# STATISTICAL THERMODYNAMICS OF VIRUS ASSEMBLY

A Thesis
Presented to
The Academic Faculty

by

Se Il Lee

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Physics

Georgia Institute of Technology
May 2010

# STATISTICAL THERMODYNAMICS OF VIRUS ASSEMBLY

Approved by:

Professor Toan T. Nguyen, Adviser
School of Physics
*Georgia Institute of Technology*

Professor Harold D. Kim
School of Physics
*Georgia Institute of Technology*

Professor Stephen C. Harvey
School of Biology
*Georgia Institute of Technology*

Professor Daniel I. Goldman
School of Physics
*Georgia Institute of Technology*

Professor Jennifer E. Curtis
School of Physics
*Georgia Institute of Technology*

Date Approved: 31 March 2010

*To my family,*

# PREFACE

This thesis is about the study of bacteriophages and HIV (Human Immunodeficiency Virus). Chapter 1 provides introduction to both bacteriophages and HIV. Bacteriophages have drawn interest from the biophysical community because they have simple structures.

In Chapter 2, the physics of genome translocation of bacteriophages in the presence of $Mg^{+2}$ counterions is studied. Experiments have shown that $MgSO_4$ salt inhibits DNA ejection from bacteriophages nonmonotonically. There is a $MgSO_4$ concentration where the minimum amount of DNA is ejected. We propose that this is the result of DNA overcharging by $Mg^{+2}$ ions. The concept of charge inversion of the double stranded DNA (dsDNA) as bacteriophages genome is used to solve this problem. This explains why $Mg^{+2}$ counterions cannot condense DNA in solution.

In Chapter 3, the problem of DNA-DNA interaction mediated by divalent counterions is studied using computer simulations. Condensation of DNA in the presence of multivalent counterions draws attention because DNA ejection shows non-monotonic behavior in with counterion. Our simulation shows that if DNA configurational entropy is restricted, divalent counterions can cause DNA reentrant condensation similar to that caused by tri- or tetra-valent counterions. DNA-DNA interaction is strongly repulsive at small or large concentration and is negligible or slightly attractive for concentrations in between. This study supports the conclusion of non-monotonic behavior of bacteriophage DNA ejection obtained in Chapter 2.

Understanding the capsid assembly process of HIV, the causative agent of Acute Immune Deficiency Syndrome (AIDS), is very important because of recent interest in capsid oriented viral therapy. The unique conical shapes of mature HIV-1 capsid

have drawn significant interest in the biological community and have started to attract attention from physicists. To understand HIV-1 capsid assembly, three problems are focused on in this thesis: a study of diversity of *in vivo* assembled HIV-1 capsids, radial distribution of RiboNucleic Acid (RNA) genomes packaged inside spherical viruses, and RNA condensation in the presence of a single nucleocapsid (NC) protein. These will be discussed in Chapter 4, Chapter 5, and Chapter 6 respectively.

In Chapter 4, how the viral membrane affects the structure of *in vivo* assembled HIV-1 capsid is studied. Previous studies showed that in a free assembly process, the HIV-1 conical shape is not thermodynamically stable. The viral envelope membrane present during assembly imposes constraints on the length of the capsid. An elastic continuum shell theory is used to approximate the energies of various HIV-1 capsid shapes (spherical, cylindrical and conical) numerically and analytically. It is shown that conical and cylindrical shapes are both thermodynamically stable with the viral envelope membrane constraint.

In Chapter 5, the problem of RNA genomes packaged inside spherical viruses is studied. The viral capsid is modeled as a hollow sphere. The attraction between the inner viral capsid and RNA molecules occurs at the inner capsid surface only and plays an important role in the RNA packaging process. For weak attraction, RNA concentration is maximum at the center of the capsid to maximize configurational entropy. For stronger attraction, RNA concentration peaks near the capsid surface. In the latter case, competition between the branching of RNA secondary structure and its adsorption to the inner capsid results in formation of a dense layer of RNA near the capsid surface. A mean-field approach depending on the adsorption strength of RNA molecules and the inner viral capsid is used to determine how RNA molecules are packaged inside the viral capsid.

In Chapter 6, the condensation of RNA molecules by a single retrovirus NC protein is studied. The core of HIV is composed of a complex of genomic RNA and NC

proteins, surrounded by a shell of capsid proteins. The interaction between RNA molecules and NC proteins is important in the reverse transcription of viral RNA, which relates to the viral infectivity. We model a single NC protein as an infinite well at the origin representing the attractive RNA-NC protein interaction. For weak adsorption of the NC protein, only a small portion of RNA is condensed near the NC protein and the boundary distance $r_0$ between a dilute and a condensed phase of RNA concentration is linearly proportional to the adsorption strength. For strong adsorption, there is more condensed RNA so that $r_0$ is extended much farther than for the weak adsorption case. The latter shows that condensed RNA screens the NC protein.

# ACKNOWLEDGEMENTS

There are many people I wish to thank who have helped me during my graduate life at Georgia Tech. Foremost among them is my advisor, Toan Nguyen, without whom I would never be here now. I would like to thank Stephen Harvey who has been very supportive of my efforts and Jennifer Curtis who has helped me learn biological physics.

I would like to extend my appreciation to Dongjoo Lee, Jiil Choi, Huseyin Kurtuldu, Serkan Dursun, Serdar Ozdemir, Nikhil Sharma, Bokwon Yoon, Rui Zhang, Cathy Tran, Deborah Ortiz, Mike Sprinkle, Jeremy Hicks, Seungjoo Nah, my brother, Woo-il Lee, and my undergraduate advisor, Kyuman Cho for their great encouragement and advice. I would like to thank my wife, Eunsuk Cha, for tolerating the past few years of my devotion to this work. I owe her special degree of gratitude. My parents and parents-in-law deserve to get my gratitude for their consistent support and prayer. I would like to give my thanks to Donnie Murray for reading and editing this dissertation, my thesis committee for reading this dissertation, as well as Andy Zangwill and Mei-Yin Chou for helping me to complete my Ph.D degree here at Georgia Tech.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER I

# INTRODUCTION

## *1.1 Bacteriophage*

Bacteriophage, any one of a number of viruses that infect bacteria, has been called the "hydrogen atom of biology" because bacteriophage is among the simplest biological objects. Phage, a shortened name of bacteriophage is commonly used. About 100 hundred years ago, bacteriophages were identified by Frederick Twork [152] and Felix d'Herelle [37] independently. Bacteriophages, like other viruses, carry their genetic information in the form of either DNA or RNA. The Hershey-Chase experiment [65] about 60 years ago confirmed that DNA is genetic material and elucidated that viral DNA leaves its protein capsid to express the proteins required for the assembly of new phages. Bacteriophages are comprised of a protective capsid containing a single molecule of nucleic acids and a hollow, long, cylindrical tail attached to the capsid.

There are two types of bacteriophage life cycles, lytic and lysogenic, as shown in Figure 1. When the tail of bacteriophage recognizes its receptor of the host cell, the nucleic acids of the phage are transfered from the capsid into the cell through the tail. Activation of one set of phase genes results in the lytic cycle to replicate its genetic material and synthesize viral proteins. Empty capsids are synthesized and are packaged with viral genome. Finally, the collar, sheath, and base plates are attached to capsids, and tail fibers are added to complete the bacteriophage assembly before the burst of the cell. The number of newly formed bacteriophages during the lytic cycle is varied between 50 and 200. The phages that induce lytic infection are called virulent phages. Activation of another set of phage genes results in the lysogenic cycle, starting the integration of viral DNA into host cell genome. An integrated

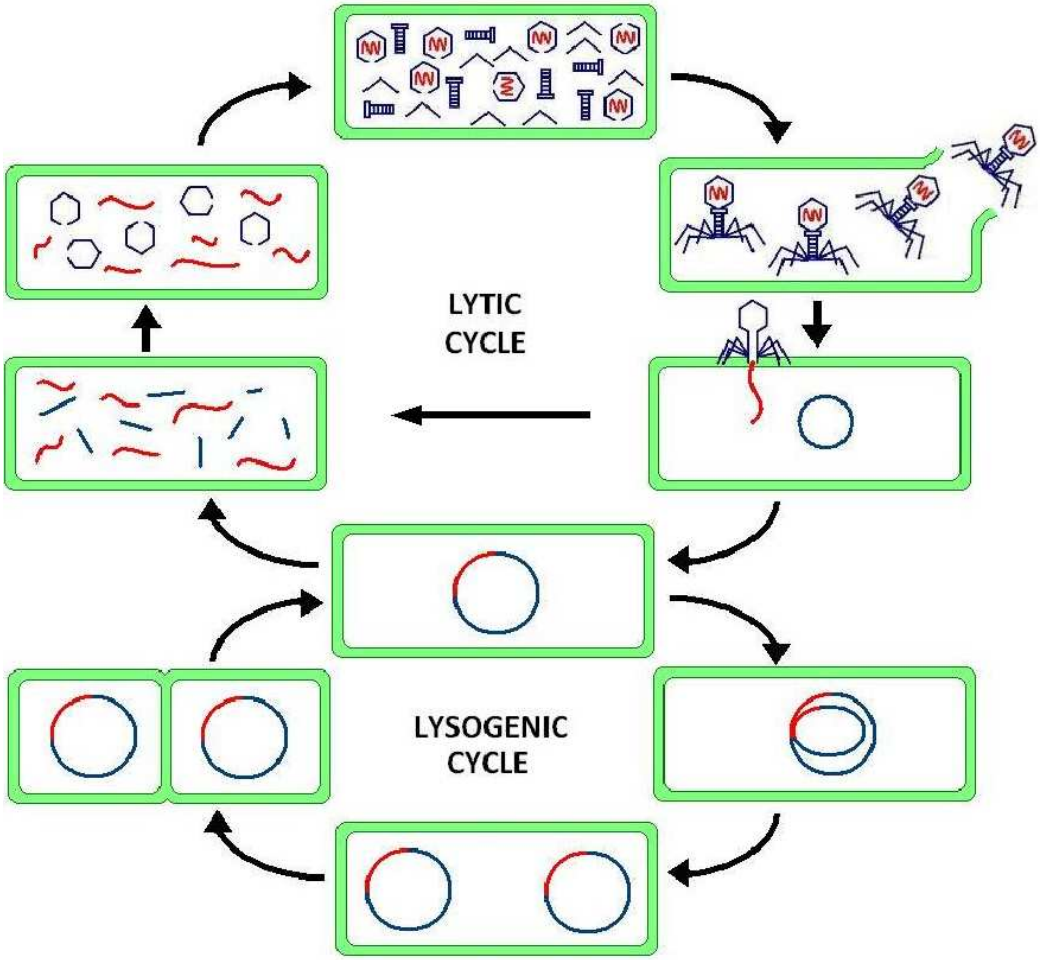


**Figure 1:** The life cycle of bacteriophage [Wikipedia Commons]. Bacteriophage follows either the lytic or lysogenic pathway depending on which set of viral genes is activated after viral DNA enters into the cell. Viral DNA is shown in red; host-cell DNA in blue;

bacteriophage is refered to a prophage. A host cell containing a prophage is called a lysogen and stays in a stable state. The phage genome is passed to all daughter cells at cell division. Completing cell division (binary fission), each cell has the phage DNA incorporated. These phages able to undergo lysogeny are known as temperate phages. Some physiological conditions (for example, DNA damage by UV induction) or some spontaneous induction can destabilize the prophage, resulting in its entry into the lytic cycle with the excised viral DNA from the host cell genome.

Most viral capsids are icosahedral, ranging in diameter from 20 to 100nm. Atomistic detail from cryo-electron microscopy [9] or X-ray studies of virus crystals [144] shows their structures. There are roughly two types of viruses. One is a a virus containing double-stranded DNA such as adenovirus, herpes, and phages T4, $\lambda$ and $\phi$29. The other is a virus containing single stranded RNA (ssRNA) such as polio, yellow fever, influenza, and Cowpea chlorotic mottle virus. There is a significant difference between these two types of viruses in assembly. Many ssRNA viruses experience a spontaneous process and capsid proteins self assemble. DNA viruses such as phages use ATP to package dsDNA into their capsids. Biologists and physicists have identified these different modes of assembly and have studied them by investigating the general physical processes such as the relationships among capsid size, charge, and genome confinement.

Two different types of experiments have been done to understand the physics of genome translocation such as (1) DNA packaging, and (2) DNA ejection.

(1) DNA packaging forces have been measured by single-molecule measurements. Smith and his collaborators measured the force exerted by the portal motor during the partial packaging of a single molecule of DNA into the capsid of phage $\phi$29 [139]. Fuller and his collaborators followed up on this measurement by adding a second optical trap which allowed them to study the packaging of the entire genome [127, 49, 48]. The relation between and packaging speed and force can be obtained by

maintaining a constant distance between beads and measuring the tension as DNA is packaged. They showed that packaging speed is highest at the initial stages and it slows down in the late stages because internal force from crowding and bending of the packaged DNA builds up, resulting in high pressure inside the capsid. It has also been confirmed that the $\lambda$ prohead undergoes a maturation process that produces an irreversible expansion of the capsid when the packaging percentage of the genome reaches 30%, and the capsid ruptures at 90 % of packaging [49].

(2) Evilevitch and his collaborators developed a quite different approach to measuring the packaging force in phage. They measured the force of DNA ejection instead of the force to package its genome. The experiments were carried out in solutions of phage $\lambda$ [43]. When LamB, a transmembrane protein involved in maltose transport in bacteria, is added to solutions of phage $\lambda$, it diffuses to the virus and attaches to the tail of phage $\lambda$. The capsid experiences a conformational change to open the capsid. All of the confined dsDNA are released into the solution of phage $\lambda$. If PEG (polyethylene glycol), a water soluble polymer, is introduced in the solution, it produces an opposing force to counterbalance the force of ejection because PEG is impermeable to the capsid. An ultra-visible absorbance measurement was used to measure how much DNA is ejected into solutions. This is done by digestion of the enzyme deoxyribonuclease(DNase), which cuts the ejected DNA into its nucleotides. They showed that the amount of DNA ejected decreases as osmotic pressure increases by increasing the concentration of PEG, and genome release is completely suppressed at around 20 atm external pressure. They also investigated the effect of crowding of DNA on DNA ejection by changing the length of the DNA genome. It has been confirmed that for the wild type (48.5 kilo base pairs (kbp)) virus, suppression occurs at 30 atm, whereas only 15 atm is required for a mutant with a 37.7 kbp of DNA [54]. Moreover, they studied the electrostatic interaction of DNA molecules on DNA ejection in the presence of multivalent counterions. They showed that the level of

osmotic pressure needed for full suppression is much less in the presence of 1mM of spermine, a tetravalent amine, than in the presence of divalent counterions such as $Mg^{+2}$ [42]. Similarly, Fuller and his collaborators showed that the packaging force in $\phi 29$ with the addition of $Mg^{+2}$ and $Cr^{+3}$ is reduced [50].

DNA packaging has also been studied using computer simulations. Simulations of DNA packaging provide insight into packaging configurations and forces associated with partially filled capsids [81, 28, 6, 140, 86, 96]. Kindt and coworkers performed Brownian dynamics simulations. In their model, the DNA is a semi-rigid string of spherical beads of Lennard-Jones particles 2.5 nm in diameter, linked to their neighbors by harmonic potentials. They showed that the force of packaging vs. the packaging fraction of DNA increases sharply when the repulsive interaction becomes dominant. For self attracting chains, donut-like toroids form at an initial stage and expand into a uniaxial spool-like structure with the addition of more beads, but they form disordered structures when there are no attractions between beads. Forrey and Muthukumar have carried out extensive Brownian dynamics simulations as well [47]. The interior capsid is modeled by planes representing the 20 faces of the icosahedron, and the chain is driven into the icosahedral capsid at a constant force (55 pN, consistent with the measured value of the $\phi 29$ motor), rather than a constant speed. They showed that in the late stages of packaging, the packaging process pauses and speeds up. For the attractive chains, the conformation in the late stages of packaging is similar to the folded toroid proposed by Hud [73]. Harvey and his collaborators have performed extensive molecular dynamics simulations of DNA packaging in phages. In their model [118], they determined the DNA-DNA interactions by fitting osmotic compression experiments [123]. Their DNA consists of 3212 beads, each of which corresponding to six consecutive base pairs. The DNA chain is linked by the harmonic bending and stretching potentials. They modeled capsids as pseudoatoms arranged

in polyhedra and the interactions between the DNA and the capsid atoms are represented by a semiharmonic restraining force. The DNA chain is driven into the capsid by a ratcheting motion in 1 nm steps, and the DNA is allowed to equilibrate after each step. They confirmed that the force of packaging DNA at each step agrees well with the observed packaging force in single molecule experiments. A decomposition of the internal energy changes in packaging shows that the electrostatic interaction is dominant and the bending energy is much smaller, which is consistent with theory. They also show that the conformation of the fully packaged DNA changes depending on capsid shape.

In the bacteriophage life cycle, genome translocation from capsid into the host cell is one of the interesting physics problems. DNA ejection in the presence of $Mg^{+2}$ counterions shows non-monotonic dependence on ion concentration [42]. In Chapter 2, DNA ejection from the capsid into solution in the presence of $Mg^{+2}$ counterions is studied based on the assumption of overcharging of DNA [89]. DNA-DNA attraction energy mediated by $Mg^{+2}$ counterions is obtained theoretically, which explaining why DNA does not condense in solution by addition of $Mg^{+2}$ counterions. This study also aims to explain the indirect experimental evidence of the charge inversion phenomenon of DNA induced by $Mg^{+2}$.

DNA-DNA electrostatic interaction mediated by divalent counterions plays an important role in DNA ejection problem of bacteriophage and is studied using computer simulation in Chapter 3. Packaging free energy of packaged and ejected segments is calculated. This problem explains DNA-DNA interaction in the presence of divalent counterions depending on the interaxial length between two DNA strands, showing DNA reentrant condensation. This study also supports one of the facts driven by our phenomenological theory in Chapter 2, that DNA can be overcharged depending on the amount of the concentration of divalent counterions, leading to the non-monotonic behavior of divalent counterions on the free energy of DNA packaging inside a DNA

bundle.

## *1.2 HIV*

### 1.2.1 AIDS

Acquired immune deficiency syndrome (AIDS) is a disease of the human immune system caused by the human immunodeficiency virus (HIV). A person is diagnosed with AIDS when their immune system fails, leading to life threatening opportunistic infections (an infection caused by pathogens). Infection with HIV is transmitted through blood, semen, vaginal fluid, preseminal fluid, and breast milk [1]. The major routes of this transmission are breast feeding, contaminated needles, unsafe sex, exchange between mother and baby during pregnancy (vertical transmission), or other exposure to bodily fluids. The first recognized cause of Acquired Immune Deficiency Syndrome (AIDS) occurred in the United States in the early 1980s. More than 25 million people died due to AIDS from 1981 to 2006 which is about 0.6% of the world's population [2]. AIDS is considered pandemic by WHO. In 2007, AIDS claimed an estimated 33.2 million lives worldwide, and estimated 2.1 million people were killed by AIDS including 330,000 children [2]. More than 75% of these deaths occured in sub-Saharan Africa, increasing poverty and retarding economic growth [11]. There is currently no known cure for AIDS, although antiviral treatment for AIDS can slow the course of the disease. Without treatment, about 90% of people infected by HIV will develop AIDS after 10-15 years [22]. Life expectancy of people infected by HIV is increased by treatment with antivirals. After HIV progresses to AIDS, the survival time with antiviral treatment is estimated to be more than 5 years [134]. Without antiviral therapy, people normally die within 1 year [106]. Antiretroviral drugs are expensive and not available in all countries, leading to the difficulty in controlling the AIDS pandemic.

### 1.2.2 HIV replication cycle

HIV replication cycle is shown in Figure 2. The life cycle includes binding and fusion, reverse transcription, integration, transcription, assembly and budding.

The cycle begins by viral entry into a target cell. HIV entry requires an interaction between a surface molecule (CD4) of the target cell and a surface unit (gp120) of HIV. This binding process involves the formation of a stable complex between gp120 and CD4. This is a highly specific interaction. The infection efficiency can be reduced by small molecular changes at the binding site of the receptor. Although CD4 is mandatory for gp120 attachment to target cells, CD4 alone is not enough for viral entry. Two kinds of co-receptors are needed for the fusion process. CCR-5 co-receptor is required by macrophage-tropic variants of HIV and CXCR-4 co-receptor is required by T-lymphotropic variants of HIV. After attachment of gp41 protein (Trans Membrane unit of HIV) to cell membrane which mediates fusion of viral envelope with cell membrane, viral genome, transcriptase, integrase, and other viral proteins start to penetrate into cell to begin replication. gp120 spreads over the infected cell membrane. Inside the cell, the capsid dissolves. The viral RNA enters the cytoplasm of a target cell. The next step is reverse transcription (the "retro" flow of genetic information from RNA to DNA) which is a unique characteristic of retrovirus. An HIV enzyme called reverse transcriptase converts the single-stranded HIV RNA to double-stranded HIV DNA using RNA as a template for DNA transcription, but it is a complex process including minus- and plus-strand transfer and RNA- and DNA-dependent polymerase transcription. The newly transcribed HIV DNA enters the host cell's nucleus, where an HIV enzyme called integrase enables the HIV DNA to be integrated into the DNA of the host cell. This integrated DNA is called provirus. When the host cell receives a signal to become active, the host's RNA polymerase transcribes one of the viral genomes and produces messenger RNA (mRNA). The viral mRNA leaves the nucleus and the host's ribosome translates the viral mRNA

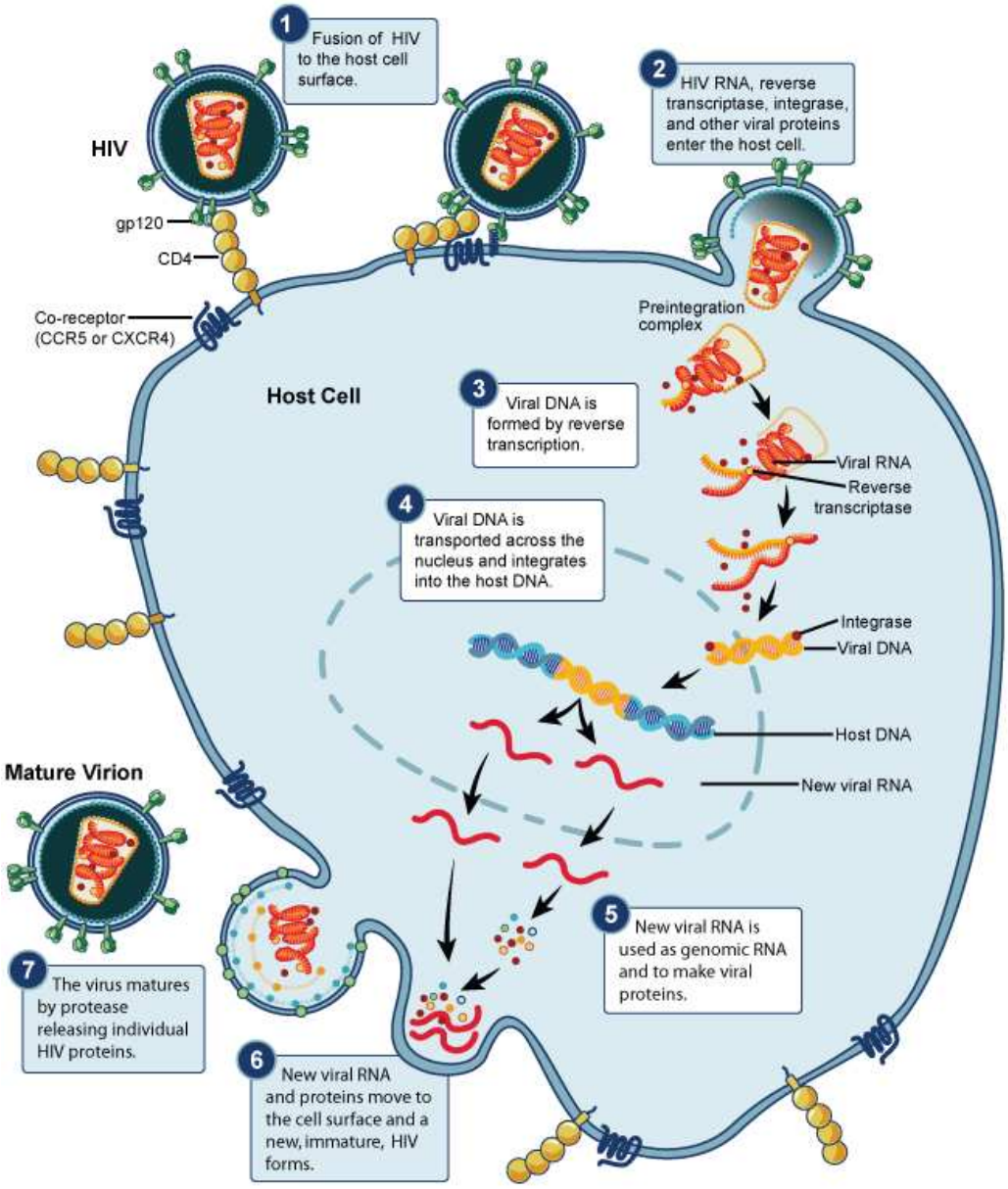


**Figure 2:** The life cycle of HIV [National Institute of Allergy and Infectious Diseases]. Viral RNA is shown in red; viral DNA, in yellow; host-cell DNA, in blue. Matrix, Capsid, and Nucleocapsid proteins are shown in blue, yellow, and light yellow spots respectively.

into viral proteins. Protease, an HIV enzyme, cuts the long HIV proteins into smaller individual proteins. Newly formed viral RNA and Gag and other proteins move to an inner cell surface of the target cell , self assemble into a spherical protein shell called "capsid", and bud out from the host cell. During the budding process, the new virus takes some part of the cell's outer envelope and takes back gp120 remaining on the host cell membrane. Gag-Gag attraction is known to be important in this process [166]. The new virion of HIV is called an immature HIV and changes into an infective mature HIV requiring a conformational change of capsid.

