji se može dobiti iz teorije tekućih kristala.

U režimu savijanja, kada je savijanje DNK dominantni doprinos slobodnoj energiji, konstruira se efektivni model temeljen na definiciji nove kvazi-čestice. Uslijed izražene kompeticije duljine ustrajnosti i radijusa zakrivljenosti, DNK je pritisnuta uz unutrašnjost zatočenja te poprima oblik petlji [22, 55]. Definiranjem petlji kao kvazi-čestica za pakiranje, konstruira se slobodna energija u ovisnosti o načinu slaganja petlji. Navedeni izrazi se poopćuju tako da uključuju međusobno presijecanje petlji kako bi mogli opisati kontinuirani prijelaz iz uređenog u neuređeno stanje, poput onog uočenog u pakiranju elastičnih žica u sfere [56, 58].

## 2.4 Zatočenje u izotropnom režimu

Diskutira se režim fleksibilnog polimera pri prelasku iz razrijeđenih u polu-razrijeđene otopine, a netom prije prijelaza u tekući kristal. Prvo se računa isključeni volumen DNK temeljem kutno ovisnog interakcijskog potencijala DNK prikazane preko nabijenih štapića [13, 14]. Radi izračuna interakcije koriste se eksperimentalno određeni parametri interakcije [62] koji uzimaju u obzir elektrostatsko zasjenjenje kao i složenije pojave poput

kondenzacije protuiona [63]. Dobiveni isključeni volumen omogućuje računanje interakcije u izotropnom režimu primjenom Flory relacije [2]. Dobivene interakcije isključenog volumena se, uz prethodno dobivene izraze za energiju savijanja polimera, primjenjuju u opisu slobodne energije prostorno zatočene DNK. DNK lanac se renormalizira [61] u efektivni polimer s dimenzijom monomera jednakoj Kuhnovoj duljini, tj. dvostrukoj duljini ustrajnosti [2].

## 2.5 Faza tekućeg kristala

Opisuje se prijelaz iz neuređenog, izotropnog, režima u stanje tekućeg kristala nematičkog uređenja [60]. Iako je poznato da je niz faznih prijelaza zatočene DNK sličan onome makroskopskog sustava, točan utjecaj energije savijanja na točku prijelaza nije poznat [65]. U svrhu razjašnjenja tog utjecaja, u Onsagerov model za tekući kristal [13, 14, 51] se ugrađuje energija savijanja ovisna o globalnom uređenju proizašla iz prijašnjih razmatranja. Uz pretpostavku da se prijelaz odvija direktno iz neuređenog u nematičko stanje, bez kolesteričkog uređenja, minimizira se Onsagerova slobodna energija u prostornom zatočenju. Uspješno se predviđa točka prijelaza u tekući kristal na usporedivim gustoćama pakiranja kao i u eksperimentima [10, 65].

Rezultati dobiveni Onsagerovim modelom stavljaju se u kontekst iznosa duljine ustrajnosti DNK [45, 68] te osjetljivosti iste na promjene u temperaturi [71]. Navedeno može objasniti nagle promjene režima, iz fleksibilnog u režim savijanja, koji su uočljivi u eksperimentima [10]. Također, diskutira se činjenica da molekularno-dinamičke simulacije predviđaju prijelaz u tekući kristal na znatno većim koncentracijama [72] od eksperimentalno izmjerenih [10].

## 2.6 Viralna ejakcija u bakteriju

Saznanja iz prethodnih sekcija se primjenjuju na proces *in vivo* izbacivanja DNK iz bakteriofaga s repovima u bakterije. DNK se tretira kao da se nalazi unutar dva vezana spremnika s različitim pripadnim slobodnim energijama. Temeljna pretpostavka je da je DNK u nekondenziranom stanju u kapsidi (prvi spremnik), a u kondenziranom stanju u stanici (drugi spremnik).

Prenapučena sredina stanice teži kondenziranju DNK [30, 37], ali i njenom izbacivanju iz stanice [93]. Primjenom Ubbink-Odijk modela [16, 42] uz proračun doprinosa kondenzacije, saznajemo da se zbog kondenzacije DNK u stanici može poništiti oko 4 – 6 atm staničnog tlaka. Dodatne korekcije zbog točnog oblika DNK kondenzata, koji ovisi o vrsti bakteriofaga i duljini njegove DNK, mogu nešto smanjiti taj tlak. S druge strane, DNK u virusnoj kapsidi ima više doprinosa koji teže izbacivanju, no ti doprinosi teže nuli kako se

smanjuje količina zatočene DNK. Diskutiraju se utjecaji elektrostatskog odbijanja, energije savijanja i prostornog zatočenja DNK u rep bakteriofaga [94]. Uz pomoć termalnih fluktuacija, pokazuje se kako doprinosi zatočenja u bakteriofagu mogu premašiti još do 1 atm tlaka unutar stanica.

Tipične Gram-pozitivne bakterije imaju unutrašnje tlakove stanica u rasponu 0.1 – 5 atm, dok su Gram-negativne u rasponu 5-25 [21, 83, 84]. Stoga, zaključuje se da je mehanizam izbacivanja viralne DNK iz bakteriofaga u Gram-pozitivne bakterije moguće objasniti primjenom termodinamičkih argumenata. Uobičajeni pristup koji priziva aktivne mehanizme (stanične motore) ili druge doprinose izbacivanju [18] nije nužan kako bi objasnio *in vivo* ekcijiju DNK.

### 3 Nukleinske kiseline i kondenzirajući proteini u prostornom zatočenju

U ovom poglavlju proučava se, za viruse rijetki, mehanizam pakiranja nukleinskih kiselina uz pomoć kondenzirajućih proteina. Poznato je da proteini koji se mogu naći u jezgrama virusa mogu imati različite uloge prilikom sastavljanja i sazrijevanja virusa [5, 98, 99]. Neki virusi kao što su bakulovirus [100], adenovirus [101], mimivirus [102] i poxvirus [103] u jezgrama sadrže i proteine za koje je poznato da se vežu na DNK. Za adenoviruse [5] i poliomavirus SV40 [26, 104] postoje direktni dokazi da ti kondenzacijski proteini potpomažu sažimanje DNK. Ove vrste virusa su također obilježene nedostatkom ikozaedarske simetrije u svojim jezgrama koja se manifestira ravnim profilom gustoće [5, 26, 105, 106, 109], u suprotnosti s drugim sličnim vrstama virusa gdje se DNK pakira u guste uređene slojeve [19, 107, 108]. Nepostojanje teorijskih modela koji opisuju nedostatak uređenja i općenito pakiranje nukleinskih kiselina s kondenzirajućim (vezujućim) proteinima (bez kondenzacije) motivira daljnje proučavanje.