Maturation is an essential process for the HIV to become infective. Although the dominant shape of HIV-1 mature capsid is conical, there exist other irregular shapes of mature capsid shown in Figure 9. This capsid morphology problem stimulates curiosity and is dealt by elasticity theory with a constraint of HIV membrane size in Chapter 4 [91]. Chapter 4 explains that conical and cylindrical capsid geometries have equal probability to appear as a mature capsid.

RNA packaging problem also seems to be interesting in HIV immature spherical capsid which contains RNA as its genome material. The RNA distribution profile is originated from interaction strength between capsid protein and RNA. Adjusting the strength explains how RNA genome is distributed inside a spherical capsid. This problem can be applied to HIV immature spherical capsid and an analytic formulation is exploited to solve it in Chapter 5.

### 1.2.3 Nucleocapsid (NC) protein

NC protein plays numerous important roles during the replication cycle. NC protein is one of the Gag proteins. HIV NC protein is a 55 amino acid long, highly basic nucleic acid-binding protein containing two conserved zinc fingers of the cys-cys-his-cys (CCHC) motif separated by a short basic linker [16, 55, 32, 94, 126]. This highly conserved CCHC sequence is found either once or twice in most retroviral

NC proteins [16, 29] and strongly affects the nucleic acid binding specificity of the NC protein. For example, a mutant form NC protein with ser-ser-his-ser (SSHS) motif is unable to bind zinc and shows a lack of ability to facilitate the minus strand transfer step in reverse transcription [60]. Besides reverse transcription, NC protein is required for many HIV infection processes such as RNA genome dimerization, tRNA primer annealing, viral assembly, and DNA integration. Mature NC proteins coat the genomic RNA, in a dimeric form, in the viral particle. NC protein helps expedite strand transfer reactions by reducing the pausing rate of reverse transcription by destabilization the RNA secondary structures [5, 33, 77, 80, 116, 130, 131, 143, 163]. The NC domain in the Gag precursor plays an important role in specific recognition and packaging of RNA, since mutations to zinc fingers can reduce genome packaging [4, 39, 53, 60, 76, 104, 119, 159, 167]. Studies show that the annealing reaction between the tRNA and protein binding site is extremely slow without NC protein [61]. The NC protein, the genomic RNA, and cellular proteins, and membrane microdomains are also involved in the assembly of HIV in human cells [109] and the zinc fingers of HIV-1 NC were found to be important determinants of Gag assembly and localization in endosomes [56].

The research about the interaction between NC protein and RNA seems to be needed to help understand the role of NC protein in many replication processes. The interaction between a single NC protein and RNA is studied in Chapter 6 with the same analytic formulation applied in Chapter 5 with a different setup. This study explains that the high adsorption strength of NC protein shows a screening effect.

# CHAPTER II

# INHIBITION OF DNA EJECTION FROM BACTERIOPHAGE BY $MG^{+2}$ COUNTERIONS

## *2.1 Introduction*

Most bacteriophage, or viruses that infect bacteria, are composed of a DNA genome coiling inside a rigid, protective capsid. It is well-known that the persistence length $l_p$ of DNA is about 50 nm, comparable to or even larger than the inner diameter of the viral capsid. The genome of a typical bacteriophage is about 10 microns or 200 persistence lengths. Thus, the DNA molecule is considerably bent and strongly confined inside the viral capsid, resulting in a substantially pressurized capsid with an internal pressure as high as 50 atm [139, 43, 24, 118]. It has been suggested that this pressure is the main driving force for the ejection of the viral genome into the host cell when the capsid tail binds to the receptor in the cell membrane and subsequently opens the capsid. This idea is supported by various experiments both *in vivo* and *in vitro* [43, 24, 93, 19, 108, 122, 41, 42]. The *in vitro* experiments additionally revealed possibilities of controlling the ejection of DNA from bacteriophages. One example is the addition of PEG (polyethyleneglycol), a large molecule that is incapable of penetrating the viral capsid. A finite PEG concentration in solution produces an osmotic pressure on the capsid. This in turn leads to a reduction or even complete inhibition of the ejection of DNA.

Since DNA is a strongly charged molecule in aqueous solution, the screening condition of the solution also affects the ejection process. At a given external osmotic pressure, by varying the salinity of solution, one can also vary the amount of DNA ejected. Interestingly, it has been shown that monovalent counterions such as NaCl

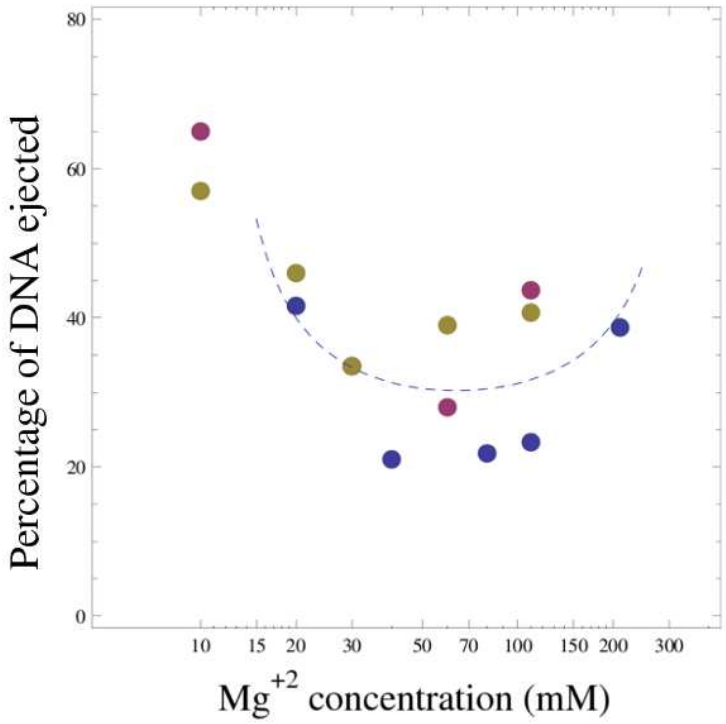


**Figure 3:** Inhibition of DNA ejection depends on $MgSO_4$ concentration for bacteriophage $\lambda$ at 3.5 atm external osmotic pressure. Solid circles represent experimental data [42, 44]. The three different colors correspond to three different sets of data. The dashed line is a theoretical fit of our theory.

have negligible effect on the DNA ejection process [43]. In contrast, multivalent counterions such as $Mg^{+2}$, $CoHex^{+3}$ (cobalt hexamine), $Spd^{+3}$ (spermidine) or $Spm^{+4}$ (spermine) exert strong effect. One such result is shown in Figure 3, where the solid circles represent experimental data for the percentage of ejected DNA from bacteriophage $\lambda$ (at 3.5 atm external osmotic pressure) as a function of $MgSO_4$ concentration [42]. Evidently, the effect of multivalent counterions on the DNA ejection is non-monotonic. There is an optimal $Mg^{+2}$ concentration where the least DNA genome is ejected from the phages. Similar qualitative behavior is observed for other multivalent counterions.

In this chapter, we focus on understanding the electrostatics involved in the inhibition of DNA ejection by $Mg^{+2}$ counterions. We propose that the non-monotonic behavior observed in Figure 3 is the result of $Mg^{+2}$ ions inducing an effective attraction between DNA segments inside the capsid, and the so-called overcharging of DNA by multivalent counterions in free solution. The proposed $Mg^{+2}$-mediated attraction between neighboring DNA segments inside the capsid is a central argument of this chapter and needs to be clarified. It is well-known that $Mg^{+2}$ ions do not condense free DNA molecules in aqueous solution [124, 74]. However, we argue that due to the entropic confinement of the viral capsid, DNA is strongly bent and thermal fluctuations of DNA molecule are strongly suppressed compared to those in free solution. It is due to this unique setup of the bacteriophage, where DNA is pre-packaged by a motor protein during virus assembly, that $Mg^{+2}$ ions can induce attractions between DNA. It should be mentioned that $Mg^{+2}$ counterions are shown *experimentally* to be able to condense DNA in another confined system: the condensation of DNA in two dimensions [84]. Therefore it is not surprising that $Mg^{+2}$ ions can cause DNA-DNA attractions inside the capsid (a zero-dimensional system).

The overall electrostatics of $Mg^{+2}$ modulated DNA ejection from bacteriophages is the following: Due to strong electrostatic interaction between DNA and $Mg^{+2}$ counterions, the counterions condense on the DNA molecule. As a result, the net charge of DNA ($\eta^*$ per unit length) which is the sum of the "bare" DNA charges ($\eta_{bare} = -1e/1.7\text{Å}$) and the charges of condensed counterions becomes smaller in magnitude than the "bare" charge. There are strong correlations between the condensed counterions at the DNA surface which cannot be described using standard Poisson-Boltzmann mean-field theory. Strongly correlated counterion theories, various experiments and simulations [136, 57, 105] have showed that when these strong correlations are taken into account, $\eta^*$ is not only smaller than $\eta_{bare}$ in magnitude but can even have opposite sign: this is known as the *charge inversion* phenomenon.

Specifically, the degree of condensation, hence $\eta^*$, depends logarithmically on the concentration of multivalent counterions, $N$. As $N$ increases from zero, $\eta^*$ becomes less negative, neutral and eventually positive. We propose that the multivalent counterion concentration $N_0$, where DNA net charge is neutral, corresponds to the optimal inhibition due to $Mg^{+2}$ induced DNA-DNA attraction inside the capsid. At lower or higher concentrations, $\eta^*$ is either negative or positive. As a charged molecule at these concentrations, DNA prefers to be in solution to lower its electrostatic self-energy (due to its geometry, the capacitance of DNA molecule is higher in free solution than in the bundle inside the capsid). Accordingly, this leads to a higher percentage of ejected viral genome. The dashed line in Figure 3 is a fit of our theoretical result to the experimental data of $MgSO_4$. The optimal $Mg^{+2}$ concentration is shown to be $N_0 = 64$ mM. The $Mg^{+2}$-mediated attraction between DNA double helices is found to be -0.004 $k_BT$/base ($k_B$ is the Boltzmann constant and $T$ is the temperature of the system).

## *2.2 Derivation of the DNA ejected length*

We begin by writing the total energy of the DNA molecule as the sum of the energy of the DNA segment ejected outside the capsid with length $L_o$ and the energy of the DNA segment remaining inside the capsid with length $L_i = L - L_o$, where $L$ and $L_o$ are the ejected and the total length of the viral DNA genome respectively:

$$E_{tot}(L_o) = E_{in}(L_i) + E_{out}(L_o) \tag{1}$$

### 2.2.1 Energy outside capsid

Because the ejected DNA segment is under no confinement, we neglect contributions from bending energy and approximate $E_{out}$ by the electrostatic energy of a free DNA of the same length in solution. Treating the DNA molecule as a uniformly charged

cylinder with radius $a$ and linear charge density $\eta^*$, one obtains:

$$E_{out}(L_o) = -L_o(\eta^{*2}/D)\ln(1 + r_s/a), \tag{2}$$

where $D = 78$ is the dielectric constant of water and $r_s$ is the Debye-Hückel screening length of the solution. The negative sign signifies the fact that the system of the combined DNA and the condensed counterions is equivalent to a cylindrical capacitor under constant charging potential. The net linear charge density of DNA, $\eta^*$, is a function of the counterion concentration $N$ [136, 57]:

$$\eta^* = -(\eta_c/2Z)\ln(N_0/N)/\ln(1 + r_s/a), \tag{3}$$

where $\eta_c = Dk_BT/e$ is Manning critical charge density and $Z$ is the counterion valence. The constant concentration at which DNA is neutral, $N_0$, can be interpreted as the concentration of counterions next to the condensed counterion layer on the DNA surface. A simple derivation for $\eta^*$ can be obtained by dividing the counterion population into two groups, a "bound" (condensed) counterion layer on the DNA and a "free" counterion population in solution [113]. The distribution of the latter is assumed to obey Boltzmann statistics:

$$N(r) = N\exp[-Ze\phi(r)/k_BT] \tag{4}$$

with $\phi(r)$ being the electrostatic potential at radial distance r from DNA central axis. Denoting $N_0 = N(a)$, one immediately gets

$$\phi(a) = -(k_BT/Ze)\ln(N_0/N). \tag{5}$$

On the other hand, the surface potential $\phi(a)$ of a charged cylinder with charge density $\eta^*$ in Debye-Hückel approximation is given by [162].

$$\phi(a) = \frac{2\eta^*}{D}\frac{K_0(a/r_s)}{(a/r_s)K_1(a/r_s)} \simeq \frac{2\eta^*}{D}\ln\left(1 + \frac{r_s}{a}\right) \tag{6}$$

where $K_{0,1}$ are Bessel functions (this expression is twice the value given in Ref. [162] because we assume that the screening ion atmosphere does not penetrate the DNA cylinder.) Eliminating $\phi(a)$ from Eqs.(5) and (6), one gets Eq. (3).

Obtaining the concentration $N_0$ using first principle calculations is a complicated and non-trivial task [136, 57, 105]. In general, it depends on the correlation between "bound" counterions at the DNA surface and its competition with the counterion entropy. However, in practical situations, DNA is almost neutralized ($|\eta_{bare}/\eta^*| >> 1$) by the counterions. Therefore, $N_0$ can be very well assumed to be independent of $N$ and $\eta^*$. Within the scope of this chapter, we treat it as a phenomenological constant concentration whose value is obtained by fitting the result of our theory to the experimental data.

### 2.2.2 Energy inside capsid

The energy of the DNA segment inside the viral capsid comes from the bending energy of the DNA coil and the interaction between neighboring DNA double helices:

$$E_{in}(L_i, d) = E_{bend}(L_i, d) + E_{int}(L_i, d) \tag{7}$$

where $d$ is the average DNA-DNA interaxial distance. To calculate $E_{bend}$, we employ the viral DNA packaging model used previously [122, 128, 81]. In this model, the genome coils co-axially inward with the neighboring DNA helices forming a hexagonal lattice with lattice constant $d$ (Figure 4). For a spherical capsid, this model gives:

$$E_{bend}(L_i, d) = \frac{4\pi l_p k_B T}{\sqrt{3}d^2} \left\{ -\left( \frac{3\sqrt{3}L_i d^2}{8\pi} \right)^{1/3} + R \ln \frac{R + \left(3\sqrt{3}L_i d^2/8\pi\right)^{1/3}}{\left(R^2 - \left(3\sqrt{3}L_i d^2/8\pi\right)^{2/3}\right)^{1/2}} \right\}, \tag{8}$$

where $R$ is the radius of the inner surface of the viral capsid.

To calculate $E_{int}(L_i, d)$, we notice that multivalent counterions can induce condensation of free DNA in solution [124, 74]. DNA molecules in such a condensate are almost neutralized by the counterions and are arranged in a hexagonal lattice (similar

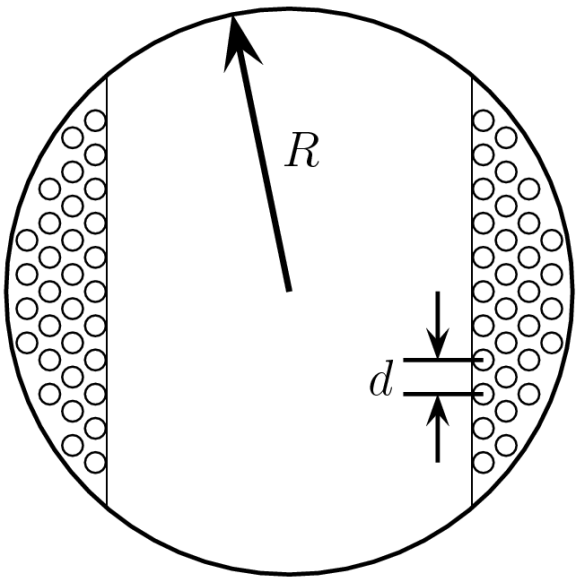


**Figure 4:** A very simplified model of bacteriophage genome packaging. The viral capsid is modeled as a rigid spherical cavity. The DNA inside coils co-axially inward. Neighboring DNA helices form a hexagonal lattice with lattice constant $d$. A sketch for a cross section of the viral capsid is shown.

to our viral DNA packaging arrangement) with an equilibrium interaxial distance $d_0$. The value of $d_0$ depends on the both valency and the type of counterions used, but is typically about 2.8 nm. As mentioned earlier we argue that, in the confinement of the viral capsid, $Mg^{+2}$ also induces attraction between DNA segments. Neglecting finite size effect, we approximate the interaction energy the viral DNA coiling inside the capsid as

$$E_{int}(L_i, d_0) = -L_i\epsilon, \tag{9}$$

where $-\epsilon$ is the DNA-DNA attraction per unit length. Like the aforementioned parameter $N_0$, we treat $\epsilon$ and $d_0$ as constant fitting parameters of our theory. In total, we have three fitting parameters $(N_0, \epsilon, d_0)$ and three fitting constraints (the two coordinates of the minimum and the curvature of the curve $L_o(N)$ in Figure 3). Thus our theory does not contain unnecessary degrees of freedom.

Obviously, due to the strong confinement of the viral capsid, the interaxial distance

$d$ between neighboring DNA double helices inside the capsid is smaller than the equilibrium distance $d_0$ inside the condensate. The experiments from Ref.([124]) provided an empirical formula that relates the restoring force to the difference $d_0 - d$. Integrating this restoring force with $d$, one obtains an expression for the interaction energy between DNA helices for a given interaxial distance $d$:

$$E_{int}(L_i, d) = L_i\sqrt{3}F_0\left[(c^2 + cd)\exp\left(\frac{d_0 - d}{c}\right) - (c^2 + cd_0) - \frac{1}{2}(d_0^2 - d^2)\right] - L_i\epsilon \quad (10)$$

where the empirical values of the constants $F_0$ and $c$ are 0.5 pN/nm$^2$ and 0.14 nm respectively.

### 2.2.3 Free energy minimization

Equation (1) together with equations (2), (7),(8) and (10) provide the complete expression for the total energy of the DNA genome of our theory. For a given external osmotic pressure, $\Pi_{\rm osm}$, and a given multivalent counterion concentration, N, the equilibrium value for the ejected DNA genome length $L_o^*$ is the length that minimizes the total free energy $G(L_o)$ of the system, where

$$G(L_o) = E_{tot}(L_o) + \Pi_{\rm osm}L_o\pi a^2. \quad (11)$$

Here, $L_o\pi a^2$ is the volume of the ejected DNA segment in aqueous solution. The specific procedure is following. The energy $E_{in}(L - L_o, d)$ of the DNA segment inside the capsid is minimized with respect to $d$ to acquire the optimal DNA-DNA interaxial distance for a given DNA ejected length, $d^*(L_o)$. Then, we substitute $E_{tot}(L_o) = E_{in}[L - L_o, d^*(L_o)] + E_{out}(L_o)$ into Eq. (11) and optimize $G(L_o)$ with respect to $L_o$ to obtain the equilibrium ejected length $L_o^*(\Pi_{\rm osm}, N)$. The details of the derivation to the length $L_o^*(\Pi_{\rm osm}, N)$ is shown in Appendix A. By fitting this $L_o^*$ with experiment data we can obtain the values for the neutralizing counterion concentration, $N_0$, the $Mg^{+2}$-mediated DNA-DNA attraction, $-\epsilon$, and the equilibrium DNA-DNA distance $d_0$. The result of fitting our theoretical ejected length $L_o^*$ to the experimental data of

Ref. [42] is shown in Figure 3. In this experiment, wild type bacteriophages $\lambda$ was used, so $R = 29$ nm and $L = 16.49$ $\mu$m [9]. $\Pi_{\rm osm}$ is held fixed at 3.5 *atm* and the $Mg^{+2}$ counterion concentration is varied from 10 mM to 200 mM. The fitted values are found to be $N_0 = 64$ mM, $\epsilon = 0.004$ $k_BT$ per nucleotide base, and $d_0 = 2.73$ nm.

## *2.3 Discussion*

The strong influence of the multivalent counterions on the process of DNA ejection from bacteriophage appears in several aspects of our theory and is easily seen by setting $d = d_0$, thus neglecting the weak dependence of $d$ on $L_i$ and using Eq. (9) for DNA-DNA interaction inside the capsid. Firstly, the attraction strength $\epsilon$ appears in the expression for the Gibbs free energy, Eq. (11), with the *same* sign as $\Pi_{\rm osm}$. In other words, the attraction between DNA strands inside capsid acts as an additional "effective" osmotic pressure preventing the ejection of DNA from bacteriophage. This switch from repulsive DNA-DNA interaction for monovalent counterion to attractive DNA-DNA interaction for $Mg^{+2}$ leads to an experimentally observed decrease in the percentage of DNA ejected from 50 % for monovalent counterions to 20% for $Mg^{+2}$ counterions at optimal inhibition ($N = N_0$). Secondly, the electrostatic energy of the ejected DNA segment given by Eq. (2) is logarithmically symmetrical around the neutralizing concentration $N_0$. This is clearly demonstrated in Figure 3 where the log-linear scale is used. This symmetry is very similar to the behavior of another system that exhibits charge inversion phenomenon, the non-monotonic swelling of macroion by multivalent counterions [138].

It is very descriptive to compare our fitting values for $\epsilon$ and $N_0$ to those obtained for other multivalent counterions. Fitting done for the DNA condensation experiments by $Spm^{+4}$ and $Spd^{+3}$ shows $\epsilon$ to be 0.07 and 0.02 $k_BT$/base respectively [124, 113]. For our case of $Mg^{+2}$, a divalent counterion, and bacteriophage $\lambda$ experiment, $\epsilon$ is found to be 0.004 $k_BT$/base. This is quite reasonable since $Mg^{+2}$ is a much weaker

counterion. Furthermore, $N_0$ was found to be 3.2 mM for the tetravalent counterion, 11 mM for the trivalent counterion. Our fit of $N_0 = 64$ mM for divalent counterions again is in favorable agreement with these independent fits. Note that, in the limit of high counterion valency ($Z \to \infty$), $N_0$ is shown to vary *exponentially* with $-Z^{3/2}$ [136, 57, 105]. The large increase in $N_0$ from 3.2 mM for tetravalent counterions to 64 mM for divalent counterions is not surprising.

The fitted value $-\epsilon = -0.004\ k_BT$ per base explains why $Mg^{+2}$ ions cannot condense DNA in free solution. It corresponds to an attraction of $-1.14 k_BT$ per persistence length. Since thermal fluctuation energy of a polymer is about $k_BT$ per persistence length, this attraction is too weak to overcome thermal fluctuations. It therefore cannot condense free DNA in solution. Only in the confinement of the viral capsid, this attraction effect appears in the ejection process. It should be mentioned that computer simulation of DNA condensation by idealized divalent counterions [99] does show a weak short-range attraction comparable to our $\epsilon$.

The phenomenological constants $-\epsilon$ and $N_0$ depend strongly on the strength of the correlations between multivalent counterions on the DNA surface. The stronger the correlations, the greater the value $\epsilon$ and the smaller the concentration $N_0$. Evidently, for the marginal case of $Mg^{+2}$ counterions, co-ions can also play important role affecting these correlations. Experimentally, the non-monotonic behavior is observed for tri-, tetra-valent counterions and for $Mg^{+2}$ and $SO_4^{-2}$ co-ions. But, up to 100 mM concentration, the minimum is not observed for $Mg^{+2}$ with $Cl^-$ co-ions. However, we believe $Mg^{+2}$ correlations on the surface of DNA are reduced in the presence of $Cl^-$ co-ions, leading to a smaller $\epsilon$ and a much higher $N_0$ ($N_0$ depends *exponentially* on the correlation strength), thus the minimum is not observed up to 100 mM. This is consistent with experimental fact that DNA ejection inhibition at $N = N_0$ for $MgSO_4$ salt is stronger than inhibition for all concentrations of $MgCl_2$ salt up to 100 mM.

# CHAPTER III

# DNA-DNA ELECTROSTATIC INTERACTION MEDIATED BY DIVALENT COUNTERIONS

## *3.1 Introduction*

The problem of DNA condensation in the presence of multivalent counterions has seen a strong revival of interest in recent years. This is because of the need to develop effective ways of gene delivery for the rapidly growing field of genetic therapy. DNA viruses such as bacteriophages provide excellent study candidates for this purpose. One can package genomic DNA into viruses, then deliver and release the molecule into targeted individual cells. Recently there is a large biophysical literature dedicated to the problem of DNA condensation (packaging and ejection) inside bacteriophages [83].