#### 3.1 Organizacija DNK i jezgrenih proteina u adenovirusima

Poznate su dvije temeljne teorije organizacije DNK u adenovirusima: jedna koja pretpostavlja dobro definiranu ikozaedarsku strukturu [114, 115, 116] te druga koja pretpostavlja formiranje "narukvice" poput one u kromatinu [117, 118, 119]. No, kako ne postoje dokazi ikozaedarske strukture, oba predložena modela su upitna [5, 105, 106]. U svrhu objašnjenja naizgled nasumične raspodjele kondenzirajućih proteina u jezgrama adenovirusa, vrši se statistička analiza eksperimentalnih podataka o uređenju nakupina jezgrenih proteina (tzv. "adenosomi").

## Statistička analiza položaja jezgrenih proteina

Položaji proteinskih čestica (adenosoma) dobiveni su iz Cryo-EM slika neoštetećenih virusnih čestica [97]. Eksperimenti ukazuju na postojanje oko 200 adenosoma, u skladu s drugim eksperimentima [118], za koje se pretpostavlja da vežu ili namataju adenovirusnu DNK. Prostorna raspodjela adenosoma prati se pomoću raspodjele vjerojatnosti nalaženja prvih susjeda na nekoj uzajamnoj udaljenosti, vjerojatnosti nalaženja drugog adenosoma na nekoj radijalno simetričnoj udaljenosti te putem radijalno simetrične gustoće čestica. Analiza pokazuje da iako su adenosomi neuređeno posloženi unutar kapside adenovirusa, položaji i međusobni odnosi adenovirusa nisu posljedica nasumičnog pakiranja. Dodatno, analiza ukazuje na dimenziju adenosoma koja je u skladu s očekivanom dimenzijom čestica dobivenih u razorenim kapsidama adenovirusa [118]. Interakcija između adenosoma je izrazito mekana, bez indikacija čvrstog vezanja kakvog bi očekivali u kromatinskom modelu s dobro definiranim duljinama veza. U gustoći adenosoma je vidljiv osiromašeni sloj uz površinu kapside. Gustoća pakiranja adenosoma i DNK je izrazito velika, te upućuje na prenapučenost koja može utjecati na interakcije.

## Modeliranje adenosoma

Kako bismo dodatno opisali unutrašnjost adenovirusa, koristi se Langevinova dinamika [129, 130]. Adenosomi se tretiraju kao efektivne čestice čija je meka interakcija [131] određena medijem DNK u kojem se nalaze. Time se prisustvo DNK tretira implicitno kroz efektivne interakcije adenosoma. Usporedba efektivnog modela čestica te efektivnog modela čestica adenosoma povezanih harmoničkom oprugom (veza), ukazuje kako model s dobro definiranim vezama ne odgovara eksperimentalnim rezultatima u proučavanju statističkim indikatorima.

## Organizacija jezgri kao mješavine efektivnih čestica

Temeljem modeliranja, utvrđeno je kako se adenosomi u kapsidi ponašaju kao efektivne čestice (kvazi-čestice) bez međusobnog vezanja. Prenapučeno okruženje DNK i proteina modificira osnovne interakcije. Povezanost DNK i proteina, ukoliko postoji, se može jedino manifestirati preko veza fleksibilnih ili slabo definiranih duljina. Model nam također omogućuje procjene donje granice unutaršnjeg tlaka viralne kapside od oko 0.1 atm, što je nekoliko redova veličina manje od tlaka izmjerenog kod gusto pakirane DNK u bakteriofagima [9, 19, 134].

## 3.2 Mješavine polimera i kondenzirajućih proteina u prostornom zatočenju

Dosadašnji modeli sažete DNK i kondenzirajućih proteina bili su temeljeni na implicitnoj prisutnosti DNK putem stvaranja veza i modificiranja interakcija između proteina [26, 97]. Temeljem modela analognog sustava bez prostornog zatočenja [137], očekuje se zanimljivi fazni dijagram. Stoga proučavamo takve mješavine u kontekstu pakiranja u adenovirusima [97] i poliomavirusima [26], s ciljem boljeg razumijevanja procesa pakiranja i unutarnje organizacije.

### Simuliranje putem molekularne dinamike

Mješavina polimera i sfera, koje predstavljaju kondenzirajuće proteine, se modelira putem Langevinove dinamike [129, 130]. Istovrsne čestice interagiraju putem odbojnih međudjelovanja, dok je interakcija kondenzirajućih proteina i polimera privlačna i nespecifična. Sve repulzivne interakcije su prikazane preko isključivo odbojnog Lennard-Jonesovog međudjelovanja, dok je veza između pojedinih kuglica ovisna o njihovoj međusobnoj udaljenosti [138] i lokalnoj savijenosti kako bi prikazivala krutost konture [76]. Navedena mješavina je zatočena u sferu koja sa svim česticama međudjeluje odbojno.

### Statistički indikatori

U svrhu opisivanja unutarnje organizacije polimera i kondenzirajućih čestica (proteina) koristimo dva indikatora unutarnje strukture te dva indikatora međusobne povezanosti polimera i kondenzirajućih čestica. Kao indikatori unutarnje organizacije koriste se radijalno simetrična gustoća vjerojatnosti nalaženja čestica  $\rho$  te radijalno simetrična raspodjela vjerojatnosti nalaženja čestica  $R$ . Kao indikatore međusobne povezanosti definiramo faktor omatanja  $w$  i faktor povezanosti  $\gamma$ . Faktor omatanja definira se kao najveća kontinuirana duljina polimera koji je u direktnom kontaktu (vezi) s istom kondenzirajućom česticom. Faktor povezanosti se definira kao najkraća udaljenost po konturi polimera koja povezuje dvije vezujuće čestice podijeljena s njihovom prostornom udaljenošću. Svi navedeni indikatori se konstruiraju temeljem više reprezentativnih konfiguracija sustava.

### Tlakovi pri pakiranju

Unutrašnji tlak u kapsidama koristi se kao indikator stabilnosti virusa [140, 141, 142]. U tu svrhu diskutira se kako različiti parametri kondenzirajućih proteina i polimera utječu na tlak prije proučavanja unutarnje organizacije. Prvo se proučava ponašanje tlaka u slučaju pakiranja samog polimera unutar zatočenja. Simulacije potvrđuju postojanje dva režima u ovisnosti o krutosti polimera: pri malim krutostima, kad je duljina ustrajnosti

manja od dimenzije sfernog zatočenja, polimer pokazuje skaliranja u skladu s teorijom fleksibilnih polimera [2, 12]. U slučaju kad je duljina ustrajnosti veća od dimenzije sfernog zatočenja, uočava se skaliranje dominirano duljinom ustrajnosti polimera, ali različito od onoga koje se dobiva u modelima srednjeg polja. U ovisnosti o koncentraciji, potvrđeno je postojanje polurazrijeđenog režima za fleksibilne polimere [11] te Flory režima za krute polimere [2, 12].

Prisustvo kondenzirajućih proteina uz polimer utječe na tlak. Ovisno o koncentraciji polimera, postoji optimalna koncentracija kondenzirajućih proteina koja smanjuje tlak u usporedbi sa zatočenim samim polimerom. Polimer s kondenzirajućim proteinima i dalje zadržava osnovne režime s obzirom na svoju duljinu ustrajnosti. Skaliranja s koncentracijom i duljinom ustrajnosti ostaju ista kao i bez proteina, osim u slučajevima kad je protein uspješan u smanjenju ukupnog tlaka.