Because DNA is a strongly charged molecule in aqueous solution, the process of ejection of DNA from bacteriophages can be strongly influenced by the screening condition of the solution. By varying the salinity of solution, one can vary the amount of DNA ejected. Interestingly, monovalent counterions such as $Na^{+1}$ have negligible effect on the DNA ejection process [43]. In contrast, multivalent counterions such as $Mg^{+2}$, $CoHex^{+3}$, $Spd^{+3}$ or $Spm^{+4}$ exert strong and non-monotonic effects [42]. There is an optimal counterion concentration, $c_{Z,0}$, where the least DNA genome is ejected from the phages. For counterion concentration, $c_Z$, higher or lower than this optimal concentration, more DNA is ejected from phages. The case of divalent counterions is more marginal. The non-monotonicity is observed for $MgSO_4$ salt but not for $MgCl_2$ salt up to the concentration of 100 mM. The hydration effects have been proposed to explain for such dependence on the type of divalent salts [42].

The non-monotonic influence of multivalent counterions on DNA ejection from viruses is expected to have the same physical origin as the phenomenon of reentrant DNA condensation in free solution [113, 133, 117]. Due to strong electrostatic interactions, multivalent counterions condense on the surface of DNA molecules. As a result the apparent charge, $\eta^*$, of a DNA molecule is not its bare charge, $\eta_0 = -1e/1.7\mathring{\text{A}}$, but is the sum of $\eta_0$ and the condensed counterions' charges. At small $c_Z$, the large entropy of the counterions keeps them from condensing on DNA and $\eta^*$ is negative. However, at large $c_Z$ counterions over-condense on the DNA making $\eta^*$ positive. In both cases, Coulomb repulsion keeps DNA molecules apart. At an intermediate concentration, $c_Z \approx c_{Z,0}$, DNA molecules are almost neutralized by condensed counterions, $\eta^* \approx 0$. Short-range attractions between DNA molecules [111, 52] are able to condense them into bundles for these intermediate concentrations. This is the reentrant DNA condensation phenomenon. The counterion concentration at which DNA is neutral depends on mostly on the counterion valency and is shown to be 3.2 mM and 11 mM for tetravalent and trivalent counterions respectively [113, 124]. Although, divalent counterions are known to condense DNA only partially in free solution [74], DNA virus provides an unique experimental setup. The constraint of the viral capsid strongly eliminates configurational entropic cost of packaging DNA. This allows divalent counterions to influence DNA condensation similar to that of trivalent/tetravalent counterions. Indeed, DNA condensation by divalent counterions has also been observed in another environment where DNA configuration is constrained, namely the condensation of DNA in two dimensional systems [84]. For virus systems, theoretical fitting suggests that the DNA is neutralized at $c_{Z,0} \approx 75$ mM for divalent counterions, and the short-range DNA attraction at this concentration is $-0.004 k_B T$ per nucleotide base [88]. In this chapter, we study the problem of DNA condensation in the presence of divalent counterions using computer simulation. We calculate the electrostatic contribution to the free energy of packaging DNA into bundles. We show

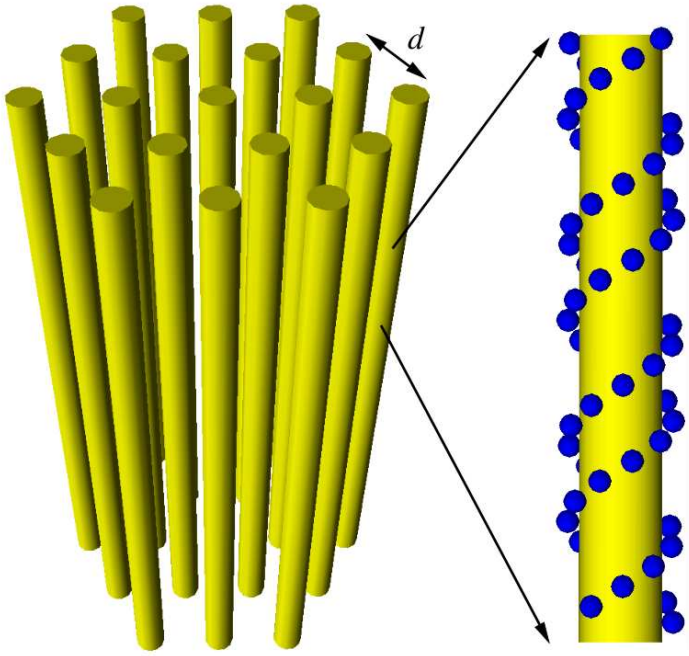


**Figure 5:** A DNA bundle is modelled as a hexagonal lattice with lattice constant $d$. Individual DNA molecule is modeled as a hard-core cylinder with negative charges glued on it according to the positions of nucleotides of a B-DNA structure.

that, if only the non-specific electrostatic contribution is included, divalent counterions can indeed induce DNA reentrant condensation like those observed for higher counterion valences. The simulation results agree reasonably well with the theoretical fit of viral ejection experiments.

## *3.2 Model*

We model the DNA bundle in hexagonal packing as a number of DNA molecules arranged in parallel along the $Z$-axis. In the horizontal plane, the DNA molecules form a two dimensional hexagonal lattice with lattice constant $d$ (the DNA-DNA interaxial distance) (Figure 5). Individual DNA molecule is modeled as an impenetrable cylinder with negative charges glued on it. The charges are positioned in accordance with the locations of nucleotide groups along the double-helix structure of a B-DNA. The hardcore cylinder has radius of 7Å. The negative charges are hard spheres of

radius 2Å, charge $-e$ and lie at a distance of 9Å from the DNA axis. This gives an averaged DNA diameter of 1 nm. The solvent water is treated as a dielectric medium with dielectric constant $\epsilon = 78$ and temperature $T = 300^o K$. The positions of DNA molecules are fixed in space. This mimics the constraint on DNA configurational entropy inside viruses and other experiments of DNA condensation using divalent counterions. The mobile ions in solution are modeled as hard spheres with unscreened Coulomb interaction (the primitive ion model). The coions have radius of 2Å and charge $-e$. The divalent counterions have radius of 2.5Å and charge $+2e$. The interaction between two ions $i$ and $j$ with radii $\sigma_{i,j}$ and charges $q_{i,j}$ is given by

$$U = \begin{cases} q_i q_j / \epsilon r_{ij} & \text{if} \quad r_{ij} > \sigma_i + \sigma_j \\ \infty & \text{if} \quad r_{ij} < \sigma_i + \sigma_j \end{cases} \tag{12}$$

where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ is the distance between the ions.

The simulation is carried out using the periodic boundary condition. Unless explicitly stated, a periodic simulation cell with $N_{DNA} = 12$ DNA molecules in the horizontal $(x, y)$ plane and 3 full helix periods in the $z$ direction is used. The dimensions of the box are $L_x = 3d$, $L_y = 2\sqrt{3}d$ and $L_z = 102$Å. The long-range electrostatic interactions between charges in neighboring cells are treated using the Ewald summation method. In Ref. [99, 59], it is shown that the macroscopic limit is reached when $N_{DNA} \geq 7$. Our simulation cell contains 12 DNA helices, hence it has enough DNA molecules to eliminate the finite size effect. We did a test run with 1, 4, 7 and 12 DNA molecules to verify that this is indeed the case. These smaller simulations are also used to check the correctness of our computer program by reproducing the results of DNA systems studied in Ref. [99, 59].

In practical situation, the DNA bundle is in equilibrium with a water solution containing free mobile ions at a given concentration. Therefore we simulate the system using Grand Canonical Monte-Carlo (GCMC) simulation. The number of ions are not constant during the simulation. Instead their chemical potentials are

fixed. The chemical potentials are chosen in advance by simulating a DNA-free salt solution and adjusting them so that the solution has the correct ion concentrations. In a simulation, the ions are inserted into or removed from the system in groups to maintain the charge neutrality [155]. We insert/remove one divalent counterion and two monovalent coions at the same time. Following Ref. [155], instead of using two independent chemical potentials, $\mu_{+2}$ and $\mu_{-1}$, for individual ion species we use only one chemical potential, $\mu_{+2\text{ salt}} = \mu_{+2} + 2\mu_{-1}$, in the Metropolis acceptance criteria of a particle insertion/deletion move.

In this chapter, we simulate DNA bundles at varying concentration $c_Z$. The limit of small $c_Z$ poses a nontrivial challenge. If we simulate DNA bundle in the presence of only divalent salt, then even at $c_z = 0$, there would be non-zero amount of divalent counterions inside the bundle due to the charge neutrality requirement. This situation is clearly non-physical. In reality, there always a finite amount of monovalent counterions from the buffer solution or from the deprotonization of DNA bases. According to the mass action law, when $c_Z$ is smaller than a certain value, the monovalent counterions will replace the divalent ones in neutralizing DNA charges. Therefore, in this chapter, to properly simulate the DNA bundle at small $c_Z$ limit, we include *both* divalent and monovalent slats in the simulation. The standard GCMC simulation method is generalized to this system by allowing insertion/removal of both kinds of salts in a simulation run. For simplicity, they both assumed to have the same coions. In addition to the chemical potential of a divalent salt molecule, $\mu_{+2\text{ salt}}$, we also use the chemical potential of a monovalent salt molecule, $\mu_{+1\text{ salt}} = \mu_{+1} + \mu_{-1}$, in the Metropolis criteria. Both $\mu_{+1\text{ salt}}$ and $\mu_{+2\text{ salt}}$ are adjusted so that the monovalent salt bulk concentration, $c_1$, in the DNA-free solution is always at 50 mM and $c_Z$ is at the desired value. Typical standard deviations in the final salt concentrations are about 10%.

## 3.3 Results and Discussion

In general, it is not trivial to calculate the free energy of assembling a DNA bundle in a Monte-Carlo simulation because the entropy cannot be calculated explicitly. To overcome this problem, we employed the Expanded Ensemble method [99, 59]. This scheme allows us to calculate the difference of the system free energy at different volumes by sampling these volumes in a simulation run. By sampling two nearly equal volumes, $V$ and $V + \Delta V$, and calculate the free energy difference $\Delta\Omega$, we can calculate the total pressure of the system:

$$P(T, V, \{\mu_\nu\}) = -\left.\frac{\partial\Omega(T, V, \{\mu_\nu\})}{\partial V}\right|_{T,\{\mu_\nu\}} \simeq -\frac{\Delta\Omega}{\Delta V}. \tag{13}$$

Here $\{\mu_\nu\}$ are the set of chemical potentials of different ion species. The osmotic pressure of the DNA bundle is then obtained by subtracting the total pressure of the bulk DNA-free solution, $P_b(T, V, \{\mu_\nu\})$, from the total pressure of the DNA system:

$$P_{\rm osm}(T, V, \{\mu_\nu\}) = P(T, V, \{\mu_\nu\}) - P_b(T, V, \{\mu_\nu\}) \tag{14}$$

In Figure 6, the osmotic pressure of DNA bundle at different $c_Z$ is plotted as a function of the interaxial DNA distance, $d$. Because this osmotic pressure is directly related to the "effective" force between DNA molecules at that interaxial distance [99, 59], Figure 6 also serves as a plot of DNA-DNA interaction. As one can see, when $c_Z$ is greater than a value around 20 mM, there is a short-range attraction between two DNA molecules as they approach each other. This is the well-known phenomenon of like-charge attraction between macroions [111, 52]. It is the result of the electrostatic correlations between counterions condensed on the surface of each DNA molecule. The attraction appears when the distance between these surfaces is of the order of the lateral separation between counterions (about 14Å for divalent counterions). The maximal attraction occurs at the distance $d \simeq 28$Å, in good agreement with various theoretical and experimental results [124, 122]. For smaller $d$, the

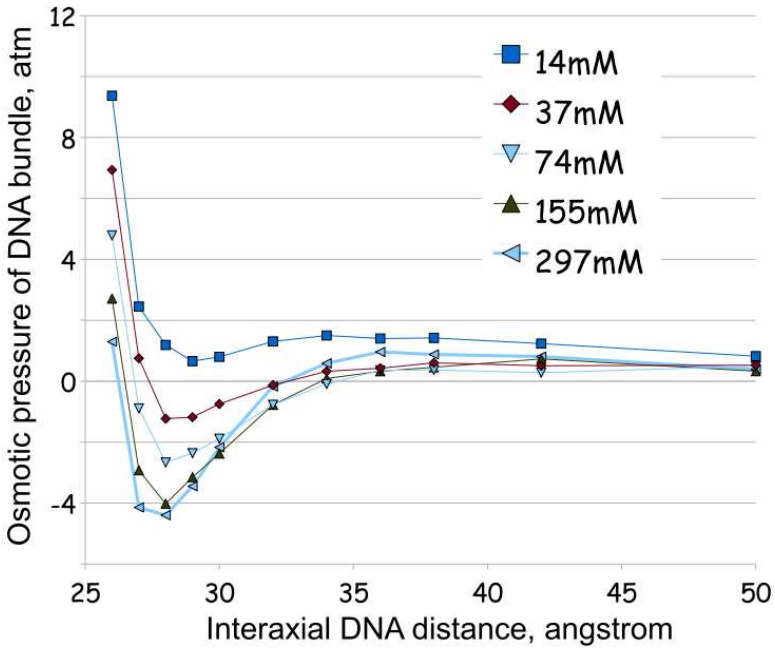


**Figure 6:** The osmotic pressure of the DNA bundle as a function of the interaxial DNA distance $d$ for different divalent counterion concentration $c_Z$ shown in the inset. The solid lines are guides to the eye.

DNA-DNA interaction experiences sharp increase due to the hardcore repulsion between the counterions. One also see that the depth of attractive force between DNA molecules saturates at around -4 atm as $c_Z$ increases. This saturation is easily understood. At small $c_Z$, there are both monovalent and divalent counterions present in the bundle. As $c_Z$ increases, divalent counterions replace monovalent ones in the bundle as the later ions are released into the bulk solution to increase the overall entropy of the solution. However, charge neutrality condition of the DNA macroscopic bundle and the hardcore repulsion between ions limit how many divalent counterions can be present inside the bundle. Once all monovalent counterions are released into solution (replaced by divalent counterions), further increase in $c_Z$ does not significantly change the number of divalent counterions in the bundle. This leads to the observed saturation of DNA-DNA short-range attraction with increasing $c_Z$.

It is also very illustrative to look at the DNA-DNA "effective" interaction at

larger $d$. At these separations, the distribution of counterions in the bundle can be considered to be composed of two populations: condensed layers of counterions near the surfaces of the DNA molecules and diffused layers of counterions further away. Of course, there is no definite distance that separates condensed from diffused counterions. Nevertheless, it is reasonable to expect the thickness of the condensed counterion layer to be of the order of the average lateral distance between counterions on the DNA surface. So for $d > 35\mathring{\text{A}}$, both counterion populations are present and one expects DNA-DNA interaction to be the standard screened Coulomb interaction between two charged cylinders with charge density $\eta^*$. As evident from Figure 6, at small $c_Z$, such as for $c_z$ =14 mM, DNA-DNA interaction is repulsive. As $c_Z$ increases, DNA-DNA interaction becomes less repulsive and reach a minimum around 75 mM. As $c_Z$ increases further, DNA-DNA repulsion starts to increase again. This is the same behavior as that of the phenomenon of reentrant DNA condensation by multivalent counterions [113, 136, 57].

The non-monotonic dependence of DNA-DNA "effective" interaction on the counterion concentration is even more clear if one calculates the free energy, $\mu_{DNA}$, of packaging DNA into bundles. This free energy is nothing but the difference between the free energy of a DNA molecule in a bundle and that of an individual DNA molecule in the bulk solution ($d = \infty$). It can be calculated by integrating the pressure with the volume of the bundle. Per DNA nucleotide base, the packaging free energy is given by:

$$\mu_{DNA}(d) = \frac{l}{L_z N_{DNA}} \int_{\infty}^{d} P_{\text{osm}}(d') dV = \frac{l}{N_{DNA}} \int_{\infty}^{d} P_{\text{osm}}(d') \frac{2L_x L_y}{d'} dd' \qquad (15)$$

here $l = 1.7\mathring{\text{A}}$ is the distance between DNA nucleotides along the axis of the DNA. The numerical result for $\mu_{DNA}(d)$ at the optimal bundle lattice constant $d = 28\mathring{\text{A}}$ is plotted in Figure 7 as a function of the $c_Z$. Due to the limitation of computer simulations, the numerical integration is performed from $d = 28\mathring{\text{A}}$ to $d = 50\mathring{\text{A}}$ only. However, this will not change the conclusion of this chapter because the omitted

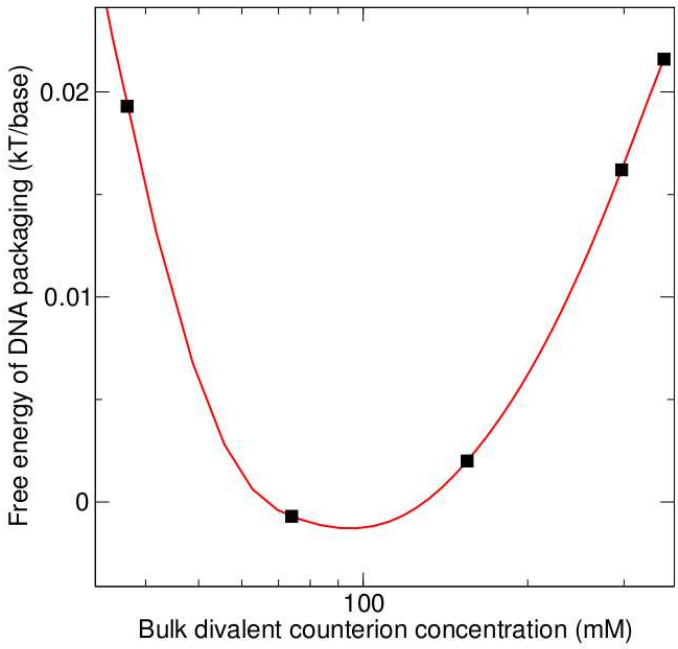


**Figure 7:** The free energy of packaging DNA molecules into hexagonal bundles as a function of the divalent counterion concentration. The points are results of numerical integration of $P_{\rm osm}$ from Figure 6. The solid line is a simple cubic spline interpolation.

integration from $d = 50\mathring{\rm A}$ to $d = \infty$ only gives an almost constant shift to $\mu_{DNA}$.

Once again, the non-monotonic dependence of the electrostatic contribution to DNA packaging free energy is clearly shown. There is an optimal concentration, $c_{Z,0}$, where the free energy cost of packaging DNA is lowest. It is even negative indicating the tendency of the divalent counterions to condense the DNA. At smaller or larger concentrations of the counterions, the free energy cost of DNA packaging is higher and positive. These results are consistent with the correlation theory of DNA reentrant condensation by multivalent counterions [113, 136, 57]. For small $c_Z$, DNA molecules are undercharged ($\eta^* < 0$). For large $c_Z$, DNA molecules are overcharged ($\eta^* > 0$). To condense the DNA molecules, one has to overcome the Coulomb repulsion between them. Therefore, the free energy cost of packaging is positive. For $c_Z \approx c_{Z,0}$, the DNA molecules are almost neutral, $\eta^* \approx 0$. The Coulomb repulsion is negligible and the free energy cost of condensing DNA molecules is lowest. Furthermore, the

like-charge attraction among DNA molecules mediated by the counterions [111, 52] is dominant in this concentration range, causing the electrostatic packaging free energy to become negative. Figure 7 gives $c_{Z,0} \approx 90$ mM and the short-range attraction among DNA molecules at this concentration is shown to be $-0.001 k_B T$/base. These values are slightly off from previous theoretical fit of viral DNA ejection experiments [88]. However, we believe these small quantitative differences are due to our choice of the system physical parameters such as ion size [100]. Nevertheless, the non-monotonic electrostatic influence of divalent counterions on DNA-DNA "effective" interaction is clearly demonstrated in our simulation.

# CHAPTER IV

# DIVERSITY OF *IN VIVO* ASSEMBLED HIV-1 CAPSIDS

## *4.1 Introduction*

Assembly and maturation of human immunodeficiency viruses (HIV) are complex and essential processes in the life cycle of this AIDS-causing virus. HIV belongs to the retroviral family of viruses, which all have similar structural components, assembly and infection pathway. Initially, retroviruses assemble as noninfectious particles that contain a spherical capsid composed of the viral structure poly-protein Gag. Upon budding, the capsid is enclosed in a membrane derived from the cell plasma membrane. This immature virion then undergoes a dramatic rearrangement to form the infectious capsid viral particles. This maturation is triggered by the cleavage of Gag into three new proteins, namely, matrix (MA), capsid (CA) and nucleocapsid (NC) proteins. The maturation of HIV is shown in Figure 8. As the viral particle matures, MA remains associated with the plasma membrane, NC associates with the viral RNA genomes, and CA self-assembles into a closed capsid shell that surrounds the RNA-NC complex [27].

Most traditional viruses have either spherical or cylindrical shapes. They are also very robust, highly reproducible in *in vitro* experiments. In contrast, retroviral capsids in general, and HIV capsid in particular, are highly heterogeneous in both shapes and sizes. Although the immature capsids of retroviruses are spherical, the shapes of the mature capsids vary significantly between different viruses, and between different virions of the same virus. TEM images of HIV-1 capsid show that, besides spherical and cylindrical, these capsids can have conical shapes as well (see Figure 9) [15]. This shape has triggered immense interest in the biology community as evident

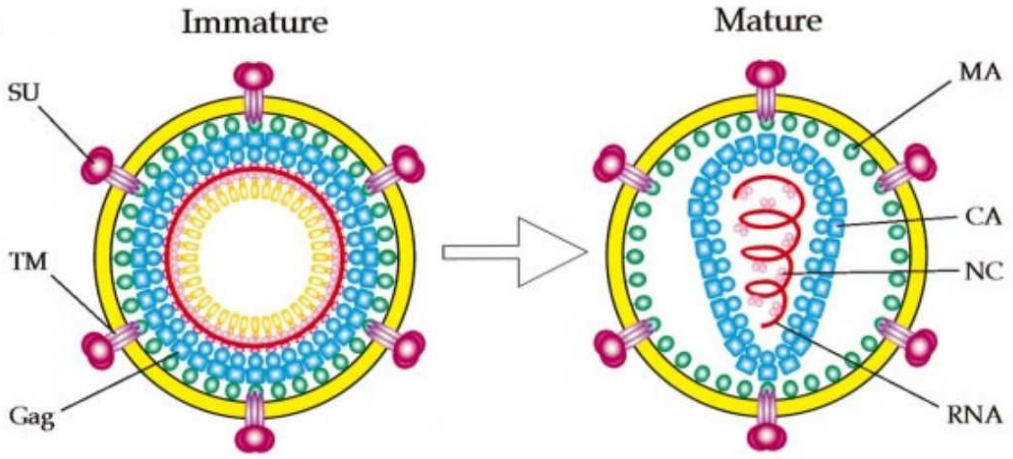


**Figure 8:** Schematic pictures of immature and mature HIV-1 [135]. Gag proteins are rearranged into three major proteins (MA, CA, and NC) during maturation process. Capsid requires a conformational change during maturation process. SU and TM are surface unit and trans membrane protein respectively.

by the growing number of scientific research to understand how the capsid proteins are arranged on such shape, and what physical factors favor the assembly of conical shape over the spherical or cylindrical ones. In the physics community, this conical shape of virus starts to receive broader attention recently. Several alternative explanations for the observed conical shape of the HIV virus were suggested. Henley proposed a kinetical approach to the formation of conical shape [66], where the conical shape is one consequence of an irreversible growth mechanism. The model, however, does not explain how conical shapes are easily obtained in *in vivo* than *in vitro* assembly. Nguyen et al studied a thermodynamic model of the assembly [114, 115]. Using a simple continuum shell approximation to the capsid to calculate the energies of spherical, cylindrical and conical shape they showed that in free assembly experiments (such as *in vitro* condition of empty capsid assembly), the most stable shape is either spherical or cylindrical depending on the spontaneous curvature of the shell. Only when the capsid volume is fixed, the conical shape becomes more stable. This result agreed qualitatively with experimental data showing the increase in conical shape

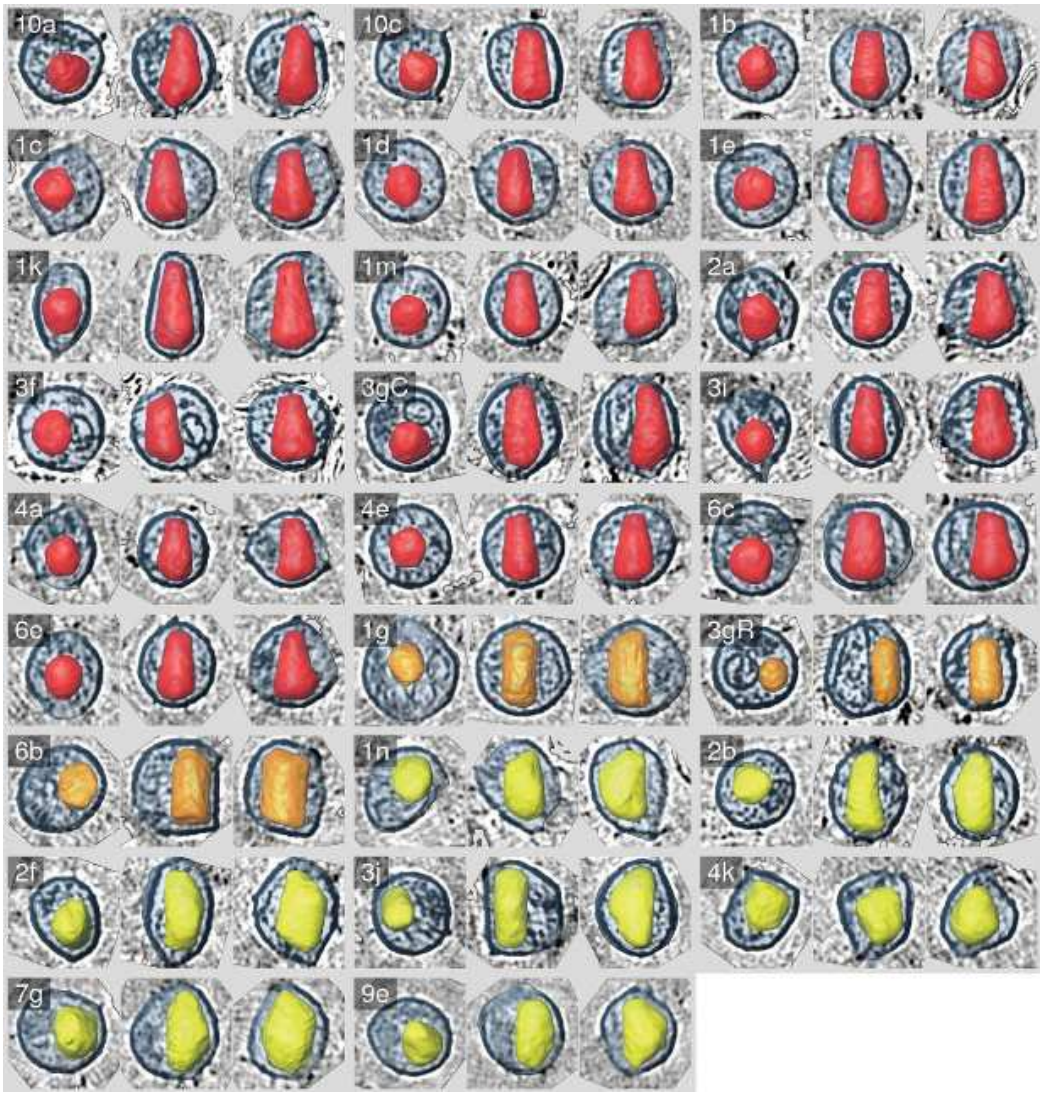


**Figure 9:** CryoTEM tomography images showing the diversity of HIV-1 capsids. Three views are shown along three orthogonal directions. Conical capsids are shown in red, Cylindrical capsids in orange, and other irregular shape capsids in yellow. Each capsid is enclosed by the lipid membrane in blue as the constraint of elongation of capsid during maturation process. These various shapes are found from a single cell culture. The size of each box is $160 \times 160\ nm^2$ [15].

when HIV capsid is assembled *in vitro* in the presence of RNA molecules [51]. Without RNA, the probability of CA proteins assembling into conical shape is very low (5%). However, in the presence of finite RNA molecules which are packaged inside capsid (hence provide a form of volume constraint on the assembly), the conical shapes account for about 20-30% of the total capsid formed.

In this chapter, we extend the continuum shell model to study the *in vivo* assembly of HIV-1 capsid. For this assembly process, HIV-1 capsid is formed inside a viral plasma membrane. This presents an additional constraint on the process which prohibits the formation of highly elongated cylindrical capsid. We investigate the possibility that, similar to the volume constraint in previous study, this membrane constraint can also increase the probability that assembled capsid has conical shape. We found that this is indeed the case, however only marginally. The physical picture is much more interesting. Within numerical errors, it is observed that the energies of optimal conical shape and optimal cylindrical shape are the same or within the thermal energy $k_BT$ from each other. In other words, thermodynamically both conical and cylindrical shapes have similar probability to appear as a result of the self-assembly process during maturation. We propose that this is why *in vivo* assembled HIV-1 capsids are so diverse in shapes and sizes even though *in vitro* assembled HIV-1 capsids are dominantly cylindrical.

## *4.2 Extended isometric construction of viral capsids*

A simple continuum elastic shell model to study energy of various viral capsid shapes (spherical, cylindrical and conical) inside the constraint of the viral membrane is used. Both analytical and computational calculations are used. For both calculations, an important starting point is the geometrical construction of the non-spherical capsids using the generalized Caspar-Klug (CK) construction [114, 115]. By using this construction, it is implicitly assumed that CA proteins are arranged in two-dimensional

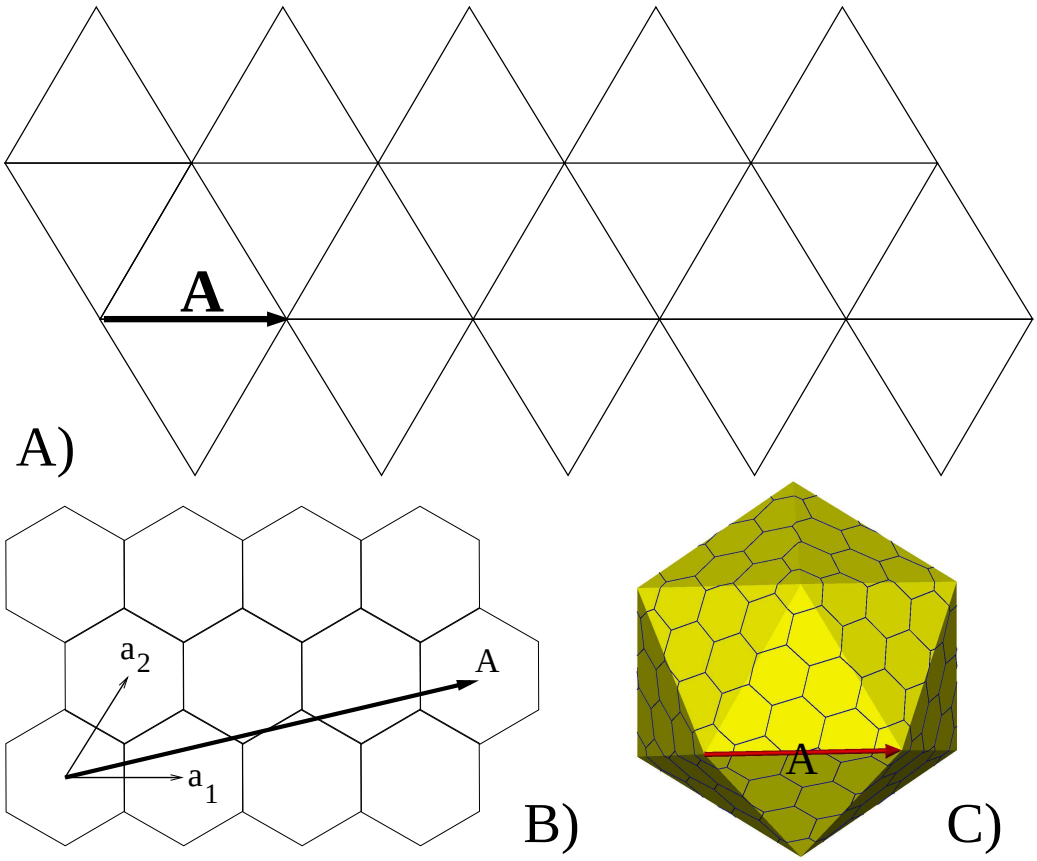


**Figure 10:** Caspar Klug construction of icosahedral shell. (a) Folding template for an icosahedral shell consisting of 20 equilateral triangles. The template is specified by the lattice vector $\vec{A} = h\vec{a}_1 + k\vec{a}_2$ of a hexagonal lattice with basis vectors $\vec{a}_1$ and $\vec{a}_2$. (b) An example with $\vec{A} = 3\vec{a}_1 + k\vec{a}_2$. (c) The icosahedron folded from the folding template for this lattice vector, which relates to $T = h^2 + k^2 + hk = 13$. Pentagons are located on 12 vertices and 6 hexagons occupy on each 20 equilateral triangles. The number of proteins is 60T = 780 [115].

hexagonal lattice on the mature capsid. Even though detail organization of CA proteins on mature capsid is not yet available, many lower resolution structural studies of HIV-1 capsid [51, 120] indirectly support this assumption. Since this geometrical construction is important to proper calculation of capsid energy, let us give its detail procedure below, starting from the standard CK construction.

### 4.2.1 CK construction for spherical capsid

Caspar and Klug deployed their famous geometrical construction of capsid subunits as a way of studying icosahedral capsid. In this construction, the icosahedron is created by folding corresponding inextensible template cut from a triangular lattice (Figure 10). Each so-called quasi-equivalent icosahedron is specified by the triangular lattice vector $\vec{A}$ that connects the neighboring vertices of the folding template (Figure 10A). Like any lattice vectors, $\vec{A}$ in turn is specified by a pairs of two integer numbers $(h, k)$ such that $\vec{A} = h\vec{a_1} + k\vec{a_2}$, where $\vec{a}_{1,2}$ are the two basis vectors of the triangular lattice (Figure 10B). One can see that the capsid made of the two basis units is composed of the 12 pentamers (made of five CA proteins) at the vertices of the icosahedron, and the hexamers (made of six CA proteins) that cover the faces of the icosahedron. Simple calculation shows that the number of capsomers is $10T + 2$ where $T = h^2 + k^2 + hk$ is the triangulation number (or T-number) of the quasi-equivalent icosahedral shape. The total number of proteins that make up the icosahedral capsid is $60T$.

### 4.2.2 Generalized CK construction for non-spherical capsids

The CK construction of icosahedral capsids has been extended to non-icosahedral shapes (cylindrical and conical) by Toan T. Nguyen [114, 115]. The corresponding templates are shown in Figure 11 and Figure 12 respectively.

Cylindrical capsid: Unlike icosahedral template, template for quasi-equivalent non-spherical capsid requires two lattice vectors $\vec{A}$ and $\vec{B}$ in order to be uniquely specified. In the case of cylindrical template (Figure 11A), these vectors can be specified on the lattice by two pairs of integer numbers $(m, n|h, k)$ such that

$$\vec{A} = n(h\vec{a}_1 + k\vec{a}_2),$$
$$\vec{B} = m(h\vec{b}_1 + k\vec{b}_2),$$

where the basis vectors $\vec{a}_{1,2}$ and $\vec{b}_{1,2}$ are shown in Figure 11B. In general, the ratio $m/n$ can be regarded as the asphericity of the cylindrical capsid (the ratio between

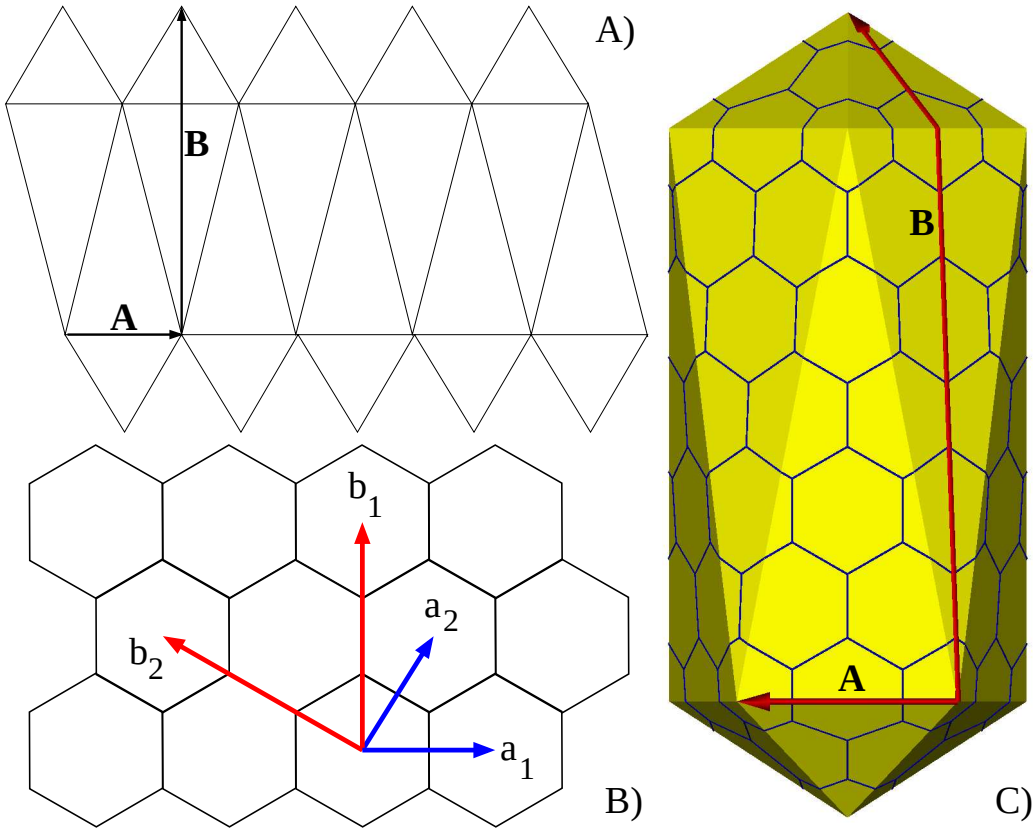


**Figure 11:** Isometric construction of a spherocylindrical shell (the spherocylinder reduces to an icosahedron for $m = n$). (a) Folding template for an isometric cylinder. (b) The folding template is specified by the two orthogonal basis lattice vectors $\vec{A} = n(h\vec{a}_1 + k\vec{a}_2)$ and $\vec{B} = m(h\vec{b}_1 + k\vec{b}_2)$, with $m > n$ two positive integers and $(h, k)$ two non-negative integers. (c) shows the isometric spherocylinder with $(m, n, h, k) = (4, 2, 1, 0)$ [115].

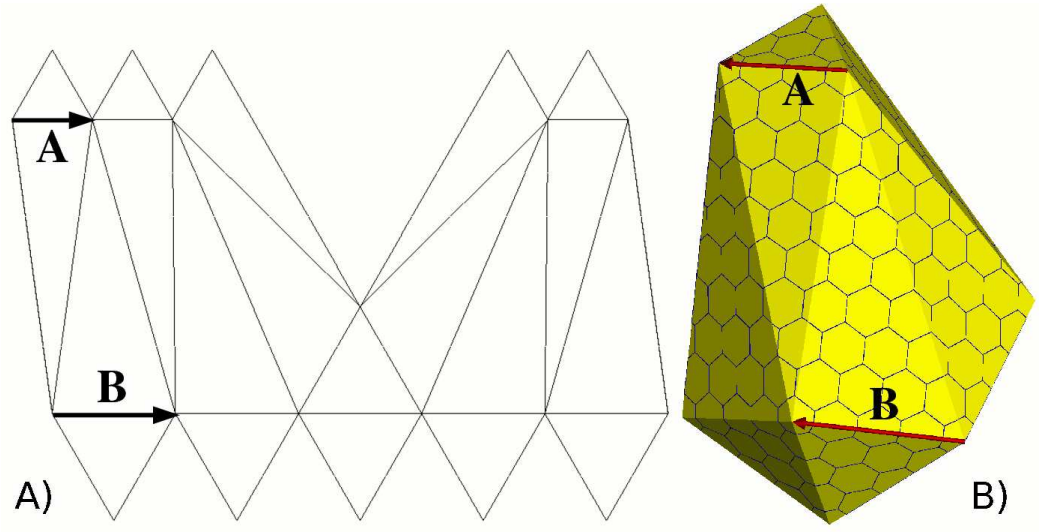


**Figure 12:** Isometric construction of a 7-5 cone (for $m = n$, the isometric 7-5 cone again drops to an icosahedron). (a) Folding template for an isometric 7-5 cone. The template is specified by two parallel vectors $\vec{A} = n(h\vec{a}_1 + k\vec{a}_2)$ and $\vec{B} = m(h\vec{b}_1 + k\vec{b}_2)$. (b) The case $(m, n, h, k) = (4, 3, 1, 0)$ [115].

the height and width of the cylinder). It is easy to verify that when this ratio is 1, the standard CK construction is recovered. One could also define a T-number $T = mn(h^2 + k^2 + hk)$ such that the number of capsomers is $10T + 2$ and the number of proteins is $60T$, similar to the spherical icosahedral capsid.

Conical capsid: In the case of conical capsid template, the two vectors are shown in Figure 12A. They can also be specified by two pairs of integer numbers $(m, n|h, k)$ such that (see Figure 12B):

$$\vec{A} = n(h\vec{a}_1 + k\vec{a}_2),$$

$$\vec{B} = m(h\vec{a}_1 + k\vec{a}_2).$$

The conical shape is asymmetric with two non-equal caps at the two ends. The ratio $m/n$ can be regarded as the ratio between the radii of these caps. When m/n=1, once again, the standard CK construction is recovered. One can also define the T-number of conical capsid as $T = (2m^2 - n^2)(h^2 + k^2 + hk)$. Like other cases, the number of capsomers is $10T + 2$ and the number of proteins is $60T$.

## *4.3 Evaluation of capsid energy*

### 4.3.1 Capsid energy from elastic theory of viral capsids

Experiments show that the number of CA proteins in a retroviral capsid is quite large (HIV mature capsid is made of about 1300 proteins [15] and the dimension of the capsid is much larger than the size of individual proteins. Because of this, we approximate the capsid as a continuum elastic shell. The energy of such shell is the sum of two contributions, an "in plane" stretching energy, $H_s$, and an "out-of-plane" bending energy $H_b$. The stretching energy is given by

$$H_s = \frac{1}{2}\int dS \left(2\mu u_{ij}^2 + \lambda u_{ii}^2\right), \tag{16}$$

where $u_{ij}$ is the strain tensor for the in-plane displacement of the shell and $\mu$ and $\lambda$ are two material constants, known as Lame coefficients which are related to the area modulus by $B = \lambda + 2\mu$, and to the (2D) Young modulus by $Y = 4\mu(\mu+\lambda)/(2\mu+\lambda)$. The integration is performed over the area S of the shell.

The bending energy $H_b$ of an elastic shell is given by

$$H_b = \frac{1}{2}\int dS \left[\kappa(H - C_0)^2 + 2\kappa_G K\right], \tag{17}$$

where $\kappa$ is the bending modulus of the shell, $H = 1/R_1 + 1/R_2$ is the mean curvature, and $C_0$ is the spontaneous curvature. To the lowest order, this spontaneous curvature is associated with the absence of "up-down" mirror symmetry of individual CA proteins. In the second term, $K = 1/R_1R_2$ is the Gaussian curvature and $\kappa_G$ is the Gaussian bending modulus. For a closed surface with fixed topology such as viral capsids, this Gaussian curvature is integrated out to a constant. We therefore drop this term from consideration. Thus, an elastic shell can be described by four phenomenological material parameters, the two Lame constants $\lambda$ and $\mu$, the bending modulus $\kappa$ and the spontaneous curvature $C_0$. Given these four parameters and a given capsid shape, one can calculate and optimize its energy as a function of the

aspherity $(m/n)$. The dominant shape in an assembly process is the shape with the lowest optimized energy.

The minimization of the total elastic energy, $H_s + H_b$, with respect to capsid shape leads to a set of coupled nonlinear equations, so-called Föpple-von Kármán (FvK) equations. This system of equations in general cannot be solved (thus, the capsid elastic energy cannot be calculated) exactly. In this chapter, we calculate the energy by two different methods, a computational one and a complementary analytical approximate one. The two methods are independent of each other for the most part. They are used to provide a consistency check of our results. The numerical method focuses only on the cylindrical and 7-5 conical shapes (where there are 7 pentamers on the larger cap and 5 pentamers on the smaller cap). The analytical method allows us to calculate energy of conical shapes with different cone angles. As we see in the results and discussion later, both methods lead to the conclusion that the membrane constraint makes both conical and cylindrical shapes similar in energy.

### 4.3.2 Numerical calculation of capsid elastic energy

The exact solution to the FvK continuum elastic equations and calculation of the elastic energy of various capsid shells can be obtained by numerical method as follows. The continuum shell is discretized the shell into a closed triangular net. The discretization follows exactly the isometric construction of the capsid described earlier. This allows us to preserve the underlying symmetry of the stress patterns of the actual protein shell, especially the presence of twelve pentamers on the capsid. Upon discretization, the "in plane" energy, Eq. (16), can be written as the sum of pair-wise interactions between neighboring nodes i and j of the net:

$$H_s = \frac{\epsilon}{2} \sum_{<i,j>} (|\vec{r}_i - \vec{r}_j| - a)^2. \tag{18}$$

Here a is the equilibrium spacing of the interaction potential, $|\vec{r}_i - \vec{r}_j|$ is the distance between two neighboring nodes (i and j), and $\epsilon$ is the strength of the interaction.

In the continuum limit where the number of discretization nodes becomes infinite, $\epsilon$ is proportional to the Young modulus of the shell, $\epsilon = \sqrt{3}Y/2$. The "out-of-plane" bending energy, Eq. (17), can be written as the sum of pair-wise interactions between neighboring triangles of the net:

$$H_b = k \sum_{<I,J>} [1 - \cos(\theta_{IJ} - \theta_0)] \tag{19}$$

Here $\theta_{IJ}$ is the dihedral angle between normal directions of two neighboring triangles I and J. In the continuum limit, the energy scale $k$ is proportional to the bending modulus, $k = 2\kappa/\sqrt{3}$, and the preferred angle $\theta_0$ is proportional to the spontaneous curvature of the continuum shell, $\theta_0 = \sqrt{3}C_0 a/2$. If number of nodes of our discretizing mesh is large enough, we could simulate the corresponding continuum (spherical or non-spherical) shell accurately. Previous study [114] shows that the continuum limit is reached when the number of nodes exceeds 10000. In this limit, for given material parameters and capsid shape, the capsid energy does not depend on specific value of the set $(m, n|h, k)$ but only on the asphericity $m/n$. Therefore, in this chapter, unless stated explicitly, we choose $h = 1$ and $k = 0$. The reference icosahedral capsid has $m = n = 55$ corresponding to 30252 nodes.

Starting from the isometric discretization, we then optimize the position of the nodes in space to minimize the total energy of the system using the conjugate gradient method. For given material parameters, $\epsilon, k$, given capsid area and membrane radius, the energy of a shell shape is calculated as a function of the asphericity m/n. The shape with the optimum $m/n$ whose energy is lowest is chosen. Comparing energy of spherical shell with the energies of the optimum cylindrical shell and the optimum conical shell, we can determine the most thermodynamically stable shape corresponding to this given set of parameters. Experimentally, these parameters depend on specific CA protein (or their mutations) interactions. Mapping available experimental results of assembled shapes, we can work out the strength of these interactions.