## Unutarnja organizacija

Statistički indikatori omogućuju uvid u unutarnju organizaciju te usporedbu s eksperimentalnim rezultatima elektronske mikroskopije [97, 121] i raspršenja X-zraka pri malim kutovima [26]. Proučava se utjecaj snage vezanja proteina, duljine ustrajnosti polimera, dimenzije zatočenja te veličine kondenzirajućih čestica na unutarnju organizaciju mješavina kondenzirajućih proteina i polimera. Naglasak je na proučavanju dimenzija zatočenja koje odgovaraju adenovirusima i poliomavirusima te duljini ustrajnosti koja odgovara DNK.

Nadalje se promatra utjecaj različitih količina kondenzirajućih proteina te polimera na međusobne odnose kondenzirajućih proteina. Statistički indikatori gustoće i radijalne distribucije čestica ukazuju na postojanje režima u kojem kondenzirajući proteini na sebe omotaju različite dijelove polimera te međudjeluju indirektno – samo posredstvom svojih polimerskih omotača. Navedeno potvrđuje kvazičestičnu sliku korištenu pri modeliranju organizacije adenosoma u adenovirusima. Također, uočena je i pojava osiromašenog sloja čestica uz površinu sfernog zatočenja, sukladno analognoj situaciji kod adenovirusa.

Indikatori međusobne povezanosti polimera i kondenzirajućih proteina, faktor omatanja  $w$  i faktor povezanosti  $\gamma$ , pokazuju da su kondenzirajući proteini neizravno povezani putem polimera. Ne postoji jasna kontura DNK koja ih naizmjenice povezuje, kao kod kromatina u slobodnom prostoru [113]. Polimer se omata na kondenzirajući protein, te faktor omatanja prvenstveno ovisi o stupnju prenapučenosti prostora.

## Otvaranje kapsida

U svrhu proučavanja kako prostorno zatočenje utječe na pakiranje polimera i virusa, uspoređuju se statistički indikatori međusobne povezanosti sa i bez prostornog zatočenja.



Naglasak je na korištenju parametara koji bi najbliže odgovarali situaciji u adenovirusima. Ukidanje zatočenja za jako vezujuće proteine pokazuje relaksirane, ali i dalje kompaktne, strukture. Indikatori međusobne povezanosti ukazuju da nedostatak zatočenja i prenapučenosti, značajno smanjuje indeks omatanja kondenzirajućih proteina s polimerom. Postojanje kompaktnih struktura je sukladno s eksperimentima koji pokazuju kompaktne jezgre adenovirusa nakon razbijanja kapsida u *in vivo* uvjetima [121].

### Posljedice za pakiranje u virusima

Diskutiraju se rezultati prethodnih sekcija na unutarnju organizaciju virusa za koje postoje eksperimentalni podaci: adenoviruse [97] i poliomaviruse [26, 104]. Zaključuje se da je ovdje razvijeni model u stanju predvidjeti organizaciju jezgri tih virusa usprkos svojoj jednostavnosti. Model uspješno predviđa oblik profila gustoće čestica bez ikozaedralne simetrije te postojanje kvazi-čestica kondenzirajućih proteina. Nedostatak direktnog vezanja između proteina može upućivati na promjenu načina kako kondenzirajući proteini poput kromatina međudjeluju s DNK u prenapučenoj sredini. Prenapučenost sredine potiče direktnije kontakte između proteina i polimera te možda može mijenjati prirodu vezanja s nukleinskim kiselinama [122]. Postoji mogućnost primjene modela na druge, za sada neistražene, jezgre virusa: bakulovirus [100], mimivirus [102] i poxvirus [103].

## 4 Zaključak

Proučavajući različite faze sažete DNK karakterizirali smo kondenziranu DNK u živim stanicama te prostorno zatočenu DNK u virusnim kapsidama. Diskutiran je utjecaj rigidnosti molekule DNK, veličine zatočenja te elektrostatskog zasjenjenja. Navedeni teorijski formalizam je primijenjen na proces izbacivanja viralne DNK iz glave bakteriofaga u stanicu. Usprkos uvriježenim pretpostavkama da proces izbacivanja virusne DNK zahtjeva uključivanje aktivnih staničnih mehanizama, poput staničnih motora, pokazali smo da je za uspješnu ejakciju u Gram pozitivne bakterije dovoljno primjeniti termodinamičke argumente. Objašnjenje je temeljeno na ravnoteži potpomažućeg utjecaja DNK kondenzirane unutar stanice, zbog prostorne prenapučenosti i kondenzirajućih proteina, te energetske cijene zatočenja unutar malih virusa.

Iako je pakiranje DNK u prostornom zatočenju proučavano, malo je poznato o tome kako prisustvo kondenzirajućih proteina utječe na organizaciju (strukturu) unutrašnjosti kapsida. Adenovirusi i poliomavirusi spadaju u kategoriju virusa u kojima posebni proteini potpomažu pakiranje DNK u kapside. Detalji mehanizma pakiranja su nepoznati, no poznato je da pakirane sredice nemaju ikozaedralnu simetriju kao što je slučaj s kapsidama i mnogim drugim virusima te vrste. U svrhu boljeg razumijevanja pakiranja, analizirali smo eksperimentalne podatke o položaju nakupina proteina i DNK (tzv. adenosomi)

u jezgrama adenovirusa. Iz statističke analize međusobnog odnosa položaja adenosoma pokazali smo da su sredice virusa, iako naizgled nasumične, posložene po određenom obrascu. Svojstva statističkih raspodjela ukazuju na to da adenovirusi međudjeluju putem mekog potencijala uzrokovanog modifikacijama DNK medija u kojem se nalaze. Ne nalazimo dokaze o postojanju povezanosti između parova adonosoma, poput one koju bismo očekivali kod nukleosoma. Zaključujemo da su rezultatne interakcije posljedica toga da je unutrašnjost kapside prenapučena sredina koja mijenja efektivne interakcije između čestica.

Dosadašnje modeliranje jezgri virusa, koje sadrže nukleinske kiseline i kondenzirajuće proteine, bilo je temeljeno na modelima koji tretiraju prisustvo DNK implicitno kroz mijenjanje interakcija i stvaranja veza između proteina. Primjenom Langevinove dinamike simulirali smo mješavinu kondenzirajućih proteina koji se nespecifično vežu na polimer i eksplicitnog polimera (DNK ili RNK) u sferičnom prostornom zatočenju. Medij, uslijed prostorne prenapućenosti, mijenja efektivne interakcije. Posljedica je da zbog zatočenja polimer teži prekrivanju kondenzirajućih proteina. Pokazali smo da prisustvo polimera stvara kvazičestice poput adenosoma, potvrđujući valjanost efektivnog modela kojeg smo primijenili na adenovirusima. Također, ne pronalazimo stvaranje dobro definiranih veza između kondenzacijskih proteina. Veze između proteina se pokazuju nasumičnim i indirektnim. Modelom možemo uspješno objasniti općenite trendove u profilima gustoće i korelacije položaja jezgrinih proteina u adenovirusima te primijeniti principe na druge vrste poput poliomavirusa.