### 4.3.3 Analytical approximation to elastic energy

Since the numerical computation focuses only on 7-5 conical shapes, we use a complimentary analytical calculation which can address any conical shapes in general. This approximation was proposed by in Ref. [95] and generalized to non-spherical capsid and to include non-zero spontaneous curvature in Ref. [114]. It is based on the fact that there are exactly 12 pentamers on viral capsids. Each pentamer can be considered a five-fold disclination defect on the 2D hexagonal lattice of the CA proteins. For typical range of the capsid protein elastic parameters, the elastic stress is significant only near the center of these disclinations. Thus, one can approximate the energy of the spherical capsid as the sum of elastic energies of 12 five-fold disclinations centered at the pentamers. This gives the energy of a spherical capsid by [114]:

$$\frac{E_0(\gamma,\alpha)}{\kappa} = 6B\left[1+\ln\left(\frac{\gamma}{\gamma_B}\right)\right] + C(\gamma,\alpha) + D(\gamma,\alpha) \tag{20}$$

where the background bending energy $D(\gamma,\alpha)$ of the core region of the disclinations,

$$D(\gamma,\alpha) = 8\pi - 4\sqrt{\pi}\alpha\sqrt{\gamma_B/\gamma} + (\alpha^2/2)\gamma_B/\gamma \tag{21}$$

and the contribution, $C(\gamma,\alpha)$ , to the bending energy of the outer region of the disclination due to non-zero $C_0$ is

$$C(\gamma,\alpha) = 6B\left[-\frac{2\alpha}{\sqrt{\pi}}\left(\sqrt{F(\gamma)}-\sqrt{\frac{\gamma_B}{\gamma}}\right)+\frac{\alpha^2}{4\pi}\left(F(\gamma)-\frac{\gamma_B}{\gamma}\right)\right] \tag{22}$$

where $F(\gamma) = \frac{1-\gamma_B/\gamma(1-3\cos\theta_1/\tan\theta_1)}{3\cos\theta_1/\tan\theta_1}$ with $\theta_1 = \arcsin(\frac{5}{6}) \sim 56\,\text{deg}$ equal to the largest cone angle consistent with forming a truncated cone from a hexagonal lattice. The elastic energy depends on two dimensionless parameters $\gamma$ and $\alpha$. $\gamma = YS/\kappa$ ($S$ is the capsid area) is the FvK number, which gives the relative strength of stretching and bending energies of the capsid shell and $\alpha = C_0 S^{1/2}$ is spontaneous curvature parameter which is proportional to the preferred angle $\theta_0$. $\gamma_B$ and $B$ are two numerical numbers. Although the theoretical values for them are 1935 and $11\pi/30$ respectively

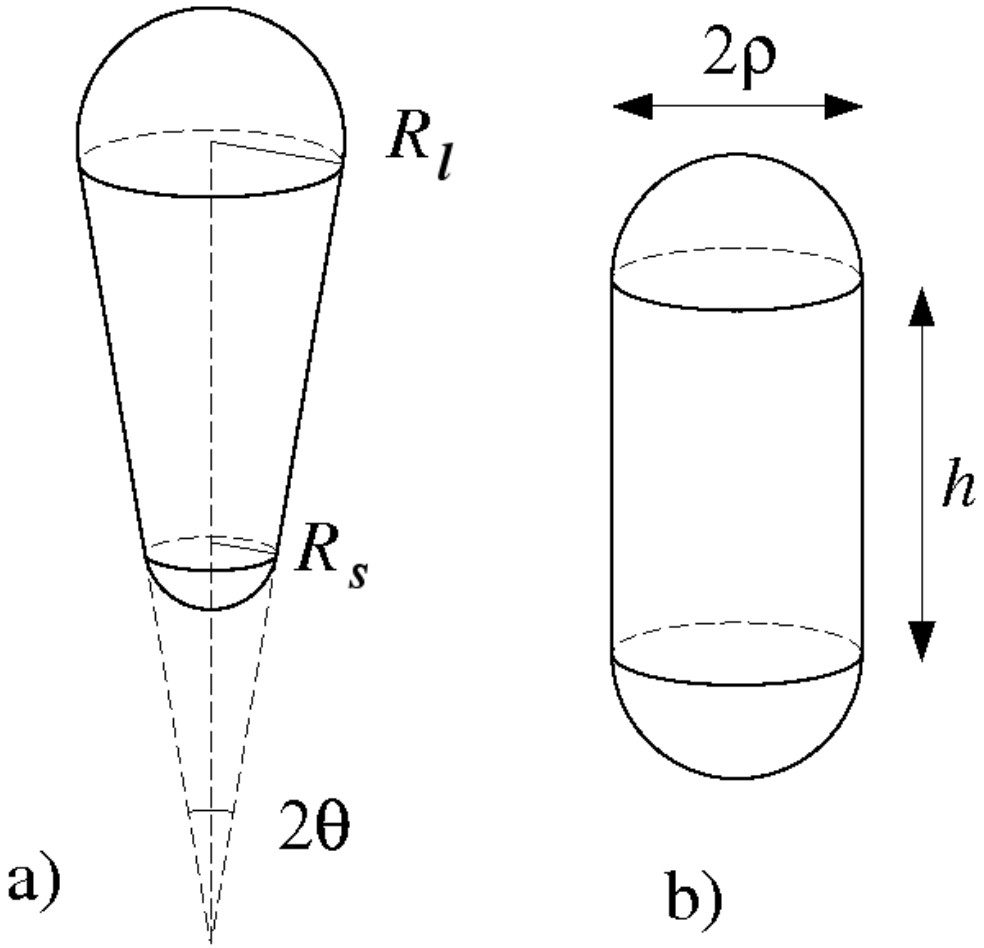


**Figure 13:** Approximate shape for non-spherical capsid. (a) Construction of a smooth shell by connecting a larger hemisphere of radius $R_l$ on the top to a smaller hemisphere of radius $R_s$ on the bottom by a cone that is cotangent to the two spheres. The cone aperture angle is $2\theta$. Curvature is discontinuous on the surface along the two matching circles. (b) Construction of a smooth spherocylindrical shell with height h and cylinder radius $\rho$ [114].

[95], they are treated as fitting parameters to numerical calculations. It should be noted that Eq. (20) is valid only for $\gamma > \gamma_B$. It can be shown from experimental estimates (as we will see in the more detail in the discussion section) that this is indeed the case for retroviruses.

Non-spherical capsids (cylindrical and conical) can be approximated by a closed surface consisting of two spherical caps connected by a smooth body (cylinder or cone) with aperture angle $2\theta$ (see Figure 13). The 12 disclinations are distributed on the caps. The smooth body is constructed from a hexagonal lattice of CA proteins, hence it contains no disclinations (see previous section on generalized CK construction). For cylindrical capsid, each cap has 6 disclinations. For conical capsid, there are two partially spherical caps at the two ends with radii $R_l$ and $R_s$ respectively. The larger cap has $12 - M$ disclinations and the smaller cap has $M$ disclinations where $M = 2, 3, 4$, or 5 (the special case $M = 1$ corresponds to spherical capsid and M=6 to cylindrical capsid). Simple geometrical calculation gives the cone angle $\theta_M$ as $\sin\theta_M = 1 - M/6$. The non-spherical capsid is now characterized by two FvK numbers corresponding to the radii of the caps, $\gamma_{l,s} = Y 4\pi R_{l,s}^2/\kappa$. The elastic energy of non-spherical capsid is approximated as the sum of three parts: the energy of two spherical caps, $E_0(\gamma, \alpha)/\kappa$, given by Eq. (20) but scaled by the appropriate number of disclinations, and the energy of the smooth connecting body:

$$\frac{E_M(\gamma_l, \gamma_s, \alpha)}{\kappa} = \frac{12 - M}{M}\frac{E_0(\gamma_l, \alpha)}{\kappa} + \frac{M}{12}\frac{E_0(\gamma_s, \alpha)}{\kappa} + D_M(\gamma_l, \gamma_s, \alpha) \tag{23}$$

The energy, $D_M(\gamma_l, \gamma_s, \alpha)$ , of the smooth body is given by:

$$D_M(\gamma_l, \gamma_s, \alpha) = \frac{\cos\theta_M}{2\tan\theta_M}\left[\pi \ln\frac{\gamma_l}{\gamma_s} - 2\sqrt{\pi}\alpha\left(\sqrt{\frac{\gamma_l}{\gamma}} - \sqrt{\frac{\gamma_s}{\gamma}}\right) + \frac{\alpha^2}{4}\left(\frac{\gamma_l}{\gamma} - \frac{\gamma_s}{\gamma}\right)\right] \tag{24}$$

The Eqs.(20) and (23) form the basis of our analytical approximation to calculate the energy of various capsids.

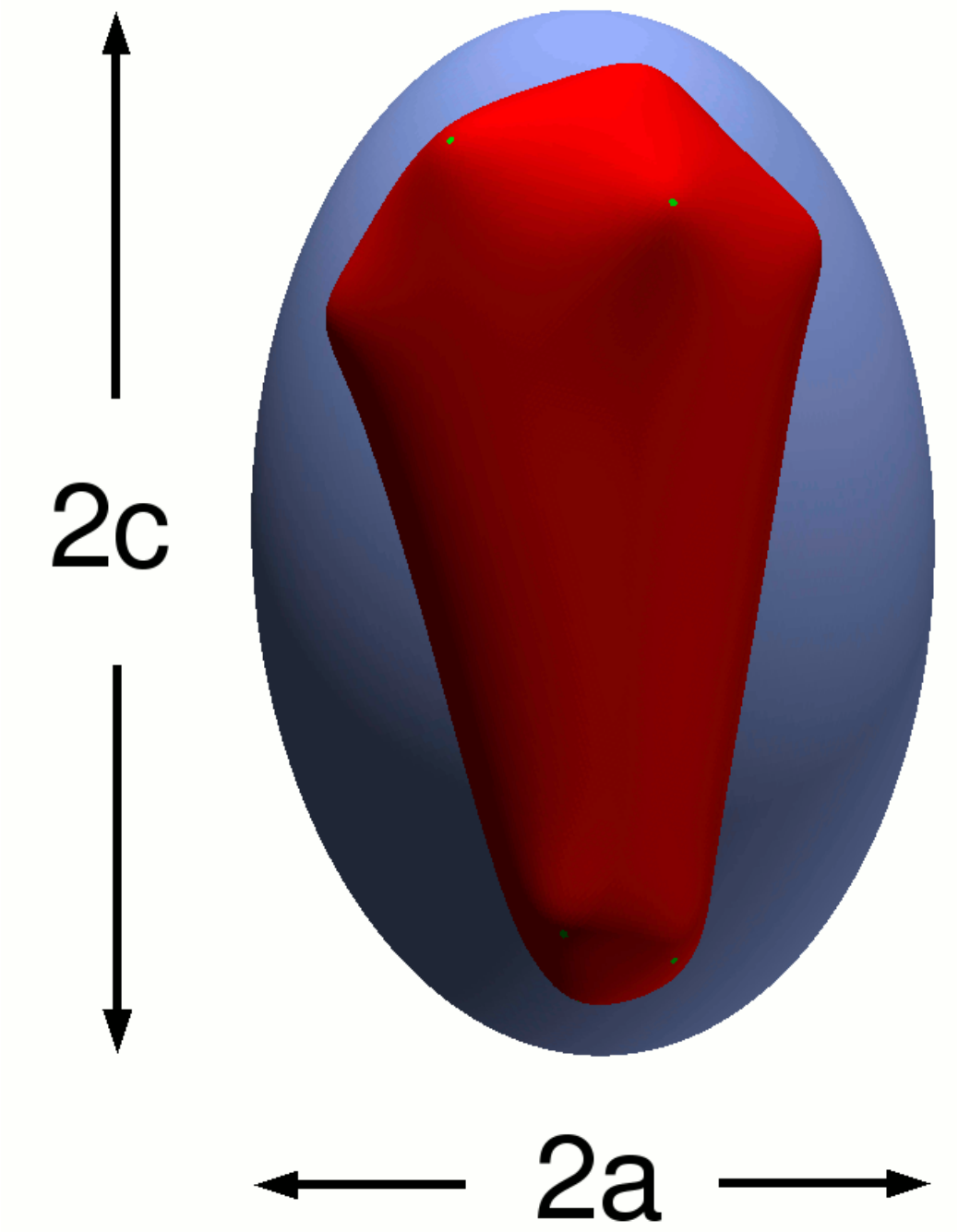


**Figure 14:** Visual image of a virus accommodating a conical capsid during capsid maturation. Immature HIV-1 changes its spherical membrane to an ellipsoidal membrane shape by pushing the inner membrane with the maturing capsid. The lengths of the long and short semi-axes of the membrane are $c$ and $a$.

## 4.4 Model for the membrane and its energy

To study the effect of the viral membrane on the assembly process of retroviral capsids, we model the membrane as a flexible ellipsoidal shell. As capsid matures from spherical to cylindrical or conical, the membrane is elongated along the long axis of the capsid (Figure 14). Denote $c$ as the polar radius (larger radius) , $a$ the equatorial radius (shorter radius) and $x = a/c$ the reciprocal of the aspect ratio of the ellipsoid. The membrane energy is described by its surface tension $\sigma$, osmotic pressure $P$ and bending rigidity $\kappa_m$ as:

$$H_m = -\sigma S(c,x) + PV(c,x) + H_{bend}. \tag{25}$$

Here $S(c,x)$ and $V(c,x)$ are surface area and enclosed volume of the membrane respectively:

$$\begin{aligned} S(c,x) &= 2\pi(cx)^2 + \left(1 + \frac{\arcsin\sqrt{1-x^2}}{x\sqrt{1-x^2}}\right), \\ V(c,x) &= \frac{4\pi}{3}c^3x^2. \end{aligned} \tag{26}$$

$H_{bend}$ (the bending energy of the membrane) can be obtained with $H$ (mean curvature) as

$$H_{bend} = 2\int dS\kappa_m H^2, \tag{27}$$

where

$$H = \frac{c^2}{16a^2}\frac{\left(3a^2 + c^2 + (a^2-c^2)\cos(2\theta)\right)^2}{\left(a^2\cos^2\theta + c^2\sin^2\theta\right)^3}. \tag{28}$$

To proceed Eq. (27) further, we need to calculate $\int dS$ using $z = c\cos\theta$:

$$\begin{aligned} \int dS &= 2\pi a\int_{-c}^{c}\sqrt{1 + \frac{(a-c)(a+c)z^2}{c^4}}dz \\ &= 2\pi ac\int_0^{\pi}\sqrt{1 + \frac{(a-c)(a+c)}{c^2}\cos^2\theta}\sin\theta d\theta. \end{aligned} \tag{29}$$

Inserting Eq. (28) and Eq. (29) into Eq. (27), we can obtain the following equation:

$$\begin{aligned} H_{bend} &= \frac{\kappa_m \pi c^3}{a} \int_0^{\pi} d\theta \sqrt{1 - \left(\frac{c^2 - a^2}{c^2}\right) \cos^2\theta} \sin\theta \\ &\times \left( \frac{1}{a^2 \cos^2\theta + c^2 \sin^2\theta} + \frac{2a^2}{(a^2 \cos^2\theta + c^2 \sin^2\theta)^2} + \frac{a^4}{(a^2 \cos^2\theta + c^2 \sin^2\theta)^3} \right) \end{aligned} \tag{30}$$

Integrating each term of the above equation, $H_{bend}$ can be finally written by

$$H_{bend} = 2\kappa_m \pi \left( \frac{7 + 2x^2}{3} + \frac{\arcsin\sqrt{1 - x^2}}{x\sqrt{1 - x^2}} \right). \tag{31}$$

This is the simplest physical model that mimics the constraint of viral membrane on capsid assembly. Note that the membrane surface tension and osmotic pressure are not independent, $P = \sigma/2R_m$. For a given elongation $c$, the optimal $x$ (and hence the membrane energy) is obtained by minimizing $H_m$.

## *4.5 Results*

### 4.5.1 Numerical computations of spherical, cylindrical and 7-5 conical capsids

It is known that in the absence of the viral membrane, the fraction of conical capsids in self-assembly experiments is very small. Numerical simulation confirms this observation [114, 115]. The authors showed that at small spontaneous curvature parameter, $\alpha$, the dominant shape is spherical. As $\alpha$ increases beyond some critical value, $\alpha_c$, the capsid shell undergoes a first order transition and the most dominant shape is cylindrical with $h/\rho \sim 2$. This critical $\alpha_c$ varies non-monotonously with $\gamma$ (decreases then increases as $\gamma$ increases, see Figure 15). As $\alpha$ increases further, the optimal cylindrical shape becomes more elongated and remains lower in energy than both conical and spherical ones. An important observation is that the conical shape is never a thermodynamically stable shape in this unconstrained assembly. Our numerical calculation reproduces their phase diagram (the most stable capsid shape as a function of the dimensionless parameters $\gamma$ and $\alpha$) by setting the membrane surface

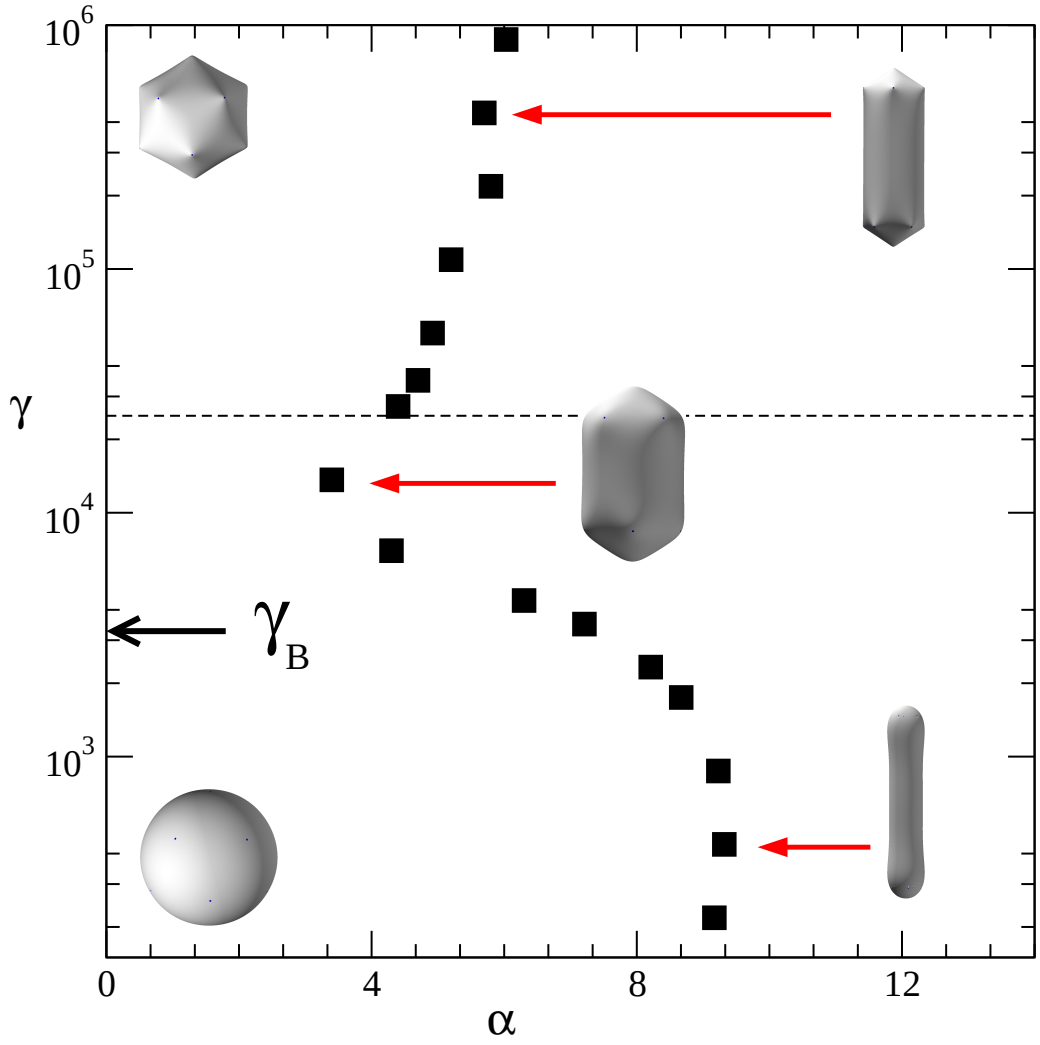


**Figure 15:** Shape phase diagram without membrane. The vertical axis is the FvK number $\gamma = YS/\kappa$; the horizontal axis is the dimensionless spontaneous curvature $\alpha = 2\theta_0 S^{1/2}/\sqrt{3}a$. For low $\alpha$, icosahedral shells are stable for all FvK numbers. The black squares are the shape phase boundary between spherical shells and the cylindrical shells obtained from numerical simulation. The cylindrical shape becomes more stable than the spherical shapes with high $\alpha$. The dotted line corresponds to the FvK number of HIV-1 capsids ($\gamma \sim 21000 > \gamma_B$) [114].

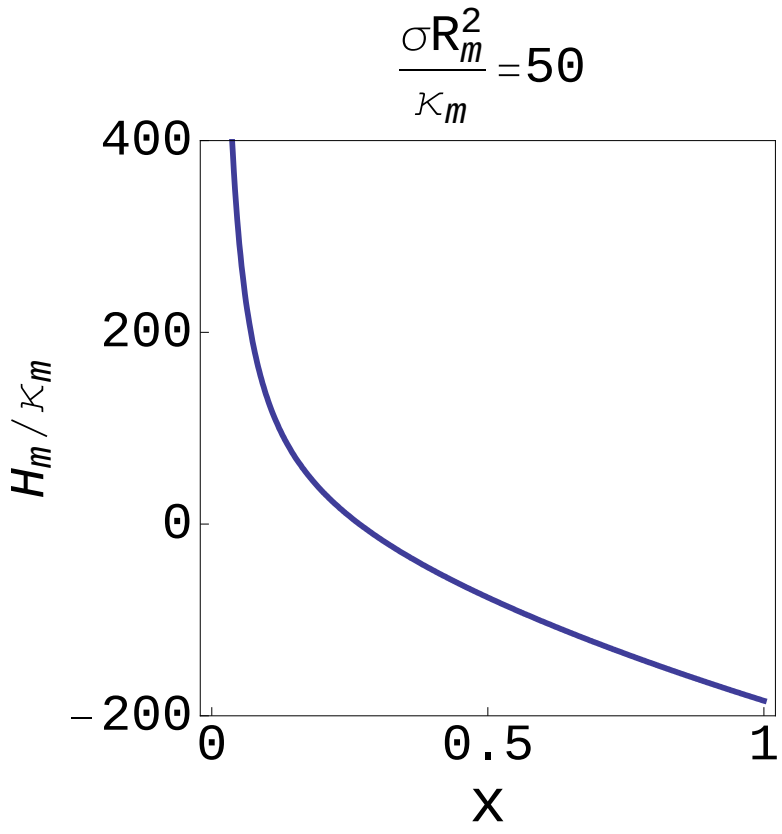


**Figure 16:** Dependence of the membrane energy $H_m$, in terms of the bending rigidity $\kappa_m$, on $x = a/c$, the reciprocal of the aspect ratio of the membrane for dimensionless membrane tension parameter $\left(\frac{\sigma R_m^2}{\kappa_m} = 50\right)$ which is in the range for HIV-1.

tension $\sigma$ to zero. The situation is very different when the capsid matures in the presence of the constraining viral envelop membrane. This membrane must elongate along the long axis of the capsid during the maturation process and is deformed into an ellipsoidal membrane. Thus additional energetic cost occurs when the capsid elongates. For the case of HIV-1 viral membrane, we can use typical experimental values: $\sigma \sim 1pN/nm$, $R_m \sim 60 - 80nm$, $\kappa_m \sim 20 - 40k_BT$. The range of $\sigma R_m^2/\kappa_m$ is from 21.95 to 78.05. In Figure 16, this energetic cost of deforming the viral membrane (normalized by the bending rigidity $\kappa_m$) is plotted as a function of $x$ (the membrane asphericity), for the dimensionless membrane tension parameter, $\sigma R_m^2/\kappa_m = 50$ which

is in the range of $\sigma R_m^2/\kappa_m$ for HIV-1. The membrane energy $H_m$ continuously decreases and is minimal at $x = 1$. Note that the decreasing behavior of $H_m$ with increasing $x$ does not depend on the strength of the dimensionless membrane surface tension parameter. Thus, HIV-1 membrane retains its spherical shape during the expansion to accommodate the capsid and after its maturation process. This simplifies our expression for the membrane energy significantly. The bending energy of a sphere is well-known to be a constant, $8\pi\kappa_m$. It does not depend on the size of a sphere. We can drop out the bending energy term for HIV-1 membrane from consideration. The only term in the membrane deformation cost which depends on capsid length $L_{max}$ is:

$$H_m = 4\pi\sigma(L_{max}/2 - R_m)^2. \tag{32}$$

### 4.5.2 The effect of the length constraint on the thermodynamics of retroviral mature shape is shown in the range for HIV-1

The capsid energy (normalized by the bending modulus $\kappa$) without the membrane ($\sigma R_m^2/\kappa = 0$) is plotted as function of $L_{max}$ for different values of $\alpha$ in Figure 17. The green line corresponds to the conical shape, and the blue line corresponds to the cylindrical shape. At small $\alpha$ (Figure 17A), the cylindrical shape is considerably lower in energy than the conical. As $\alpha$ increases, the energies of conical and cylindrical capsids become identical for $L_{max}$ close to $2R_m$ (Figure 17B). However, the optimal cylindrical shape (cylindrical shape with lowest energy) is much lower in energy than the optimal conical shape. This optimal cylindrical shape has m/n ratio $\sim 5$, and is highly elongated. In Figure 18 , the capsid energy in the presence of the membrane constraint with $\sigma R_m^2/\kappa = 63.81$ is plotted. The elongated shape becomes prohibitively high in energy and the optimal shape has $L_{max}$ close to $2R_m$. In this case, we can see that conical and cylindrical are very similar in energy. The optimal conical shape (green curve) is even smaller than the optimal cylindrical shape (blue curve). However, the difference is within numerical error.

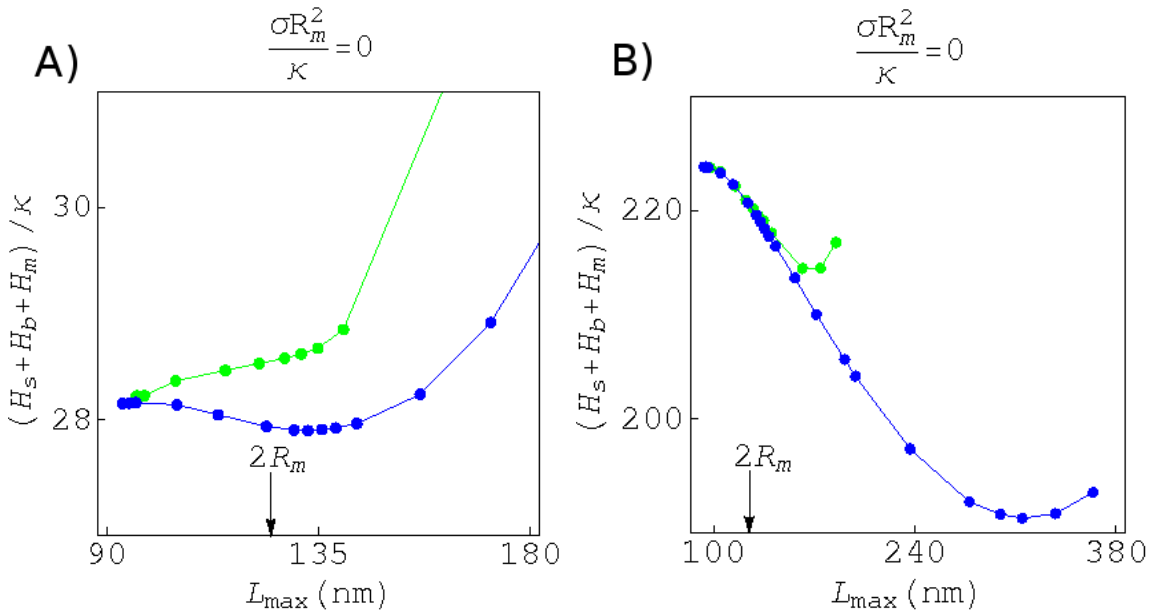


**Figure 17:** Dependence of the elastic energy, in terms of the bending constant $\kappa$, of the spherocylindrical capsid (blue line) and conical capsid (green line) on the length of the capsid at two different $\alpha$ without membrane from numerical calculation (NBG). (a) The dimensionless spontaneous curvature $\alpha$ is 5. The spherocylindrical shape starts to be more stable than spherical. (b) is drawn at $\alpha = 15$. Conical and spherocylindrical shape are identical in energy up to $L_{max} = 2R_m$.

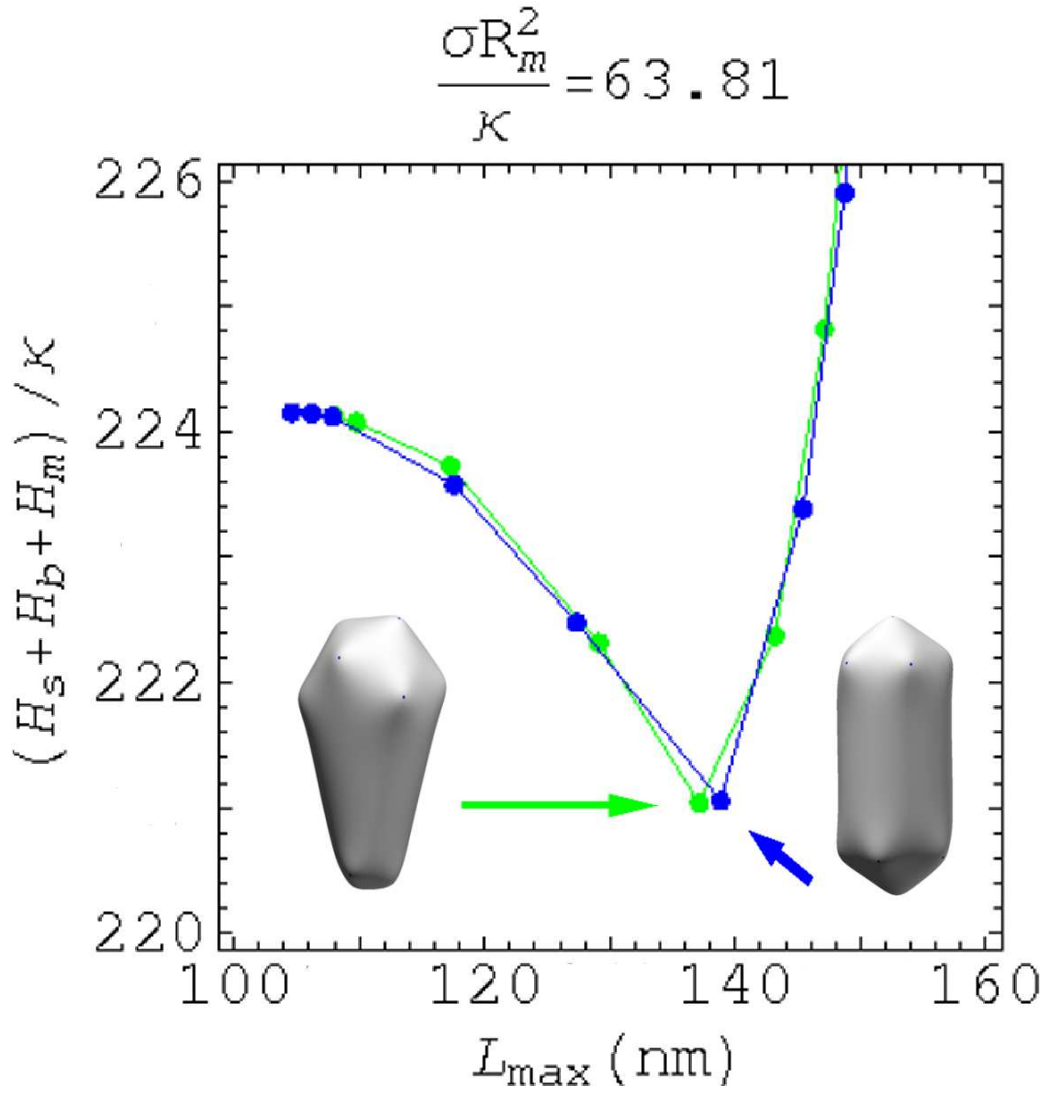


**Figure 18:** Dependence of the total energy of the system, in terms of the bending constant $\kappa$, the spherocylindrical capsid (blue line) and conical capsid (green line) on the length of the capsid for the case $\left(\frac{\sigma R_m^2}{\kappa_m} = 63.81\right)$. Two energies are identical in energy up to $2R_m$.

### 4.5.3 Analytical approximation for capsid energy

Numerical result shows that the presence of the viral membrane imposes certain restriction on the length of the capsid depending on the strength of the membrane surface tension. For HIV-1, available experimental data show that the diameter of the viral membrane ($2R_m$) and the cylindrical and conical capsid diameter ($L_{max}$) are very similar, $2R_m/L_{max} \sim 1$. The lowest energy capsids, therefore, will have their largest diameter, $L_{max}$, only slightly bigger than $2R_m$. Since these capsids are most relevant ones in an assembly process, it is natural to consider only them and approximate their energy by expanding $E_M(\gamma, \alpha, \epsilon)/\kappa$ in terms of the small parameter $\epsilon = (L_{max} - 2R)/2R$ ($R$ is spherical capsid radius) and keep only terms up to the second order. After some mathematical manipulation we get:

$$\frac{E_M(\gamma,\alpha,\epsilon)}{\kappa} \simeq \frac{E_M(\gamma,\alpha,0)}{\kappa} + \epsilon\frac{E_M^{(1)}(\gamma,\alpha,0)}{\kappa} + \epsilon^2\frac{E_M^{(2)}(\gamma,\alpha,0)}{\kappa} \tag{33}$$

The first term is simply the energy of the spherical capsid at the same capsid area given by Eq. (20) , thus is independent of $M$ (identical for cylindrical and conical capsids). It is found that the second term is also independent of $M$:

$$\frac{E_M^{(1)}(\gamma,\alpha,0)}{\kappa} = -12B + 4\pi + \alpha\left[-4\sqrt{\pi} + \frac{6B}{\sqrt{11\pi}}\frac{20}{\sqrt{10+\gamma_B/\gamma}}\right] + \alpha^2\left(1 - \frac{30B}{11\pi}\right). \tag{34}$$

However, the third term in the expansion is dependent on $M$, and is the main energy term responsible for the difference in energies between capsids:

$$\begin{aligned}\frac{E_M^{(2)}(\gamma,\alpha,0)}{\kappa} &= \left[\frac{-132(5M^2-60M+144)}{22(12-M)M}B + \frac{6(M^2-12M+48)}{(12-M)M}\pi\right] \\ &+ \alpha\left[2\sqrt{\pi} - \frac{6B}{\sqrt{11\pi}}\left(200 + \frac{10(M^2-12M+144)}{(12-M)M}\frac{\gamma_B}{\gamma}\right)\Big/\left(10+\frac{\gamma_B}{\gamma}\right)^{\frac{3}{2}}\right] \\ &+ \alpha^2\left[\frac{-(M^2-12M+144)}{2(12-M)M} + \frac{30(M^2-12M+144)}{22(12-M)M}\frac{B}{\pi}\right].\end{aligned} \tag{35}$$

Then, we can write Eq. (33) as

$$\begin{aligned}
\frac{E_M(\gamma,\alpha,\epsilon)}{\kappa} &\simeq \frac{E_M(\gamma,\alpha,0)}{\kappa} + \epsilon\frac{E_M^{(1)}(\gamma,\alpha,\epsilon)}{\kappa} + \epsilon^2\frac{E_M^{(2)}(\gamma,\alpha,\epsilon)}{\kappa} \\
&= 6B\left[1+\ln\left(\frac{\gamma}{\gamma_B}\right)\right] + 8\pi + \alpha\left[\frac{12B}{\sqrt{\pi}}\sqrt{\frac{\gamma_B}{\gamma}} - 4\sqrt{\pi}\sqrt{\frac{\gamma_B}{\gamma}}\right. \\
&\qquad \left. - 12B\sqrt{\frac{10}{11\pi}\left(1+\frac{\gamma_B}{10\gamma}\right)}\right] + \alpha^2\left[\frac{15}{11\pi}B + \frac{\gamma_B}{2\gamma} - \frac{15B}{11\pi}\frac{\gamma_B}{\gamma}\right] \\
&+ \epsilon\left\{-12B + 4\pi + \alpha\left[-4\sqrt{\pi} + \frac{6B}{\sqrt{11\pi}}\frac{20}{\sqrt{10+\gamma_B/\gamma}}\right] + \alpha^2\left(1-\frac{30B}{11\pi}\right)\right\} \\
&+ \epsilon^2\left\{\left[\frac{-132(5M^2-60M+144)}{22(12-M)M}B + \frac{6(M^2-12M+48)}{(12-M)M}\pi\right]\right. \\
&+ \alpha\left[2\sqrt{\pi} - 6B11\pi\left(200 + \frac{10(M^2-12M+144)}{(12-M)M}\frac{\gamma_B}{\gamma}\right)\Big/\left(10+\frac{\gamma_B}{\gamma}\right)^{\frac{3}{2}}\right] \\
&+ \left.\alpha^2\left[\frac{-(M^2-12M+144)}{2(12-M)M} + \frac{30(M^2-12M+144)}{22(12-M)M}\frac{B}{\pi}\right]\right\}. \qquad (36)
\end{aligned}$$

The detail derivation is shown in Appendix B.

## *4.6 Discussion*

Let us start this section by showing the expected range of elastic parameters for HIV-1 capsid. Many elastic fitting of various viruses [107] indicated that typical ratio between Young modulus and the bending rigidity of viral shell is about $Y/\kappa \sim 1\ nm^{-2}$ in general (with about 30 % variation between viruses). The total area of HIV-1 capsid is about 21000 $nm^2$ with standard deviation 9000 $nm^2$. Thus the range of the FvK number $\gamma$ for HIV-1 is estimated to be $12000 \sim 30000$. The spontaneous curvature of CA protein shell can be deduced from the fact that cylindrical capsid formed in experiment has diameter of about 40 $nm$, leading to $C_0 \sim 1/20$ $n$m and the range of $\alpha \sim$ $(5.48 \sim 8.66)$. According to Figure 15, the optimal shape for a self-assembly experiment is a cylindrical one. The elastic model thus confirms the dominance of cylindrical shape in *in vitro* experiment. It should be noted that our calculation using these HIV-1 parameters shows that conical shape is also lower in energy than spherical shape. However, the highly elongated cylindrical is the most stable one because (at

high spontaneous curvature parameter $\alpha$) it lowers the total bending of the capsid.

The situation is very interesting when HIV-1 capsid matures in the presence of the strong envelope membrane constraint. All elongated capsid greater than the membrane diameter $2R_m$ becomes very expensive in energy and only capsid with $L_{max} \sim 2R_m$ remains. The analytical formula, Eq. (33), shows that the energies of both cylindrical and conical with different cone angle $\theta_M$ are identical up to the first order. The third term of Eq. (33) is tabulated in Table 1 for theoretical value $B = 11\pi/30$ and the semi-empirical value $B = 1.27$ in the range of HIV-1. If one uses the theoretical value $B = 11\pi/30$, the terms proportional to $\alpha^2$ in the third expansion term are identically zero. This term only depends on the first order of $\alpha$. The energy difference between $E_5^{(2)}(\gamma, \alpha, 0)$ and $E_6^{(2)}(\gamma, \alpha, 0)$ (7-5 conical and cylindrical) is $0.07\kappa \sim 0.7\ k_BT$ which is less than $1\ k_BT$, the thermal energy in the whole range of $\alpha$ for HIV-1. On the other hand, if one uses the semi-empirical value $B \sim 1.27$, then the third term not only depends on $\alpha$, but also $\alpha^2$. Both $E_5^{(2)}(\gamma, \alpha, 0)/\kappa$ and $E_6^{(2)}(\gamma, \alpha, 0)/\kappa$ calculated with semi-empirical value $B \sim 1.27$ is higher than those with the theoretical value $B = 11\pi/30$ because $\alpha^2$ is not zero. In the semi-empirical case, the difference between the energy of the 7-5 conical shape and the cylindrical shape is less than the thermal energy $1\ k_BT$ up to $\alpha \sim 6.6$. This energy difference reaches $2.9\ k_BT$ at the highest value $\alpha = 8.66$ of the range of HIV-1 because of the quadratic of $\alpha$ term. However, the difference between $E_5^{(2)}(\gamma, \alpha, 0)/\kappa$ and $E_6^{(2)}(\gamma, \alpha, 0)/\kappa$ in the medium range of $\alpha$ for HIV-1 is also $\sim 1\ k_BT$. Therefore, our analytical expansion to the energy for conical and cylindrical shape calculated with both the theoretical and semi-empirical value shows that the energy difference between both shapes are negligible. On the other hand, if one uses the semi-empirical value $B \sim 1.30$, then the coefficient of the $\alpha^2$ term in $E_5^{(2)}(\gamma, \alpha, 0)/\kappa$ is 0.200 while in $E_6^{(2)}(\gamma, \alpha, 0)/\kappa$, it is 0.192. In other words, for this value of $B$, $E_5^{(2)}(\gamma, \alpha, 0)/\kappa$ would be slightly smaller than $E_6^{(2)}(\gamma, \alpha, 0)/\kappa$ at small $\alpha$ and slightly larger at large $\alpha$. When we use the word

**Table 1:** $E_M^{(2)}(\gamma, \alpha, 0)/\kappa$ for different $M$'s for empirical and theoretical values $B$.

| $B = 1.27$ | $E_2^{(2)}(\gamma, \alpha, 0)/\kappa$ | $E_3^{(2)}(\gamma, \alpha, 0)/\kappa$ | $E_4^{(2)}(\gamma, \alpha, 0)/\kappa$ | $E_5^{(2)}(\gamma, \alpha, 0)/\kappa$ | $E_6^{(2)}(\gamma, \alpha, 0)/\kappa$ |
|---|---|---|---|---|---|
| $\alpha = 5$ | $-5.70$ | $-5.59$ | $-5.55$ | $\mathbf{-5.53}$ | $\mathbf{-5.52}$ |
| $\alpha = 6$ | $-6.85$ | $-7.80$ | $-8.23$ | $\mathbf{-8.42}$ | $\mathbf{-8.48}$ |
| $\alpha = 7$ | $-7.38$ | $-9.57$ | $-10.55$ | $\mathbf{-11.00}$ | $\mathbf{-11.13}$ |
| $\alpha = 8$ | $-7.26$ | $-10.89$ | $-12.51$ | $\mathbf{-13.26}$ | $\mathbf{-13.48}$ |
| $\alpha = 9$ | $-6.51$ | $-11.77$ | $-14.11$ | $\mathbf{-15.20}$ | $\mathbf{-15.52}$ |
| $B = \frac{11\pi}{30}$ | $E_2^{(2)}(\gamma, \alpha, 0)/\kappa$ | $E_3^{(2)}(\gamma, \alpha, 0)/\kappa$ | $E_4^{(2)}(\gamma, \alpha, 0)/\kappa$ | $E_5^{(2)}(\gamma, \alpha, 0)/\kappa$ | $E_6^{(2)}(\gamma, \alpha, 0)/\kappa$ |
| $\alpha = 5$ | $-8.27$ | $-7.10$ | $-6.57$ | $\mathbf{-6.33}$ | $\mathbf{-6.26}$ |
| $\alpha = 6$ | $-12.16$ | $-10.99$ | $-10.47$ | $\mathbf{-10.22}$ | $\mathbf{-10.15}$ |
| $\alpha = 7$ | $-16.05$ | $-14.88$ | $-14.36$ | $\mathbf{-14.11}$ | $\mathbf{-14.04}$ |
| $\alpha = 8$ | $-19.94$ | $-18.77$ | $-18.25$ | $\mathbf{-18.00}$ | $\mathbf{-17.93}$ |
| $\alpha = 9$ | $-23.83$ | $-22.66$ | $-22.14$ | $\mathbf{-21.90}$ | $\mathbf{-21.82}$ |

"slightly", we mean that, even for $\alpha \sim 8$, the difference in value between these two different terms, and hence the difference between the energy of the conical shape and the cylindrical shape is less than $0.1\kappa \sim 1k_BT$, the thermal energy. This shows that the cylindrical and conical has the same probability to appear after *in vivo* maturation process. This agrees well with available experimental data.

Our numerical calculation also shows that the energies of conical and cylindrical shapes are nearly identical in the range of HIV-1 capsid. In Figure 17, we showed that cylindrical capsid has lower energy than the conical capsid at small $\alpha$ but the energies of the conical capsid and the cylindrical capsid are almost same until these capsids matures up to $2R_m$ at high $\alpha$. Also shown is the average value for the diameter of the HIV viral membrane, $2R_m$.

The optimal cylindrical shape in this case (without membrane) has length of 350 $nm$. In Figure 18, the total energy in the presence of the membrane is plotted as a function of the length of the capsid (for the typical surface tension or osmotic pressure of HIV-1 capsid). The 350 $nm$ cylindrical shell becomes very high in energy and is no longer the most stable one. Now the optimal cylindrical and conical capsids

have the same energies (the conical one is actually slightly smaller in energy but it is within our numerical error). More impressive is the fact that up to 40% of elongation, cylindrical and conical shapes have the same energy. This shows that the analytical expansion we used in Eq. (33) has very large range of validity.

In summary, we introduce a way to study the morphologies of HIV-1 capsid using Caspar-Klug construction for spherical capsid and the generalized isometric construction for non-spherical capsids. We suggested evidences for the question of diversity of *in vivo* assembled HIV-1 capsids by using the continuum elastic theory and the flexible elliptical membrane model. Both our numerical calculations for cylindrical and 7-5 conical capsids and analytical calculations for cylindrical and all conical capsids support that the presence of the viral membrane plays an important role in the diversity of *in vivo* retroviral capsids. Without the membrane, the cylindrical is the most dominant shape in CA protein assembly. However, with the membrane, retroviral conical capsids and cylindrical capsids have similar energy and equal probability to appear in the assembly.

# CHAPTER V

# RADIAL DISTRIBUTION OF RNA GENOMES PACKAGED INSIDE SPHERICAL VIRUSES

## *5.1 Introduction*

Viruses attract broad interests from physics community due to their ability of spontaneous self assembly. Many viruses can be produced both *in vivo* and *in vitro* as highly robust and monodisperse particles. As a result, beside biomedical applications, understanding virus assembly can also have novel promising applications in nanofabrication. At the basic level, viruses consist of viral genomes (RNA or DNA molecules) packaged inside a protective protein shell (viral capsid). The structures of viral capsids for most viruses are well understood from high-resolution experiments using cryoelectron microscopy or X-ray analysis [151, 78], as well as theoretical studies [95, 115]. Single-stranded RNA (ssRNA) viruses also package their genome spontaneously during assembly. Several theoretical studies have demonstrated that the interaction between capsid proteins and RNA nucleotide basis plays an important role in the RNA packaging process, both energetically and kinetically [71, 14, 132, 156, 72]. However, unlike the structural study of viral capsid, there is still a lack of general understanding of structure of packaged RNA. In references [14, 156, 72], different models of RNA packaging inside viruses were studied. However, all these works treat RNA molecules as *linear* flexible polymers. In this chapter, we want to address the question of how RNA molecules are arranged inside a spherical virus, *explicitly* taking into account the *branching* degree of freedom of RNA secondary structure. We focus on a particular class of ssRNA viruses where the interaction between capsid proteins and RNA molecules is non−specific and occurs dominantly at the inner surface of the

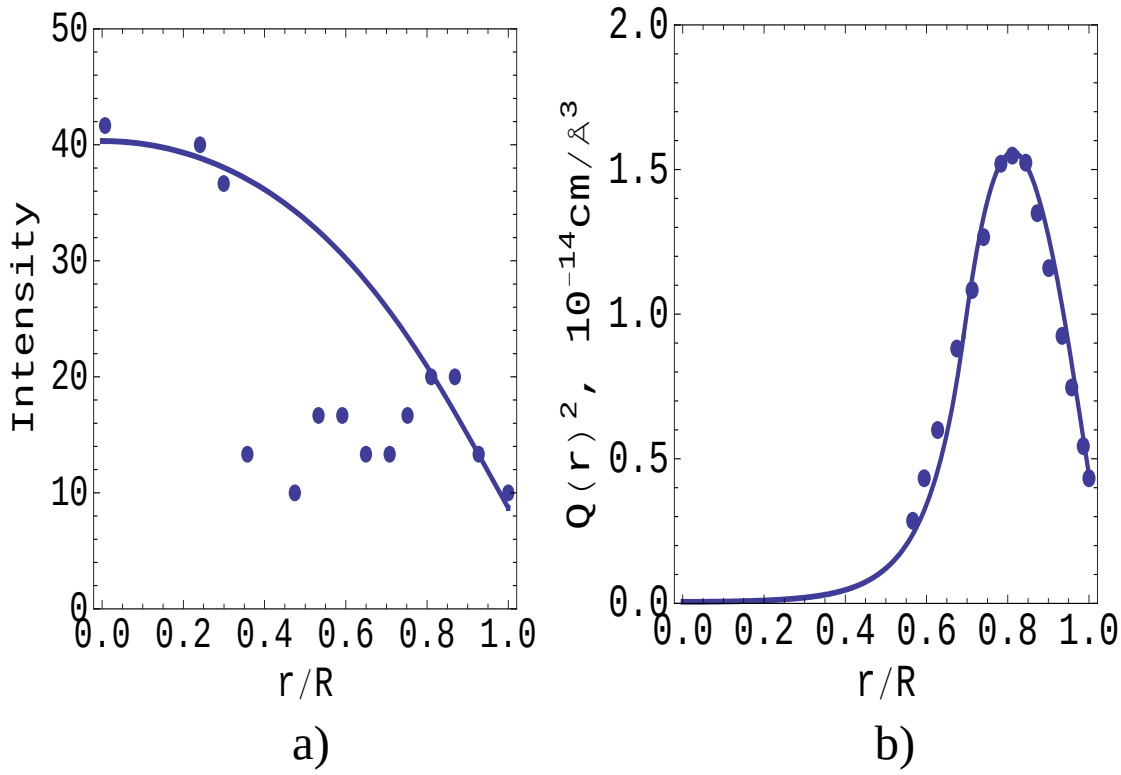


**Figure 19:** Two different profiles for RNA monomer concentration inside spherical viruses. Points are experimental data and solid lines are theoretical fit. a) Profile II, Eq. (48), fitted to RNA concentration of Dengue virus obtained from cryoelectron microscopy experiment [85]. b) Profile III, Eq. (51), fitted to RNA concentration of bacteriophage $MS_2$ obtained from small angle neutron scattering experiment [75].

capsid. This is the case for viruses where basic amino acids are located on the surface and electrostatic interaction is strongly screened in the bulk solution (examples of such viruses are bacteriophage $MS_2$, Q Beta, Dengue, Immature Yellow Fever,... generally viruses belonging to group B and C mentioned in Ref [72]). (In some viruses such as pariacoto virus [145], the viral capsid forces some fraction of RNA molecules to adopt it dodecahedron structure. In that case, the theory presented below should be applied to the free fraction of these RNAs.) Even though RNA-capsid interaction only occurs at the surface, RNA radial concentration profiles and the amount of RNA packaged inside a virus can be dictated by the strength of this interaction. The main result of this chapter is that there are two different profiles for the radial RNA

nucleotide concentration. For small capsid attraction, the RNA concentration is maximum at the center of the capsid. A representative virus (the Dengue virus) for this profile is shown in Figure 19a. For larger capsid attraction, the RNA concentration is maximum at a distance close to (but always smaller than) the inner capsid radius. A representative virus (the bacteriophage $MS_2$) for this profile is shown in Figure 19b. For the later case, the RNA molecules form a dense layer at the inner capsid surface. The thickness of this layer varies very slowly (logarithmic) with the capsid radius. As a result, the amount of RNA packaged inside such viruses is proportional to the capsid area (or the number of capsid proteins) instead of its volume.