Modeliranje mješavina nukleinskih kiselina i kondenzirajućih proteina će u budućnosti imati sve veću važnost kako se eksperimentalne tehnike budu približavale direktnom opažanju neuređenih sredica [26, 97, 153]. Razvijeni model može se primijeniti na testiranje kako različite vrste kondenzirajućih proteina [122] utječu na unutarnju organizaciju i tlak kapsida. Otvorena je i mogućnost simuliranja eksperimenata stiskanja kapsida mikroskopom atomskih sila [135] s ciljem utvrđivanja odnosa proteina i nukleinskih kiselina [123]. Ovakve studije omogućuju proučavanje fenomena poput stanične prenapućenosti te utjecaja kojeg prenapučena sredina ima na pakiranje kromosoma [148, 149] ili savijanje proteina [27].



# Bibliography

- [1] M. Rubinstein and R. H. Colby, *Polymer Physics* (OUP Oxford, 2003).
- [2] P.-G. de Gennes, *Scaling Concepts in Polymer Physics*, 1st ed. (Cornell University Press, Ithaca and London, 1979).
- [3] J. F. Marko and E. D. Siggia, “Bending and twisting elasticity of DNA,” *Macromolecules* **27**, 981–988 (1994).
- [4] D. Andelman, in *Soft Condens. Matter Phys. Mol. Cell Biol.*, edited by W. C. K. Poon and D. Andelman (Taylor & Francis, 2006) pp. 97–122.
- [5] C. San Martin, “Latest insights on adenovirus structure and assembly.” *Viruses* **4**, 847–77 (2012).
- [6] B. N. Fields, D. M. Knipe, and P. M. Howley, *Fields’ Virology*, v. 1 (Wolters Kluwer Health/Lippincott Williams & Wilkins, 2007).
- [7] I. Kurtboke, ed., *Bacteriophages* (InTech, 2012).
- [8] E. Kutter and A. Sulakvelidze, eds., *Bacteriophages: Biology and Applications* (CRC Press, 2004).
- [9] A. Siber, A. L. Božič, and R. Podgornik, “Energies and pressures in viruses: contribution of nonspecific electrostatic interactions.” *Phys. Chem. Chem. Phys.* **14**, 3746–65 (2012).
- [10] A. Leforestier, “Polymorphism of DNA conformation inside the bacteriophage capsid.” *J. Biol. Phys.* **39**, 201–13 (2013).
- [11] A. Cacciuto and E. Luijten, “Self-avoiding flexible polymers under spherical confinement.” *Nano Lett.* **6**, 901–5 (2006).
- [12] T. Sakaue, “Semiflexible Polymer Confined in Closed Spaces,” *Macromolecules* **40**, 5206–5211 (2007).
- [13] T. Odijk, “Theory of lyotropic polymer liquid crystals,” *Macromolecules* **19**, 2313–2329 (1986).

- [14] N. N. Oskolkov, P. Linse, I. I. Potemkin, and A. R. Khokhlov, “Nematic ordering of polymers in confined geometry applied to DNA packaging in viral capsids.” *J. Phys. Chem. B* **115**, 422–32 (2011).
- [15] D. Svenšek and R. Podgornik, “Confined chiral polymer nematics: Ordering and spontaneous condensation,” *EPL* **100**, 66005 (2012).
- [16] J. Ubbink and T. Odijk, “Deformation of toroidal DNA condensates under surface stress,” *Europhys. Lett.* **33**, 353–358 (1996).
- [17] A. Leforestier, A. Siber, F. Livolant, and R. Podgornik, “Protein-DNA interactions determine the shapes of DNA toroids condensed in virus capsids.” *Biophys. J.* **100**, 2209–16 (2011).
- [18] I. Molineux and D. Panja, “Popping the cork: mechanisms of phage genome ejection,” *Nat. Rev. Microbiol.* **11**, 194–204 (2013).
- [19] M. De Frutos, A. Leforestier, and F. Livolant, “Relationship between the genome packing in the bacteriophage capsid and the kinetics of DNA ejection,” *Biophys. Rev. Lett.* **9**, 81–104 (2014).
- [20] D. E. Smith, S. J. Tans, S. B. Smith, S. Grimes, D. L. Anderson, and C. Bustamante, “The bacteriophage straight phi29 portal motor can package DNA against a large internal force.” *Nature* **413**, 748–52 (2001).
- [21] Y. Deng, M. Sun, and J. W. Shaevitz, “Direct Measurement of Cell Wall Stress Stiffening and Turgor Pressure in Live Bacterial Cells,” *Phys. Rev. Lett.* **107**, 158101 (2011).
- [22] P. K. Purohit, M. M. Inamdar, P. D. Grayson, T. M. Squires, J. Kondev, and R. Phillips, “Forces during bacteriophage DNA packaging and ejection.” *Biophys. J.* **88**, 851–66 (2005).
- [23] D. Van Valen, D. Wu, Y.-J. Chen, H. Tuson, P. Wiggins, and R. Phillips, “A single-molecule Hershey-Chase experiment.” *Curr. Biol.* **22**, 1339–43 (2012).
- [24] A. Lošdorfer Božič, A. Šiber, and R. Podgornik, “How simple can a model of an empty viral capsid be? Charge distributions in viral capsids,” *J. Biol. Phys.* **38**, 657–671 (2012).
- [25] A. Šiber and R. Podgornik, “Thermodynamics of nanospheres encapsulated in virus capsids,” *Phys. Rev. E* **81**, 1–11 (2010).

- 
- [26] G. Saper, S. Kler, R. Asor, A. Oppenheim, U. Raviv, and D. Harries, “Effect of capsid confinement on the chromatin organization of the SV40 minichromosome,” *Nucleic Acids Res.* **41**, 1569–1580 (2013).
- [27] A. Zaccone and E. M. Terentjev, “Theory of molecular crowding in Brownian hard-sphere liquids,” *Phys. Rev. E* **85**, 061202 (2012).
- [28] F. Höfling and T. Franosch, “Anomalous transport in the crowded world of biological cells.” *Rep. Prog. Phys.* **76**, 046602 (2013).
- [29] J. Pelta, D. Durand, J. Doucet, and F. Livolant, “DNA mesophases induced by spermidine: structural properties and biological implications.” *Biophys. J.* **71**, 48–63 (1996).
- [30] R. Golan, L. I. Pietrasanta, W. Hsieh, and H. G. Hansma, “DNA toroids: stages in condensation.” *Biochemistry* **38**, 14069–76 (1999).
- [31] J. Pelletier, K. Halvorsen, B.-Y. Ha, R. Paparcone, S. J. Sandler, C. L. Woldringh, W. P. Wong, and S. Jun, “PNAS Plus: Physical manipulation of the Escherichia coli chromosome reveals its soft nature,” *Proc. Natl. Acad. Sci.* **109**, E2649–E2656 (2012).
- [32] J. Kim, C. Jeon, H. Jeong, Y. Jung, and B.-Y. Ha, “A polymer in a crowded and confined space: effects of crowder size and poly-dispersity,” *Soft Matter* **11**, 1877–1888 (2015).
- [33] H.-X. Zhou, “Protein folding in confined and crowded environments,” *Arch. Biochem. Biophys.* **469**, 76–82 (2008).
- [34] S. Marion and A. Šiber, “Ejecting Phage DNA against Cellular Turgor Pressure,” *Biophys. J.* **107**, 1924–1929 (2014).
- [35] I. Lifshitz, A. Grosberg, and A. Khokhlov, “Some problems of the statistical physics of polymer chains with volume interaction,” *Rev. Mod. Phys.* **50**, 683–713 (1978).
- [36] J. Kindt, S. Tzlil, a. Ben-Shaul, and W. M. Gelbart, “DNA packaging and ejection forces in bacteriophage.” *Proc. Natl. Acad. Sci. U. S. A.* **98**, 13671–4 (2001).
- [37] S. Tzlil, J. T. Kindt, W. M. Gelbart, and A. Ben-Shaul, “Forces and pressures in DNA packaging and release from viral capsids.” *Biophys. J.* **84**, 1616–27 (2003).
- [38] S. Tzlil, M. Deserno, W. M. Gelbart, and A. Ben-Shaul, “A statistical-thermodynamic model of viral budding.” *Biophys. J.* **86**, 2037–48 (2004).

- [39] V. B. Teif and K. Bohinc, “Condensed DNA: condensing the concepts.” *Prog. Biophys. Mol. Biol.* **105**, 208–22 (2011).
- [40] 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.” *Biophys. J.* **61**, 246–59 (1992).
- [41] V. A. Bloomfield, “DNA condensation by multivalent cations.” *Biopolymers* **44**, 269–82 (1997).
- [42] J. Ubbink and T. Odijk, “Polymer- and salt-induced toroids of hexagonal DNA.” *Biophys. J.* **68**, 54–61 (1995).
- [43] O. Lambert, L. Letellier, W. M. Gelbart, and J. L. Rigaud, “DNA delivery by phage as a strategy for encapsulating toroidal condensates of arbitrary size into liposomes.” *Proc. Natl. Acad. Sci. U. S. A.* **97**, 7248–53 (2000).
- [44] T. X. Hoang, A. Giacometti, R. Podgornik, N. T. T. Nguyen, J. R. Banavar, and A. Maritan, “From toroidal to rod-like condensates of semiflexible polymers.” *J. Chem. Phys.* **140**, 064902 (2014).
- [45] C. G. Baumann, S. B. Smith, V. a. Bloomfield, and C. Bustamante, “Ionic effects on the elasticity of single DNA molecules.” *Proc. Natl. Acad. Sci. U. S. A.* **94**, 6185–90 (1997).
- [46] S. K. Lam, A. Pitrou, and S. Seibert, in *Proc. Second Work. LLVM Compil. Infrastruct. HPC - LLVM '15* (ACM Press, New York, New York, USA, 2015) pp. 1–6.
- [47] T. Odijk, “Physics of tightly curved semiflexible polymer chains,” *Macromolecules* **26**, 6897–6902 (1993).
- [48] V. A. Bloomfield, “Condensation of DNA by multivalent cations: considerations on mechanism,” *Biopolymers* **31**, 1471–81 (1991).
- [49] B. Schnurr, F. Gittes, and F. MacKintosh, “Metastable intermediates in the condensation of semiflexible polymers,” *Phys. Rev. E* **65**, 061904 (2002).
- [50] A. R. Khokhlov and A. Y. Grosberg, *Statistical Physics of Macromolecules* (AIP Press, 1994).
- [51] L. Onsager, “The effects of shape on the interaction of colloidal particles,” *Ann. N. Y. Acad. Sci.* **51**, 627–659 (1949).

- 
- [52] W. Jiang, J. Chang, J. Jakana, P. Weigele, J. King, and W. Chiu, “Structure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus,” *Nature* **439**, 612–616 (2006).
- [53] W. C. Earnshaw and S. C. Harrison, “DNA arrangement in isometric phage heads,” *Nature* **268**, 598–602 (1977).
- [54] R. Calendar and S. T. Abedon, eds., *The Bacteriophages*, 2nd ed. (Oxford University Press, 2006).
- [55] P. K. Purohit, J. Kondev, and R. Phillips, “Mechanics of DNA packaging in viruses.” *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3173–8 (2003).
- [56] N. Stoop, J. Najafi, F. K. Wittel, M. Habibi, and H. J. Herrmann, “Packing of Elastic Wires in Spherical Cavities,” *Phys. Rev. Lett.* **106**, 214102 (2011).
- [57] G. Morrison and D. Thirumalai, “Semiflexible chains in confined spaces,” *Phys. Rev. E* **79**, 011924 (2009).
- [58] J. Najafi, N. Stoop, F. Wittel, and M. Habibi, “Ordered packing of elastic wires in a sphere,” *Phys. Rev. E* **85**, 061108 (2012).
- [59] P. K. Purohit, J. Kondev, and R. Phillips, “Force steps during viral DNA packaging?” *J. Mech. Phys. Solids* **51**, 2239–2257 (2003).
- [60] P. G. de Gennes and J. Prost, *The Physics of Liquid Crystals*, 2nd ed. (Clarendon Press, 1993).
- [61] A. Khokhlov and A. Semenov, “Liquid-crystalline ordering in the solution of long persistent chains,” *Physica A* **108**, 546–556 (1981).
- [62] S. Brenner and V. Parsegian, “A physical method for deriving the electrostatic interaction between rod-like polyions at all mutual angles,” *Biophys. J.* **14**, 327–334 (1974).
- [63] F. Oosawa, *Polyelectrolytes* (M. Dekker, 1971).
- [64] H. Strey, V. Parsegian, and R. Podgornik, “Equation of State for DNA Liquid Crystals: Fluctuation Enhanced Electrostatic Double Layer Repulsion,” *Phys. Rev. Lett.* **78**, 895–898 (1997).
- [65] A. Leforestier and F. Livolant, “The bacteriophage genome undergoes a succession of intracapsid phase transitions upon DNA ejection.” *J. Mol. Biol.* **396**, 384–95 (2010).