## *5.2 Model*

Inner capsid of a spherical virus is considered as a hollow sphere only including RNA molecules which interact with the inner capsid surface in order to show how RNA molecules are radially distributed inside a spherical virus. It is well known that ssRNA molecules fold on themselves due to base-pairing interaction between their nucleotides. Because nucleotide sequence of ssRNA molecules is not perfect for such pairing, their secondary structure is highly nonlinear. To the first approximation, RNA molecules are considered to be highly flexible branch polymers which can fluctuate freely over all possible branching configurations. Different branching configurations are described in the schematic way shown in Figure 20, characterized by fugacities for "bi-functional" units (linear sequences), "tri-functional units" (branching points) and "endpoints" (stem-loops or hair-pins). We assume good solvent condition with repulsive interactions between the different units (with no "tertiary" pairing). Using a mean-field approximation [112] to a field theory for solutions of branching polymers of this type [98], one can write down an expression for the free energy density of RNA solution $W[Q(\vec{r})]$ as

$$\frac{W[Q(\vec{r})]}{m} = \frac{\epsilon}{2}Q(\vec{r})^2 - \frac{w}{6}Q(\vec{r})^3 + muQ(\vec{r})^4 - hQ(\vec{r}) \ , \tag{37}$$

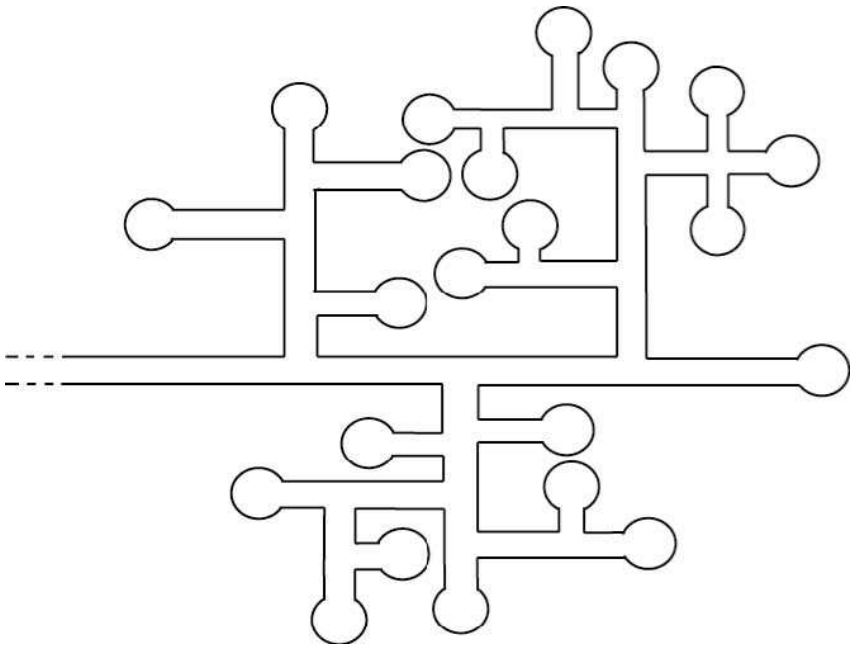

**Figure 20:** Schematic representation of the secondary structure of a single-stranded RNA molecule as a collection of linear sections, branch-points, and end-points. The molecule can freely fluctuate between different branching configurations.

where $\epsilon$, $w$, $h$ and $m$ are the fugacities of the monomers, branch points, the end-points and the whole polymers respectively. The coefficient $u$ is proportional to the second-order virial coefficient for monomer-monomer interaction (since RNA molecules are assumed to be in good solvent, $u$ is positive). $Q(\vec{r})$ is the order parameter of the field theory and is proportional to the concentration of end-points. Note that if one sets $w = 0$ (the branching degree of freedom is suppressed), Eq. (37) recovers the well known expression for the free energy density of a solution of linear polymers [35]. Based on this mean-field expression, it is suggested that RNA are prone to a surface condensation which is different from that of linear polymer[112]. In this chapter, we will use the mean-field expression, Eq. (37), to study how RNA molecules are packaged inside a virus. For simplicity, we model the viral capsid as a hollow sphere with inner radius $R$. We also assume that RNA molecules are radially distributed inside the capsid so that $Q(\vec{r}) \equiv Q(r)$ where $r$ is the radial distance from the center

of viral capsid. As a result, the excess free energy of the RNA molecules packaged inside a capsid can be written as

$$H_{\mathrm{MF}} = H_s + \int_0^R 4\pi r^2 dr \left\{ \frac{m}{2} \left( \frac{dQ}{dr} \right)^2 + \Delta W[Q] \right\}, \tag{38}$$

with $\Delta W[Q(r)] = W[Q(r)] - W[Q_{bulk}]$. The first term in Eq. (38) denotes the interaction energy of the capsid proteins with the RNA molecules. Assuming this interaction occurs only at the inner capsid surface, $H_s$ can be written as the sum of contributions from monomers and endpoint adsorptions:

$$H_s = 4\pi R^2 m[-\gamma_1 Q(R) - \gamma_2 Q(R)^2/2], \tag{39}$$

where $\gamma_{1,2}$ are the strengths of the adsorption.

Due to the cubic term proportional to $w$ in Eq. (37), for small positive $\epsilon$, the free energy density $W(Q)$ has two minima, $Q_D$ and $Q_C$, corresponding to, respectively, the mean-field order parameter of a dilute bulk RNA solution and that of a condensed bulk RNA solution. A first-order condensation transition takes place when $W(Q_D) = W(Q_C)$. We will always assume RNA solution lies at this coexistence regime so that both the dilute and dense phases of RNA solution are close in energy. Therefore, we set bulk value $Q_{bulk} = Q_D$. The equilibrium RNA concentration profile corresponds to the profile $Q(r)$ that minimizes the Hamiltonian Eq. (38). Setting the functional derivative, $\delta H_{MF}/\delta Q$ to zero, we obtain the Euler-Lagrangian equation

$$\frac{d^2Q}{dr^2} + \frac{2}{r}\frac{dQ}{dr} - \frac{1}{m}\frac{\partial \Delta W}{\partial Q} = 0, \tag{40}$$

and a boundary condition at the inner capsid surface:

$$\left.\frac{dQ}{dr}\right|_{r=R} = \frac{H_s^{'}[Q(R)]}{4\pi R^2 m} = -\gamma_1 - \gamma_2 Q(R)\ . \tag{41}$$

The detail derivation for Euler Lagrange equation and boundary condition is shown in Appendix C.1. To proceed further, we approximate $\Delta W$ using the double parabolic

potential form[70]:

$$\Delta W(Q) = \begin{cases} \frac{1}{2}m\lambda_D^2(Q-Q_D)^2 & \text{for } Q < Q_m \\ \frac{1}{2}m\lambda_C^2(Q-Q_C)^2 & \text{for } Q > Q_m \end{cases}, \tag{42}$$

where $Q_m = (\lambda_D Q_D + \lambda_C Q_C)/(\lambda_D + \lambda_C)$ is the point where the two parabolas cross each other forming a cusp. Here, $Q_m$ can be calculated by setting two double parabolas of Eq. (42) equal:

$$\frac{1}{2}m\lambda_D^2(Q_m - Q_D)^2 = \frac{1}{2}m\lambda_C^2(Q_m - Q_C)^2 \ . \tag{43}$$

The two coefficients $\lambda_D^2, \lambda_C^2$ are the stiffness of the free energy density of RNA solution near the two minima. They are proportional to the corresponding correlation lengths of the two phases. In general, this double parabolic potential form for the free energy density breaks down near the critical temperature where the first order transition becomes second order, or when the fugacity of branch points, $w$, goes to 0 (the branching degree of freedom is suppressed and RNA molecules behave as a linear polymer). However, it was shown [112] that the mean-field expression, Eq. (37), breaks down before this limit is approached. If one stays within the limit of mean-field theory, the double parabola approximation is a reasonable approximation. We will come back to its limitation in later discussion.

## *5.3 RNA condensation inside a spherical capsid*

With this approximate form of $\Delta W$, Eq. (40) becomes linear and easy to solve. The general solution is a linear combination of $\exp(\pm\lambda_{D,C}r)/r$. We try to solve six possible concentration profiles for the RNA molecules in this chapter.

*Profile I.* If for all $r$, $Q(r) < Q_m$, we can set up the solution to the Euler equation is

$$Q(r) = C_1\frac{\exp(\lambda_D r)}{r} + C_2\frac{\exp(-\lambda_D r)}{r} + C_0, \tag{44}$$

where $C_1$, $C_2$, and, $C_0$ are constants determined by the boundary condition Eq. (41) and the potential form Eq. (42) with the condition that the free energy density at

$r = 0$, $Q(r = 0)$ should be finite. This finity condition at $r = 0$ gives $C_1 = -C_2$. We can easily obtain $C_0 = Q_D$. This relation is the result of the Euler Lagrange Equation Eq. (40). After putting Eq. (44) into the boundary condition Eq. (41), we obtain the solution to the Euler equation

$$Q(r) = -C_{10}\sinh(\lambda_D r)/(\lambda_D r) + Q_D \ . \tag{45}$$

where

$$C_{10} = \frac{(\gamma_1 + \gamma_2 Q_D)R}{\cosh(\lambda_D R) + (\gamma_2 R - 1)\sinh(\lambda_D R)/(\lambda_D R)}. \tag{46}$$

Because the interaction of the RNA monomers with the viral capsid is attractive, $\gamma_{1,2} > 0$, the coefficient $C_{10}$ is a positive quantity. According to Eq. (45), this means that for all $r$, the endpoint (and monomer) concentration in this profile is always smaller than the bulk value, $Q(r) < Q_D = Q_{bulk}$. This is a non physical situation. Therefore, we discard this solution from later consideration.

*Profile II.* The second possibility is the case that for all $r$, $Q > Q_m$. We can also write down the general solution to this case

$$Q(r) = C_1\frac{\exp(\lambda_C r)}{r} + C_2\frac{\exp(-\lambda_C r)}{r} + C_0. \tag{47}$$

$C_1$, $C_2$, and $C_0$ can also be determined by the fact that $Q(r = 0)$ should be finite and a result of the Euler Lagrange equation Eq. (40) leading to $C_1 = -C_2$ and $C_0 = Q_C$. The solving method is the same as that used in *Profile I* except for using the stiffness coefficient $\lambda_C$ for a condensed phase instead of $\lambda_D$ for a dilute phase. Accordingly, by inserting Eq. (47) into the boundary condition Eq. (41), we arrive at the solution

$$Q(r) = -C_{20}\sinh(\lambda_C r)/(\lambda_C r) + Q_C \ , \tag{48}$$

where

$$C_{20} = \frac{(\gamma_1 + \gamma_2 Q_C)R}{\cosh(\lambda_C R) + (\gamma_2 R - 1)\sinh(\lambda_C R)/(\lambda_C R)} \ . \tag{49}$$

This solution is a monotonously decreasing function of $r$ and the RNA concentration is maximum at the *center* of the capsid. Because of the requirement that $Q(R)$ must

be greater than $Q_m$, this profile is possible only for very weak adsorption (in practice, $\lambda_C R \gg 1$, this requirement means $(\gamma_1/Q_C+\gamma_2)/\lambda_D < 1$). As a result, RNA monomers want to concentrate at the center of the capsid to gain their configurational entropy (minimizing the gradient term in Eq. (38) ).

*Profile III.* The third possibility is that $Q(r)$ passes through $Q_m$ at some distant $r = r_0$ $(0 < r_0 < R)$ such that $Q(r = r_0) = Q_m$. We can interpret $r_0$ as the boundary between the dilute and the condensed phases of RNA molecules inside the capsid. The solution can be written down by

$$Q(r) = \begin{cases} C_1 \frac{\exp(\lambda_D r)}{\lambda_D r} + C_2 \frac{\exp(-\lambda_D r)}{\lambda_D r} + Q_D & \text{for} \quad r < r_0 \\ C_{31} \frac{\exp(\lambda_C r)}{\lambda_C r} + C_{32} \frac{\exp(-\lambda_C r)}{\lambda_C r} + Q_C & \text{for} \quad r_0 < r < R. \end{cases} \tag{50}$$

With $Q(r = 0)$ being finite, we can rewrite the above Equation,

$$Q(r) = \begin{cases} (Q_0 - Q_D) \frac{\sinh(\lambda_D r)}{\lambda_D r} + Q_D & \text{for} \quad r < r_0 \\ C_{31} \frac{\exp(\lambda_C r)}{\lambda_C r} + C_{32} \frac{\exp(-\lambda_C r)}{\lambda_C r} + Q_C & \text{for} \quad r_0 < r < R. \end{cases} \tag{51}$$

Now we have four unknown parameters $C_{31}$, $C_{32}$, $r_0$ and $Q_0$. We have three requirements (two continuity requirements and one condition at the interface $r_0$): (1) the density profile $Q(r)$ should be continuous and (2) the derivative of the density profile $Q'(r)$ also should be continuous. (3) the density profile $Q(r = r_0)$ at the interface between a dilute and a condensed phase for RNA molecule is equal to $Q_m$. We also have one more boundary condition Eq. (41). Using two continuity requirements $Q(r = r_0^+) = Q(r = r_0^-)$ and $Q'(r = r_0^+) = Q'(r = r_0^-)$, we can obtain the next two equations for $C_1$ and $C_2$:

$$C_{31} \frac{\exp(\lambda_C r_0)}{\lambda_C r_0} + C_{32} \frac{\exp(-\lambda_C r_0)}{\lambda_C r_0} + Q_C = (Q_0 - Q_D) \frac{\exp(\lambda_D r_0) - \exp(-\lambda_D r_0)}{2\lambda_D r_0} + Q_D, \tag{52}$$

and

$$C_{31}\left(-\frac{\exp(\lambda_C r_0)}{\lambda_C r_0^2}+\frac{\exp(\lambda_C r_0)}{r_0}\right)+C_{32}\left(-\frac{\exp(-\lambda_C r_0)}{\lambda_C r_0^2}-\frac{\exp(-\lambda_C r_0)}{r_0}\right)$$
$$=(Q_0-Q_D)\left(-\frac{\exp(\lambda_D r_0)-\exp(-\lambda_D r_0)}{2\lambda_D r_0^2}+\frac{\exp(\lambda_D r_0)+\exp(-\lambda_D r_0)}{2r_0}\right). \tag{53}$$

Solving for $C_{31}$ and $C_{32}$ from two equations (52) and (53), we obtain

$$\begin{aligned}
C_{31} &= -\exp[-(\lambda_C+\lambda_D)r_0](Q_0-Q_D)(\lambda_C/\lambda_D-1)/4 \\
&\quad +\exp[-(\lambda_C-\lambda_D)r_0](Q_0-Q_D)(\lambda_C/\lambda_D+1)/4 \\
&\quad -\exp(-\lambda_C r_0)(Q_C-Q_D)(\lambda_C r_0+1)/2, \\
C_{32} &= \exp[(\lambda_C+\lambda_D)r_0](Q_0-Q_D)(\lambda_C/\lambda_D-1)/4 \\
&\quad -\exp[(\lambda_C-\lambda_D)r_0](Q_0-Q_D)(\lambda_C/\lambda_D+1)/4 \\
&\quad -\exp(\lambda_C r_0)(Q_C-Q_D)(\lambda_C r_0-1)/2.
\end{aligned}$$

$r_0$ and $Q_0$ are two unknowns in the solution Eq. (51) with two coefficients $C_{31}$ and $C_{32}$. To find the unknowns $r_0$ and $Q_0$, the boundary condition, Eq. (41), and the condition $Q(r_0)=Q_m$. The later condition with Eq. (51) gives

$$Q_m=(Q_0-Q_D)\frac{\exp(\lambda_D r_0)-\exp(-\lambda_D r_0)}{2\lambda_D r_0}+Q_D. \tag{54}$$

Replacing $Q_m$ by $(\lambda_D Q_D+\lambda_C Q_C)/(\lambda_D+\lambda_C)$ and solving for $Q_0$, it gives

$$Q_0-Q_D=(Q_C-Q_D)\frac{\lambda_C}{\lambda_C+\lambda_D}\frac{\lambda_D r_0}{\sinh(\lambda_D r_0)}\ . \tag{55}$$

The above relation will be used in the equation obtained from the boundary condition. The boundary condition, Eq. (41) gives another equation :

$$C_{31}\left(-\frac{\exp(\lambda_C R)}{\lambda_C R^2}+\frac{\exp(\lambda_C R)}{R}\right)+C_{32}\left(-\frac{\exp(-\lambda_C R)}{\lambda_C R^2}-\frac{\exp(-\lambda_C R)}{R}\right)$$
$$=-\gamma_1-\gamma_2 Q(R)\ . \tag{56}$$

Plugging $Q(R)$ of Eq. (51) into the above equation and factorize with $C_{31}$ and $C_{32}$, we get

$$
\begin{aligned}
&[(\lambda_C+\gamma_2)R-1]C_{31}\exp(\lambda_C R)-[(\lambda_C-\gamma_2)R+1]C_{32}\exp(-\lambda_C R)\\
&=-(\gamma_1+\gamma_2 Q_C)\lambda_C R^2\ . \qquad (57)
\end{aligned}
$$

Inserting $C_{31}$ and $C_{32}$ into Eq. (57), it gives

$$
\begin{aligned}
&[(\lambda_C+\gamma_2)R-1]\exp[\lambda_C(R-r_0)]\times\\
&\left[-\frac{(\lambda_C-\lambda_D)}{4\lambda_C\lambda_D}\exp(-\lambda_D r_0)+\frac{(\lambda_C+\lambda_D)}{4\lambda_C\lambda_D}\exp(\lambda_D r_0)-\frac{(\lambda_C r_0+1)}{2\lambda_C}\frac{Q_C-Q_D}{Q_0-Q_D}\right]\\
&-[(\lambda_C-\gamma_2)R+1]\exp[-\lambda_C(R-r_0)]\times\\
&\left[\frac{(\lambda_C-\lambda_D)}{4\lambda_C\lambda_D}\exp(\lambda_D r_0)-\frac{(\lambda_C+\lambda_D)}{4\lambda_C\lambda_D}\exp(-\lambda_D r_0)-\frac{(\lambda_C r_0-1)}{2\lambda_C}\frac{Q_C-Q_D}{Q_0-Q_D}\right]\\
&+\frac{(\gamma_1+\gamma_2 Q_C)R^2}{Q_0-Q_D}=0\ . \qquad (58)
\end{aligned}
$$

Using the relation between $Q_0-Q_D$ and $Q_C-Q_D$, Eq. (55), we get

$$
\begin{aligned}
&[(\lambda_C+\gamma_2)R-1]\exp[\lambda_C(R-r_0)]\times\\
&\left[-\frac{(\lambda_C-\lambda_D)}{4\lambda_C\lambda_D}\exp(-\lambda_D r_0)+\frac{(\lambda_C+\lambda_D)}{4\lambda_C\lambda_D}\exp(\lambda_D r_0)-\frac{(\lambda_C r_0+1)}{2\lambda_C}\frac{(\lambda_C+\lambda_D)}{\lambda_C}\frac{\sinh(\lambda_D r_0)}{\lambda_D r_0}\right]\\
&-[(\lambda_C-\gamma_2)R+1]\exp[-\lambda_C(R-r_0)]\times\\
&\left[\frac{(\lambda_C-\lambda_D)}{4\lambda_C\lambda_D}\exp(\lambda_D r_0)-\frac{(\lambda_C+\lambda_D)}{4\lambda_C\lambda_D}\exp(-\lambda_D r_0)-\frac{(\lambda_C r_0-1)}{2\lambda_C}\frac{(\lambda_C+\lambda_D)}{\lambda_C}\frac{\sinh(\lambda_D r_0)}{\lambda_D r_0}\right]\\
&+\frac{(\gamma_1+\gamma_2 Q_C)R^2}{Q_C-Q_D}\frac{(\lambda_C+\lambda_D)}{\lambda_C}\frac{\sinh(\lambda_D r_0)}{\lambda_D r_0}=0\ . \qquad (59)
\end{aligned}
$$

After some simple calculations, we arrive at the equation for $r_0$:

$$
\begin{aligned}
&\left(1+\frac{\gamma_2}{\lambda_C}-\frac{1}{\lambda_C R}\right)\left[-\frac{\lambda_D r_0\exp(-\lambda_D r_0)}{\sinh(\lambda_D r_0)}+1+\frac{\lambda_D}{\lambda_C}\right]u^2\\
&-2\lambda_s R u\\
&-\left(1-\frac{\gamma_2}{\lambda_C}+\frac{1}{\lambda_C R}\right)\left[\frac{\lambda_D r_0\exp(\lambda_D r_0)}{\sinh(\lambda_D r_0)}-1-\frac{\lambda_D}{\lambda_C}\right]=0, \qquad (60)
\end{aligned}
$$

where $u=\exp[\lambda_C(R-r_0)]$ . The parameter

$$
\lambda_s=(1+\lambda_D/\lambda_C)(\gamma_1+\gamma_2 Q_C)/(Q_C-Q_D)\ , \qquad (61)
$$

is proportional to the strength of RNA adsorption at the inner capsid surface and has dimension of inverse length. The solution to Eq. (60) is

$$\begin{aligned} u &= \left\{ \lambda_s R \pm \left[ (\lambda_s R)^2 + \left( 1 - \left( \frac{\gamma_2}{\lambda_C} - \frac{1}{\lambda_C R} \right)^2 \right) \right. \right. \\ &\times \left. \left. \left( - \left( \frac{\lambda_D r_o}{\sinh(\lambda_D r_o)} \right)^2 + 2 \left( 1 + \frac{\lambda_D}{\lambda_C} \right) \frac{\lambda_D r_o}{\tanh(\lambda_D r_o)} - (1 + \frac{\lambda_D}{\lambda_C})^2 \right) \right]^{\frac{1}{2}} \right\} \Big/ \\ &\left\{ \left( 1 + \frac{\gamma_2}{\lambda_C} - \frac{1}{\lambda_C R} \right) \times \left( - \frac{\lambda_D r_o \exp(-\lambda_D r_o)}{\sinh(\lambda_D r_o)} + 1 + \frac{\lambda_D}{\lambda_C} \right) \right\} \end{aligned} \tag{62}$$

*Profile IV*. The fourth possibility is that $Q(r)$ passes through $Q_m$ at some distant $r = r_0$ $(0 < r_0 < R)$ such that $Q(r = r_0) = Q_m$ and $Q(r)$ starts from above $Q_m$ at inner capsid center and decreases as the $r$ approaches to $r_0$. The solution to the Euler Lagrange equation (40) can be set up dy

$$Q(r) = \begin{cases} C_1 \frac{\exp(\lambda_C r)}{\lambda_C r} + C_2 \frac{\exp(-\lambda_C r)}{\lambda_C r} + Q_C & \text{for} \quad r < r_0 \\ C_{41} \frac{\exp(\lambda_D r)}{\lambda_D r} + C_{42} \frac{\exp(-\lambda_D r)}{\lambda_D r} + Q_D & \text{for} \quad r_0 < r < R. \end{cases} \tag{63}$$

With the density profile $Q(r)$ at $r = 0$ being finite, we can write down the solution

$$Q(r) = \begin{cases} (Q_0 - Q_C) \frac{\sinh(\lambda_C r)}{\lambda_C r} + Q_C & \text{for} \quad r < r_0 \\ C_{41} \frac{\exp(\lambda_D r)}{\lambda_D r} + C_{42} \frac{\exp(-\lambda_D r)}{\lambda_D r} + Q_D & \text{for} \quad r_0 < r < R \end{cases} \tag{64}$$

where $Q_0 = Q(0)$. In Eq. (64), there are four unknowns which can be solved by four requirements: two continuities of the density profile and its derivative at its phase boundary at $r_0$ (1) $Q(r_0^+) = Q(r_0^-)$ and 2) $Q'(r_0^+) = Q'(r_0^-)$), 3) one condition $Q_m = Q(r_0)$, 4) the boundary condition Eq. (41). Applying the first two requirements, we can get two equations for $C_1$ and $C_2$:

$$C_{41} \frac{\exp(\lambda_D r_0)}{\lambda_D r_0} + C_{42} \frac{\exp(-\lambda_D r_0)}{\lambda_D r_0} + Q_D = (Q_0 - Q_C) \frac{\exp(\lambda_C r_0) - \exp(-\lambda_C r_0)}{2 \lambda_C r_0} + Q_C, \tag{65}$$

and

$$C_{41}\left(-\frac{\exp(\lambda_D r_0)}{\lambda_D r_0^2}+\frac{\exp(\lambda_D r_0)}{r_0}\right)+C_{42}\left(-\frac{\exp(-\lambda_D r_0)}{\lambda_D r_0^2}-\frac{\exp(-\lambda_D r_0)}{r_0}\right)$$
$$=(Q_0-Q_C)\left(-\frac{\exp(\lambda_C r_0)-\exp(-\lambda_C r_0)}{2\lambda_C r_0^2}+\frac{\exp(\lambda_C r_0)+\exp(-\lambda_C r_0)}{2r_0}\right). \tag{66}$$