- [66] G. Vroege and H. Lekkerkerker, “Phase transitions in lyotropic colloidal and polymer liquid crystals,” *Reports Prog. Phys.* **1241** (1992).
- [67] A. J. Spakowitz and Z.-G. Wang, “DNA packaging in bacteriophage: is twist important?” *Biophys. J.* **88**, 3912–23 (2005).
- [68] Y. Lu, B. Weers, and N. C. Stellwagen, “DNA persistence length revisited.” *Biopolymers* **61**, 261–75 (2002).
- [69] P. a. Wiggins, T. van der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker, and P. C. Nelson, “High flexibility of DNA on short length scales probed by atomic force microscopy.” *Nat. Nanotechnol.* **1**, 137–41 (2006).
- [70] A. K. Mazur and M. Maaloum, “DNA Flexibility on Short Length Scales Probed by Atomic Force Microscopy,” *Phys. Rev. Lett.* **112**, 068104 (2014).
- [71] S. Geggier, A. Kotlyar, and A. Vologodskii, “Temperature dependence of DNA persistence length.” *Nucleic Acids Res.* **39**, 1419–26 (2011).
- [72] A. S. Petrov, M. B. Boz, and S. C. Harvey, “The conformation of double-stranded DNA inside bacteriophages depends on capsid size and shape.” *J. Struct. Biol.* **160**, 241–8 (2007).
- [73] A. Ben-Shaul, “Entropy, energy, and bending of DNA in viral capsids.” *Biophys. J.* **104**, L15–7 (2013).
- [74] S. C. Harvey, “Comment on the letter by A. Ben-shaul: "Entropy, energy, and bending of DNA in viral capsids"."” *Biophys. J.* **106**, 489–92 (2014).
- [75] A. Ben-Shaul, “Reply to the comment by S. Harvey on "Entropy, energy, and bending of DNA in viral capsids"."” *Biophys. J.* **106**, 493–6 (2014).
- [76] C. Locker and S. Harvey, “A model for viral genome packing,” *Multiscale Model. Simul.* **5**, 1264–1279 (2006).
- [77] A. Dobrynin and M. Rubinstein, “Theory of polyelectrolytes in solutions and at surfaces,” *Prog. Polym. Sci.* **30**, 1049–1118 (2005).
- [78] C. Forrey and M. Muthukumar, “Langevin dynamics simulations of genome packing in bacteriophage.” *Biophys. J.* **91**, 25–41 (2006).
- [79] A. S. Petrov and S. C. Harvey, “Packaging double-helical DNA into viral capsids: structures, forces, and energetics.” *Biophys. J.* **95**, 497–502 (2008).
- [80] T. Odijk and F. Slok, “Nonuniform Donnan Equilibrium within Bacteriophages Packed with DNA,” *J. Phys. Chem. B* **107**, 8074–8077 (2003).

- 
- [81] T. Odijk, “Statics and dynamics of condensed DNA within phages and globules.” *Philos. Trans. A. Math. Phys. Eng. Sci.* **362**, 1497–517 (2004).
- [82] L. Ponchon, S. Mangenot, P. Boulanger, and L. Letellier, “Encapsidation and transfer of phage DNA into host cells: from in vivo to single particles studies.” *Biochim. Biophys. Acta* **1724**, 255–61 (2005).
- [83] M. Arnoldi, M. Fritz, E. Bäuerlein, M. Radmacher, E. Sackmann, and A. Boulbitch, “Bacterial turgor pressure can be measured by atomic force microscopy,” *Phys. Rev. E* **62**, 1034–1044 (2000).
- [84] C. São-José, M. de Frutos, E. Raspaud, M. A. Santos, and P. Tavares, “Pressure built by DNA packing inside virions: enough to drive DNA ejection in vitro, largely insufficient for delivery into the bacterial cytoplasm.” *J. Mol. Biol.* **374**, 346–55.
- [85] S. Köster, A. Evilevitch, M. Jeembaeva, and D. A. Weitz, “Influence of internal capsid pressure on viral infection by phage lambda.” *Biophys. J.* **97**, 1525–9 (2009).
- [86] S. Mangenot, M. Hochrein, J. Rädler, and L. Letellier, “Real-time imaging of DNA ejection from single phage particles.” *Curr. Biol.* **15**, 430–5 (2005).
- [87] P. Kemp, M. Gupta, and I. J. Molineux, “Bacteriophage T7 DNA ejection into cells is initiated by an enzyme-like mechanism.” *Mol. Microbiol.* **53**, 1251–65 (2004).
- [88] I. J. Molineux, “No syringes please, ejection of phage T7 DNA from the virion is enzyme driven,” *Mol. Microbiol.* **40**, 1–8 (2001).
- [89] D. Panja and I. J. Molineux, “Dynamics of bacteriophage genome ejection in vitro and in vivo.” *Phys. Biol.* **7**, 045006 (2010).
- [90] P. Grayson and I. J. Molineux, “Is phage DNA ‘injected’ into cells—biologists and physicists can agree.” *Curr. Opin. Microbiol.* **10**, 401–9 (2007).
- [91] M. M. Inamdar, W. M. Gelbart, and R. Phillips, “Dynamics of DNA ejection from bacteriophage.” *Biophys. J.* **91**, 411–20 (2006).
- [92] S. B. Zimmerman and L. D. Murphy, “Macromolecular crowding and the mandatory condensation of DNA in bacteria.” *FEBS Lett.* **390**, 245–8 (1996).
- [93] M. Castelnovo, R. K. Bowles, H. Reiss, and W. M. Gelbart, “Osmotic force resisting chain insertion in a colloidal suspension.” *Eur. Phys. J. E. Soft Matter* **10**, 191–7 (2003).
- [94] T. Odijk, “On the Statistics and Dynamics of Confined or Entangled Stiff Polymers,” *Macromolecules* **1344**, 1340–1344 (1983).

- [95] A. Evilevitch, J. W. Goyer, M. Phillips, C. M. Knobler, and W. M. Gelbart, "Measurements of DNA lengths remaining in a viral capsid after osmotically suppressed partial ejection." *Biophys. J.* **88**, 751–6 (2005).
- [96] A. S. Petrov, S. S. Douglas, and S. C. Harvey, "Effects of pulling forces, osmotic pressure, condensing agents and viscosity on the thermodynamics and kinetics of DNA ejection from bacteriophages to bacterial cells: a computational study." *J. Phys. Condens. Matter* **25**, 115101 (2013).
- [97] A. J. Perez-Berna, S. Marion, F. J. Chichon, J. J. Fernandez, D. C. Winkler, J. L. Carrascosa, A. C. Steven, A. Šiber, and C. San Martín, "Distribution of DNA-condensing protein complexes in the adenovirus core," *Nucleic Acids Res.* **43**, 4274–4283 (2015).
- [98] T. Dokland, "Scaffolding proteins and their role in viral assembly," *Cell. Mol. Life Sci.* **56**, 580–603 (1999).
- [99] L. Tong, "Viral Proteases," *Chem. Rev.* **102**, 4609–4626 (2002).
- [100] K. A. Tweeten, L. A. Bulla, and R. A. Consigli, "Characterization of an extremely basic protein derived from granulosis virus nucleocapsids," *J. Virol.* **33**, 866–76 (1980).
- [101] A. N. Giberson, A. R. Davidson, and R. J. Parks, "Chromatin structure of adenovirus DNA throughout infection." *Nucleic Acids Res.* **40**, 2369–76 (2012).
- [102] C. Xiao, Y. G. Kuznetso, S. Sun, S. L. Hafenstein, V. A. Kostyuchenko, P. R. Chipman, M. Suzan-Monti, D. Raoult, A. McPherson, and M. G. Rossmann, "Structural studies of the giant Mimivirus," *PLoS Biol.* **7**, 0958–0966 (2009).
- [103] R. C. Condit, N. Moussatche, and P. Traktman, "In A Nutshell: Structure and Assembly of the Vaccinia Virion," *Adv. Virus Res.* **65**, 31–124 (2006).
- [104] J. E. Germond, B. Hirt, P. Oudet, M. Gross-Bellark, and P. Chambon, "Folding of the DNA double helix in chromatin-like structures from simian virus 40." *Proc. Natl. Acad. Sci. U. S. A.* **72**, 1843–7 (1975).
- [105] M. Silvestry, S. Lindert, J. G. Smith, O. Maier, C. M. Wiethoff, G. R. Nemerow, and P. L. Stewart, "Cryo-Electron Microscopy Structure of Adenovirus Type 2 Temperature-Sensitive Mutant 1 Reveals Insight into the Cell Entry Defect," *J. Virol.* **83**, 7375–7383 (2009).
- [106] A. J. Pérez-Berná, R. Marabini, S. H. W. Scheres, R. Menéndez-Conejero, I. P. Dmitriev, D. T. Curiel, W. F. Mangel, S. J. Flint, and C. S. Martín, "Structure and Uncoating of Immature Adenovirus," *J. Mol. Biol.* **392**, 547–557 (2009).

- [107] F. P. Booy, W. W. Newcomb, B. L. Trus, J. C. Brown, T. S. Baker, and A. C. Steven, "Liquid-crystalline, phage-like packing of encapsidated DNA in herpes simplex virus." *Cell* **64**, 1007–15 (1991).
- [108] C. S. Martín, R. M. Burnett, F. de Haas, R. Heinkel, T. Rutten, S. D. Fuller, S. J. Butcher, and D. H. Bamford, "Combined EM/X-ray imaging yields a quasi-atomic model of the adenovirus-related bacteriophage PRD1 and shows key capsid and membrane interactions." *Structure* **9**, 917–30 (2001).
- [109] P. S. Shen, D. Enderlein, C. D. S. Nelson, W. S. Carter, M. Kawano, L. Xing, R. D. Swenson, N. H. Olson, T. S. Baker, R. H. Cheng, W. J. Atwood, R. Johne, and D. M. Belnap, "The structure of avian polyomavirus reveals variably sized capsids, non-conserved inter-capsomere interactions, and a possible location of the minor capsid protein VP4," *Virology* **411**, 142–152 (2011).
- [110] L. Schoonen and J. C. M. van Hest, "Functionalization of protein-based nanocages for drug delivery applications." *Nanoscale* **6**, 7124–41 (2014).
- [111] A. Amalfitano and R. J. Parks, "Separating fact from fiction: assessing the potential of modified adenovirus vectors for use in human gene therapy." *Curr. Gene Ther.* **2**, 111–33 (2002).
- [112] C. E. Thomas, A. Ehrhardt, and M. a. Kay, "Progress and problems with the use of viral vectors for gene therapy," *Nat. Rev. Genet.* **4**, 346–358 (2003).
- [113] H. Schiessel, "The physics of chromatin," *J. Phys. Condens. Matter* **15**, R699–R774 (2003).
- [114] D. T. Brown, M. Westphal, B. T. Burlingham, U. Winterhoff, and W. Doerfler, "Structure and composition of the adenovirus type 2 core." *J. Virol.* **16**, 366–87 (1975).
- [115] W. W. Newcomb, J. W. Boring, and J. C. Brown, "Ion etching of human adenovirus 2: structure of the core." *J. Virol.* **51**, 52–56 (1984).
- [116] M. L. Wong and M. T. Hsu, "Linear adenovirus DNA is organized into supercoiled domains in virus particles." *Nucleic Acids Res.* **17**, 3535–50 (1989).
- [117] J. Corden, H. M. Engelking, and G. D. Pearson, "Chromatin-like organization of the adenovirus chromosome." *Proc. Natl. Acad. Sci.* **73**, 401–404 (1976).
- [118] M. A. Mirza and J. Weber, "Structure of adenovirus chromatin." *Biochim. Biophys. Acta* **696**, 76–86 (1982).

- [119] M. Vayda, A. Rogers, and S. Flint, “The structure of nucleoprotein cores released from adenovirions,” *Nucleic Acids Res.* **1**, 441–460 (1983).
- [120] C. Devaux, P. A. Timmins, and C. Berthet-Colominas, “Structural studies of adenovirus type 2 by neutron and X-ray scattering,” *J. Mol. Biol.* **167**, 119–132 (1983).
- [121] A. J. Pérez-Berná, A. Ortega-Esteban, R. Menéndez-Conejero, D. C. Winkler, M. Menéndez, A. C. Steven, S. J. Flint, P. J. De Pablo, and C. San Martín, “The role of capsid maturation on adenovirus priming for sequential uncoating,” *J. Biol. Chem.* **287**, 31582–31595 (2012).
- [122] M. S. Luijsterburg, M. F. White, R. van Driel, and R. T. Dame, “The Major Architects of Chromatin: Architectural Proteins in Bacteria, Archaea and Eukaryotes,” *Crit. Rev. Biochem. Mol. Biol.* **43**, 393–418 (2008).
- [123] A. Ortega-Esteban, G. N. Condezo, A. J. Pérez-Berná, M. Chillón, S. J. Flint, D. Reguera, C. San Martín, and P. J. De Pablo, “Mechanics of Viral Chromatin Reveals the Pressurization of Human Adenovirus,” *ACS Nano* **9**, 10826–10833 (2015).
- [124] C. Likos, “Effective interactions in soft condensed matter physics,” *Phys. Rep.* **348** (2001).
- [125] M. Lei, A. de Graff, M. Thorpe, S. Wells, and A. Sartbaeva, “Uncovering the intrinsic geometry from the atomic pair distribution function of nanomaterials,” *Phys. Rev. B* **80**, 024118 (2009).
- [126] C. Song, P. Wang, and H. A. Makse, “A phase diagram for jammed matter,” *Nature* **453**, 629–632 (2008).
- [127] S. Nakano, D. Miyoshi, and N. Sugimoto, “Effects of molecular crowding on the structures, interactions, and functions of nucleic acids.” *Chem. Rev.* **114**, 2733–58 (2014).
- [128] K. W. Desmond and E. R. Weeks, “Random close packing of disks and spheres in confined geometries,” *Phys. Rev. E* **80**, 051305 (2009).
- [129] S. Plimpton, “Fast Parallel Algorithms for Short-Range Molecular Dynamics,” *J. Comput. Phys.* **117**, 1–19 (1995).
- [130] T. Schneider and E. Stoll, “Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions,” *Phys. Rev. B* **17**, 1302–1322 (1978).

- 
- [131] J. D. Weeks, D. Chandler, and H. C. Andersen, "Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids," *J. Chem. Phys.* **54**, 5237–5247 (1971).
- [132] S. Jun, A. Arnold, and B.-Y. Ha, "Confined Space and Effective Interactions of Multiple Self-Avoiding Chains," *Phys. Rev. Lett.* **98**, 1–4 (2007).
- [133] A. Minton, "Influence of excluded volume upon macromolecular structure and associations in 'crowded' media." *Curr. Opin. Biotechnol.* **8**, 65–9 (1997).
- [134] D. W. Bauer, J. B. Huffman, F. L. Homa, and A. Evilevitch, "Herpes Virus Genome, The Pressure Is On," *J. Am. Chem. Soc.* **135**, 11216–11221 (2013).
- [135] I. L. Ivanovska, P. J. de Pablo, B. Ibarra, G. Sgalari, F. C. MacKintosh, J. L. Carrascosa, C. F. Schmidt, and G. J. L. Wuite, "Bacteriophage capsids: Tough nanoshells with complex elastic properties," *Proc. Natl. Acad. Sci.* **101**, 7600–7605 (2004).
- [136] G. Archontis and T. Simonson, "Proton binding to proteins: a free-energy component analysis using a dielectric continuum model." *Biophys. J.* **88**, 3888–904 (2005).
- [137] G. Le Treut, F. Képès, and H. Orland, "Phase Behavior of DNA in the Presence of DNA-Binding Proteins," *Biophys. J.* **110**, 51–62 (2016).
- [138] K. Kremer and G. S. Grest, "Dynamics of entangled linear polymer melts: A molecular-dynamics simulation," *J. Chem. Phys.* **92**, 5057 (1990).
- [139] R. C. Liddington, Y. Yan, J. Moulai, R. Sahli, T. L. Benjamin, and S. C. Harrison, "Structure of simian virus 40 at 3.8-Å resolution," *Nature* **354**, 278–284 (1991).
- [140] A. Cordova, M. Deserno, W. M. Gelbart, and A. Ben-Shaul, "Osmotic shock and the strength of viral capsids." *Biophys. J.* **85**, 70–4 (2003).
- [141] R. Zandi and D. Reguera, "Mechanical properties of viral capsids," *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **72** (2005).
- [142] T. F. Anderson, C. Rappaport, and N. A. Muscatine, "On the structure and osmotic properties of phage particles." *Ann. Inst. Pasteur (Paris)*, **84**, 5–14 (1953).
- [143] Z. Jin, S. Zhao, and J. Wu, "Entropic forces of single-chain confinement in spherical cavities," *Phys. Rev. E* **82**, 041805 (2010).
- [144] P. Cifra and T. Bleha, "Free Energy of Polymers Confined in Open and Closed Cavities," *Macromol. Theory Simulations* **21**, 15–23 (2012).

- [145] C. R. Locker, S. D. Fuller, and S. C. Harvey, “DNA organization and thermodynamics during viral packing.” *Biophys. J.* **93**, 2861–9 (2007).
- [146] D. Marenduzzo, C. Micheletti, and E. Orlandini, “Biopolymer organization upon confinement.” *J. Phys. Condens. Matter* **22**, 283102 (2010).
- [147] T. Schlick, J. Hayes, and S. Grigoryev, “Toward Convergence of Experimental Studies and Theoretical Modeling of the Chromatin Fiber,” *J. Biol. Chem.* **287**, 5183–5191 (2012).
- [148] K. Maeshima, S. Hihara, and M. Eltsov, “Chromatin structure: does the 30-nm fibre exist in vivo?” *Curr. Opin. Cell Biol.* **22**, 291–297 (2010).
- [149] K. Luger, M. L. Dechassa, and D. J. Tremethick, “New insights into nucleosome and chromatin structure: an ordered state or a disordered affair?” *Nat. Rev. Mol. Cell Biol.* **13**, 436–47 (2012).
- [150] T. Baker, J. Drak, and M. Bina, “Reconstruction of the three-dimensional structure of simian virus 40 and visualization of the chromatin core,” *Proc. Natl. Acad. Sci. U. S. A.* **85**, 422–426 (1988).
- [151] D. L. Hurdiss, E. L. Morgan, R. F. Thompson, E. L. Prescott, M. M. Panou, A. Macdonald, and N. A. Ranson, “New Structural Insights into the Genome and Minor Capsid Proteins of BK Polyomavirus using Cryo-Electron Microscopy,” *Structure* **24**, 528–536 (2016).
- [152] M. Cyrklaff, A. Linaroudis, M. Boicu, P. Chlanda, W. Baumeister, G. Griffiths, and J. Krijnse-Locker, “Whole Cell Cryo-Electron Tomography Reveals Distinct Disassembly Intermediates of Vaccinia Virus,” *PLoS One* **2**, e420 (2007).
- [153] W. Wu, W. W. Newcomb, N. Cheng, A. Aksyuk, D. C. Winkler, and A. C. Steven, “Internal Proteins of the Procapsid and Mature Capsids of Herpes Simplex Virus 1 Mapped by Bubblegram Imaging,” *J. Virol.* **90**, 5176–5186 (2016).
- [154] A. Llauró, D. Luque, E. Edwards, B. Trus, J. Avera, D. Reguera, T. Douglas, P. de Pablo, and J. R. Caston, “Cargo-shell and cargo-cargo couplings govern the mechanics of artificially loaded virus-derived cages,” *Nanoscale* **8**, 9328–9336 (2016).

# Curriculum vitae

Sanjin Marion was born on July 8, 1987 in Rijeka, Croatia. In 2011 he obtained his Master's degree in Physics from the Department of Physics, Faculty of Science of the University of Zagreb in Croatia. Since 2011 he has been employed as a research assistant at the Institute of Physics in Zagreb, Croatia.

## List of publications:

1. R. D. Bulushev\*, S. Marion\*, E. Petrova\*, S. J. Davis\*, S. J. Maerkl, and A. Radenovic, "Single Molecule Localization and Discrimination of DNA-Protein Complexes by Controlled Translocation Through Nanocapillaries," *Nano Lett.* **16**, 7882–7890 (2016).
2. A. J. Perez-Berna, S. Marion, F. J. Chichon, J. J. Fernandez, D. C. Winkler, J. L. Carrascosa, A. C. Steven, A. Šiber, and C. San Martin, "Distribution of DNA-condensing protein complexes in the adenovirus core," *Nucleic Acids Res.* **43**, 4274–4283 (2015).
3. R. D. Bulushev, S. Marion, and A. Radenovic, "Relevance of the drag force during controlled translocation of a DNA-Protein complex through a glass nanocapillary," *Nano Lett.* **15**, 7118–7125 (2015).
4. S. Marion and A. Šiber, "Ejecting Phage DNA against Cellular Turgor Pressure," *Biophys. J.* **107**, 1924–1929 (2014).
5. D. Pelc, S. Marion, M. Požek, and M. Basletić, "Role of microscopic phase separation in gelation of aqueous gelatin solutions," *Soft Matter* **10**(2), 348 (2014).
6. S. Marion and B. Gumhalter, "Electron scattering by random adsorbates: A tunable decoherence mechanism in surface bands," *Phys. status solidi* **249**, 1218–1223 (2012).
7. D. Pelc, S. Marion, and M. Basletic, "Four-contact impedance spectroscopy of conductive liquid samples," *Rev. Sci. Instrum.* **82**, 5 (2011).