Solving Eq. (65) and Eq. (66), we can express $C_1$ and $C_2$:

$$\begin{aligned}
C_{41} &= -\exp[-(\lambda_C+\lambda_D)r_0](Q_0-Q_C)(\lambda_D/\lambda_C-1)/4 \\
&\quad +\exp[(\lambda_C-\lambda_D)r_0](Q_0-Q_C)(\lambda_D/\lambda_C+1)/4 \\
&\quad +\exp(-\lambda_D r_0)(Q_C-Q_D)(\lambda_D r_0+1)/2, \\
C_{42} &= \exp[(\lambda_C+\lambda_D)r_0](Q_0-Q_C)(\lambda_D/\lambda_C-1)/4 \\
&\quad -\exp[-(\lambda_C-\lambda_D)r_0](Q_0-Q_C)(\lambda_D/\lambda_C+1)/4 \\
&\quad +\exp(\lambda_D r_0)(Q_C-Q_D)(\lambda_D r_0-1)/2.
\end{aligned}$$

$r_0$ and $Q_0$ are two unknowns in the solution above and there are two requirements left. The condition $Q(r_0)=Q_m$ gives

$$Q_0-Q_C=-(Q_C-Q_D)\frac{\lambda_D}{\lambda_C+\lambda_D}\frac{\lambda_C r_0}{\sinh(\lambda_C r_0)}\ . \tag{67}$$

The above relation will be used in the equation obtained from the boundary condition. The boundary condition, Eq. (41) gives another left equation :

$$C_{41}\left(-\frac{\exp(\lambda_D R)}{\lambda_D R^2}+\frac{\exp(\lambda_D R)}{R}\right)+C_{42}\left(-\frac{\exp(-\lambda_D R)}{\lambda_D R^2}-\frac{\exp(-\lambda_D R)}{R}\right)$$
$$=-\gamma_1-\gamma_2 Q(R)\ . \tag{68}$$

Plugging $Q(R)$ of Eq. (64) into the above equation and factorize with $C_{41}$ and $C_{42}$, we get

$$[(\lambda_D+\gamma_2)R-1]C_{41}\exp(\lambda_D R)-[(\lambda_D-\gamma_2)R+1]C_{42}\exp(-\lambda_D R)$$
$$=-(\gamma_1+\gamma_2 Q_D)\lambda_D R^2\ . \tag{69}$$

Inserting $C_{41}$ and $C_{42}$ into Eq. (69), it gives

$$
\begin{aligned}
&[(\lambda_D+\gamma_2)R-1]\exp[\lambda_D(R-r_0)]\times \\
&\left[\frac{(\lambda_C-\lambda_D)}{4\lambda_C\lambda_D}\exp(-\lambda_C r_0)+\frac{(\lambda_C+\lambda_D)}{4\lambda_C\lambda_D}\exp(\lambda_C r_0)+\frac{(\lambda_D r_0+1)}{2\lambda_D}\frac{Q_C-Q_D}{Q_0-Q_C}\right] \\
&-[(\lambda_D-\gamma_2)R+1]\exp[-\lambda_D(R-r_0)]\times \\
&\left[-\frac{(\lambda_C-\lambda_D)}{4\lambda_C\lambda_D}\exp(\lambda_C r_0)-\frac{(\lambda_C+\lambda_D)}{4\lambda_C\lambda_D}\exp(-\lambda_C r_0)+\frac{(\lambda_D r_0-1)}{2\lambda_D}\frac{Q_C-Q_D}{Q_0-Q_C}\right] \\
&+\frac{(\gamma_1+\gamma_2 Q_D)R^2}{Q_0-Q_C}=0\ .
\end{aligned}
\tag{70}
$$

Using the relation between $Q_0-Q_C$ and $Q_C-Q_D$, Eq. (67), we get

$$
\begin{aligned}
&[(\lambda_D+\gamma_2)R-1]\exp[\lambda_D(R-r_0)]\times \\
&\left[\frac{(\lambda_C-\lambda_D)}{4\lambda_C\lambda_D}\exp(-\lambda_C r_0)+\frac{(\lambda_C+\lambda_D)}{4\lambda_C\lambda_D}\exp(\lambda_C r_0)-\frac{(\lambda_D r_0+1)}{2\lambda_D}\frac{(\lambda_C+\lambda_D)}{\lambda_D}\frac{\sinh(\lambda_C r_0)}{\lambda_C r_0}\right] \\
&-[(\lambda_D-\gamma_2)R+1]\exp[-\lambda_D(R-r_0)]\times \\
&\left[-\frac{(\lambda_C-\lambda_D)}{4\lambda_C\lambda_D}\exp(\lambda_C r_0)-\frac{(\lambda_C+\lambda_D)}{4\lambda_C\lambda_D}\exp(-\lambda_C r_0)-\frac{(\lambda_D r_0-1)}{2\lambda_D}\frac{(\lambda_C+\lambda_D)}{\lambda_D}\frac{\sinh(\lambda_C r_0)}{\lambda_C r_0}\right] \\
&-\frac{(\gamma_1+\gamma_2 Q_D)R^2}{Q_C-Q_D}\frac{(\lambda_C+\lambda_D)}{\lambda_D}\frac{\sinh(\lambda_C r_0)}{\lambda_C r_0}=0\ .
\end{aligned}
\tag{71}
$$

After doing some algebra, we can finally get

$$
\begin{aligned}
&\left(1+\frac{\gamma_2}{\lambda_D}-\frac{1}{\lambda_D R}\right)\left[-\frac{\lambda_C r_0\exp(-\lambda_C r_0)}{\sinh(\lambda_C r_0)}+1+\frac{\lambda_C}{\lambda_D}\right]u^2 \\
&+2\lambda_s^{'}Ru \\
&-\left(1-\frac{\gamma_2}{\lambda_D}+\frac{1}{\lambda_D R}\right)\left[\frac{\lambda_C r_0\exp(\lambda_C r_0)}{\sinh(\lambda_C r_0)}-1-\frac{\lambda_C}{\lambda_D}\right]=0,
\end{aligned}
\tag{72}
$$

where $u=\exp[\lambda_D(R-r_0)]$ . The parameter

$$
\lambda_s^{'}=(1+\lambda_C/\lambda_D)(\gamma_1+\gamma_2 Q_D)/(Q_C-Q_D) \tag{73}
$$

is proportional to the strength of RNA adsorption at the inner capsid surface and has

dimension of inverse length. The solution to Eq. (72) is

$$
\begin{aligned}
u &= \left\{ -\lambda_s^{'} R \pm \left[ \left(\lambda_s^{'} R\right)^2 + \left( 1 - \left( \frac{\gamma_2}{\lambda_D} - \frac{1}{\lambda_D R} \right)^2 \right) \right.\right. \\
&\times \left.\left. \left( - \left( \frac{\lambda_C r_o}{\sinh(\lambda_C r_o)} \right)^2 + 2 \left( 1 + \frac{\lambda_C}{\lambda_D} \right) \frac{\lambda_C r_o}{\tanh(\lambda_C r_o)} - \left( 1 + \frac{\lambda_C}{\lambda_D} \right)^2 \right) \right]^{\frac{1}{2}} \right\} \Bigg/ \\
&\left\{ \left( 1 + \frac{\gamma_2}{\lambda_D} - \frac{1}{\lambda_D R} \right) \times \left( - \frac{\lambda_C r_o \exp(-\lambda_C r_o)}{\sinh(\lambda_C r_o)} + 1 + \frac{\lambda_C}{\lambda_D} \right) \right\} \qquad (74)
\end{aligned}
$$

As we consider simple calculations, we can compare the solutions Eq. (51) of *Profile III* with Eq. (64) of *Profile IV*. Eq. (64) can be easily obtained by interchanging subscripts between a condensed phase and a dilute phase from Eq. (51) without actually solving the Euler Lagrange Equation Eq. (40) with the boundary conditions Eq. (41). Also, the rest of equations such as Eq. (67), Eq. (72) and Eq. (74) can be derived from Eq. (55), Eq. (60) and Eq. (62) respectively by substituting each other ($\lambda_C \leftrightarrow \lambda_D$, $Q_C \leftrightarrow Q_D$). This observation can reduce a little heavy calculations into simple ones.

*Profile V*. The fifth possibility is that $Q(r)$ increases and passes through $Q_m$ at $r = r_0$ $(0 < r_0 < R)$ such that $Q(r = r_0) = Q_m$ and passes again through $Q_m$ at $r = r_1$ $(r_0 < r_1 < R)$. $r_1$ can also be considered as the boundary between the dilute and the condensed phases of RNA molecules inside the capsid as $r_0$. A dilute phase changes into a condensed phase at $r_0$ and the condensed phase again changes into a dilute phase at $r_1$.

The solution to the Euler Lagrange Equation (40) can be written by

$$
Q(r) = \begin{cases}
(Q_0 - Q_D)\frac{\sinh(\lambda_D r)}{\lambda_D r} + Q_D & \text{for} \quad r < r_0 \\
C_{31}\frac{\exp(\lambda_C r)}{\lambda_C r} + C_{32}\frac{\exp(-\lambda_C r)}{\lambda_C r} + Q_C & \text{for} \quad r_0 < r < r_1 \\
C_{51}\frac{\exp(\lambda_D r)}{\lambda_D r} + C_{52}\frac{\exp(-\lambda_D r)}{\lambda_D r} + Q_D & \text{for} \quad r_1 < r < R
\end{cases} \qquad (75)
$$

where $Q_0 = Q(0)$. The coefficients $C_{31}$ and $C_{32}$ is the same as $C_{31}$ and $C_{32}$ in *Profile III* because we can apply the same continuities of $Q(r)$ and $Q'(r)$ at $r_0$, and follow the process of calculation of *Profile III* with the same requirements $Q_0 = Q(0)$ and

$Q(r_0) = Q_m$. There are still four unknown parameters $C_{51}$, $C_{52}$, $r_0$ and $r_1$. We can obtain these four unknowns with two continuities of the density profile at $r_1$, one requirement $Q(r_1) = Q_m$, and the boundary condition, Eq. (41).

Let's apply two continuity conditions 1) $Q(r_1^-) = Q(r_1^+)$ and 2) $Q'(r_1^-) = Q'(r_1^+)$. The first continuity condition gives

$$C_{31}\frac{\exp(\lambda_C r_1)}{\lambda_C r_1} + C_{32}\frac{\exp(-\lambda_C r_1)}{\lambda_C r_1} + Q_C = C_{51}\frac{\exp(\lambda_D r_1)}{\lambda_D r_1} + C_{52}\frac{\exp(-\lambda_D r_1)}{\lambda_D r_1} + Q_D. \tag{76}$$

The second continuity condition gives

$$\begin{aligned} &C_{31}\left(-\frac{\exp(\lambda_C r_1)}{\lambda_C r_1^2} + \frac{\exp(\lambda_C r_1)}{r_1}\right) + C_{32}\left(-\frac{\exp(-\lambda_C r_1)}{\lambda_C r_1^2} - \frac{\exp(-\lambda_C r_1)}{r_1}\right) = \\ &C_{51}\left(-\frac{\exp(\lambda_D r_1)}{\lambda_D r_1^2} + \frac{\exp(\lambda_D r_1)}{r_1}\right) + C_{52}\left(-\frac{\exp(-\lambda_D r_1)}{\lambda_D r_1^2} - \frac{\exp(-\lambda_D r_1)}{r_1}\right). \end{aligned} \tag{77}$$

Combining Eq. (76) and Eq. (77), we can solve for $C_{51}$ and $C_{52}$:

$$\begin{aligned} C_{51} &= C_{32}\exp[-(\lambda_C + \lambda_D)r_1](\lambda_D/\lambda_C - 1)/2 \\ &\quad + C_{31}\exp[(\lambda_C - \lambda_D)r_1](\lambda_D/\lambda_C + 1)/2 \\ &\quad + \exp(-\lambda_D r_1)(Q_C - Q_D)(\lambda_D r_1 + 1)/2, \\ C_{52} &= C_{31}\exp[(\lambda_C + \lambda_D)r_1](\lambda_D/\lambda_C - 1)/2 \\ &\quad + C_{32}\exp[-(\lambda_C - \lambda_D)r_1](\lambda_D/\lambda_C + 1)/2 \\ &\quad + \exp(\lambda_D r_1)(Q_C - Q_D)(\lambda_D r_1 - 1)/2, \end{aligned}$$

where

$$
\begin{aligned}
C_{31} &= -\exp[-(\lambda_C+\lambda_D)r_0](Q_0-Q_D)(\lambda_C/\lambda_D-1)/4 \\
&\quad +\exp[-(\lambda_C-\lambda_D)r_0](Q_0-Q_D)(\lambda_C/\lambda_D+1)/4 \\
&\quad -\exp(-\lambda_C r_0)(Q_C-Q_D)(\lambda_C r_0+1)/2, \\
C_{32} &= \exp[(\lambda_C+\lambda_D)r_0](Q_0-Q_D)(\lambda_C/\lambda_D-1)/4 \\
&\quad -\exp[(\lambda_C-\lambda_D)r_0](Q_0-Q_D)(\lambda_C/\lambda_D+1)/4 \\
&\quad -\exp(\lambda_C r_0)(Q_C-Q_D)(\lambda_C r_0-1)/2.
\end{aligned}
$$

There are still two unknowns $r_0$ and $r_1$ in the density profile. The condition $Q(r_1)=Q_m$ gives

$$
Q_m = C_{31}\frac{\exp(\lambda_C r_1)}{\lambda_C r_1} + C_{32}\frac{\exp(-\lambda_C r_1)}{\lambda_C r_1} + Q_C \ . \tag{78}
$$

Using the relation $Q_m - Q_C = -\frac{\lambda_D(Q_C-Q_D)}{\lambda_C+\lambda_D}$, Eq. (78) can be ordered

$$
C_{31}\frac{\exp(\lambda_C r_1)}{\lambda_C r_1} + C_{32}\frac{\exp(-\lambda_C r_1)}{\lambda_C r_1} = \frac{-\lambda_D}{\lambda_C+\lambda_D}(Q_C-Q_D) \ . \tag{79}
$$

Plugging $C_{31}$ and $C_{32}$ into the above equation and factorizing it with exponential terms, we can get

$$
\begin{aligned}
&-\frac{(Q_0-Q_D)(\lambda_C-\lambda_D)}{4\lambda_C\lambda_D r_1}\exp[\lambda_C(r_1-r_0)-\lambda_D r_0] \\
&+\frac{(Q_0-Q_D)(\lambda_C+\lambda_D)}{4\lambda_C\lambda_D r_1}\exp[\lambda_C(r_1-r_0)+\lambda_D r_0] \\
&-\frac{(Q_C-Q_D)(\lambda_C r_0+1)}{2\lambda_C r_1}\exp[\lambda_C(r_1-r_0)] \\
&+\frac{(Q_0-Q_D)(\lambda_C-\lambda_D)}{4\lambda_C\lambda_D r_1}\exp[-\lambda_C(r_1-r_0)+\lambda_D r_0] \\
&-\frac{(Q_0-Q_D)(\lambda_C+\lambda_D)}{4\lambda_C\lambda_D r_1}\exp[-\lambda_C(r_1-r_0)-\lambda_D r_0] \\
&-\frac{(Q_C-Q_D)(\lambda_C r_0-1)}{2\lambda_C r_1}\exp[-\lambda_C(r_1-r_0)] = \frac{-\lambda_D}{\lambda_C+\lambda_D}(Q_C-Q_D) \ .
\end{aligned} \tag{80}
$$

We can write the above equation with trigonometric functions as

$$
\begin{aligned}
&-\frac{(Q_0-Q_D)(\lambda_C-\lambda_D)}{2\lambda_C\lambda_D r_1}\sinh[\lambda_C(r_1-r_0)-\lambda_D r_0]\\
&+\frac{(Q_0-Q_D)(\lambda_C+\lambda_D)}{2\lambda_C\lambda_D r_1}\sinh[\lambda_C(r_1-r_0)+\lambda_D r_0]\\
&-\frac{(Q_C-Q_D)(\lambda_C r_0)}{\lambda_C r_1}\cosh[\lambda_C(r_1-r_0)]\\
&-\frac{(Q_C-Q_D)}{\lambda_C r_1}\sinh[\lambda_C(r_1-r_0)]=\frac{-\lambda_D}{\lambda_C+\lambda_D}(Q_C-Q_D)\ . \qquad (81)
\end{aligned}
$$

Using the relation Eq. (55) between $Q_0-Q_D$ and $Q_C-Q_D$, we can simplify Eq. (81) as

$$
\begin{aligned}
&-\frac{(\lambda_C-\lambda_D)\lambda_D r_0}{2(\lambda_C+\lambda_D)\lambda_D r_1\sinh(\lambda_D r_0)}\sinh[\lambda_C(r_1-r_0)-\lambda_D r_0]\\
&+\frac{\lambda_D r_0}{2\lambda_D r_1\sinh(\lambda_D r_0)}\sinh[\lambda_C(r_1-r_0)+\lambda_D r_0]\\
&-\frac{\lambda_C r_0}{\lambda_C r_1}\cosh[\lambda_C(r_1-r_0)]-\frac{1}{\lambda_C r_1}\sinh[\lambda_C(r_1-r_0)]=-\frac{\lambda_D}{\lambda_C+\lambda_D}\ . \qquad (82)
\end{aligned}
$$

After some algebraic calculations, we finally arrive at

$$
\begin{aligned}
&\frac{1}{2\lambda_D r_1}\frac{\lambda_D r_0}{\sinh(\lambda_D r_0)}\left\{\sinh[\lambda_C(r_1-r_0)+\lambda_D r_0]-\frac{\lambda_C-\lambda_D}{\lambda_C+\lambda_D}\sinh[\lambda_C(r_1-r_0)+\lambda_D r_0]\right\}\\
&-\frac{1}{\lambda_C r_1}\left\{\lambda_C r_0\cosh[\lambda_C(r_1-r_0)]+\sinh[\lambda_C(r_1-r_0)]\right\}+\frac{\lambda_D}{\lambda_C+\lambda_D}=0\ . \qquad (83)
\end{aligned}
$$

The boundary condition, Eq. (41) initially gives

$$
\begin{aligned}
&C_{51}\left(-\frac{\exp(\lambda_D R)}{\lambda_D R^2}+\frac{\exp(\lambda_D R)}{R}\right)+\\
&C_{52}\left(-\frac{\exp(-\lambda_D R)}{\lambda_D R^2}-\frac{\exp(-\lambda_D R)}{R}\right)=-\gamma_1-\gamma_2 Q(R)\ . \qquad (84)
\end{aligned}
$$

Plugging $Q(R)$ obtained from Eq. (75) into the above equation, we can factorize it with $C_{51}$ and $C_{52}$ as

$$
\begin{aligned}
&C_{51}\left(-\frac{\exp(\lambda_D R)}{\lambda_D R^2}+\frac{\exp(\lambda_D R)}{R}+\gamma_2\frac{\exp(\lambda_D R)}{\lambda_D R}\right)+\\
&C_{52}\left(-\frac{\exp(-\lambda_D R)}{\lambda_D R^2}-\frac{\exp(-\lambda_D R)}{R}+\gamma_2\frac{\exp(-\lambda_D R)}{\lambda_D R}\right)=-\gamma_1-\gamma_2 Q_D\ . \qquad (85)
\end{aligned}
$$

The above equation can be rewritten neatly by

$$[(\lambda_D+\gamma_2)R-1]\exp(\lambda_D R)C_{51}-[(\lambda_D-\gamma_2)R+1]\exp(-\lambda_D R)C_{52}$$
$$=(-\gamma_1-\gamma_2 Q_D)\lambda_D R^2\ . \tag{86}$$

We need to calculate $C_{51}$ and $C_{52}$ to write down the above equation completely. Plugging $C_{31}$ and $C_{32}$ into $C_{51}$ and $C_{52}$, we can obtain

$$\begin{aligned}
C_{51} &= \{\exp[(\lambda_C+\lambda_D)r_0](Q_0-Q_D)(\lambda_C/\lambda_D-1)/4\\
&\quad -\exp[(\lambda_C-\lambda_D)r_0](Q_0-Q_D)(\lambda_C/\lambda_D+1)/4\\
&\quad -\exp(\lambda_C r_0)(Q_C-Q_D)(\lambda_C r_0-1)/2\}\\
&\quad \times\exp[-(\lambda_C+\lambda_D)r_1](\lambda_D/\lambda_C-1)/2\\
&\quad +\{-\exp[-(\lambda_C+\lambda_D)r_0](Q_0-Q_D)(\lambda_C/\lambda_D-1)/4\\
&\quad +\exp[-(\lambda_C-\lambda_D)r_0](Q_0-Q_D)(\lambda_C/\lambda_D+1)/4\\
&\quad -\exp(-\lambda_C r_0)(Q_C-Q_D)(\lambda_C r_0+1)/2\}\\
&\quad \times\exp[(\lambda_C-\lambda_D)r_1](\lambda_D/\lambda_C+1)/2\\
&\quad +\exp(-\lambda_D r_1)(Q_C-Q_D)(\lambda_D r_1+1)/2,
\end{aligned} \tag{87}$$

$$\begin{aligned}
C_{52} &= \{-\exp[-(\lambda_C+\lambda_D)r_0](Q_0-Q_D)(\lambda_C/\lambda_D-1)/4\\
&\quad +\exp[-(\lambda_C-\lambda_D)r_0](Q_0-Q_D)(\lambda_C/\lambda_D+1)/4\\
&\quad -\exp(-\lambda_C r_0)(Q_C-Q_D)(\lambda_C r_0+1)/2\}\\
&\quad \times\exp[(\lambda_C+\lambda_D)r_1](\lambda_D/\lambda_C-1)/2\\
&\quad +\{\exp[(\lambda_C+\lambda_D)r_0](Q_0-Q_D)(\lambda_C/\lambda_D-1)/4\\
&\quad -\exp[(\lambda_C-\lambda_D)r_0](Q_0-Q_D)(\lambda_C/\lambda_D+1)/4\\
&\quad -\exp(\lambda_C r_0)(Q_C-Q_D)(\lambda_C r_0-1)/2\}\\
&\quad \times\exp[-(\lambda_C-\lambda_D)r_1](\lambda_D/\lambda_C+1)/2\\
&\quad +\exp(\lambda_D r_1)(Q_C-Q_D)(\lambda_D r_1-1)/2\ .
\end{aligned} \tag{88}$$

We can express the first term in the left hand side of Eq. (86) as

$$
\begin{aligned}
&[(\lambda_D+\gamma_2)R-1]\exp(\lambda_D R)C_{51} = \\
&-\frac{(Q_0-Q_D)(\lambda_C-\lambda_D)^2}{8\lambda_C\lambda_D}[(\lambda_D+\gamma_2)R-1]\exp[-\lambda_C(r_1-r_0)+\lambda_D(R+r_0-r_1)] \\
&+\frac{(Q_0-Q_D)(\lambda_C^2-\lambda_D^2)}{8\lambda_C\lambda_D}[(\lambda_D+\gamma_2)R-1]\exp[-\lambda_C(r_1-r_0)+\lambda_D(R-r_0-r_1)] \\
&+\frac{(Q_C-Q_D)(\lambda_C r_0-1)(\lambda_C-\lambda_D)}{4\lambda_C}[(\lambda_D+\gamma_2)R-1]\exp[-\lambda_C(r_1-r_0)+\lambda_D(R-r_1)] \\
&+\frac{(Q_0-Q_D)(\lambda_C+\lambda_D)^2}{8\lambda_C\lambda_D}[(\lambda_D+\gamma_2)R-1]\exp[\lambda_C(r_1-r_0)+\lambda_D(R+r_0-r_1)] \\
&-\frac{(Q_0-Q_D)(\lambda_C^2-\lambda_D^2)}{8\lambda_C\lambda_D}[(\lambda_D+\gamma_2)R-1]\exp[\lambda_C(r_1-r_0)+\lambda_D(R-r_0-r_1)] \\
&-\frac{(Q_C-Q_D)(\lambda_C r_0+1)(\lambda_C+\lambda_D)}{4\lambda_C}[(\lambda_D+\gamma_2)R-1]\exp[\lambda_C(r_1-r_0)+\lambda_D(R-r_1)] \\
&+(Q_C-Q_D)(\lambda_D r_1+1)[(\lambda_D+\gamma_2)R-1]\exp[\lambda_D(R-r_1)]/2 \\
&= [(\lambda_D+\gamma_2)R-1]\exp[\lambda_D(R-r_1)] \\
&\times\Bigg[-\frac{(Q_0-Q_D)(\lambda_C-\lambda_D)^2}{8\lambda_C\lambda_D}\exp[-\lambda_C(r_1-r_0)+\lambda_D r_0] \\
&+\frac{(Q_0-Q_D)(\lambda_C^2-\lambda_D^2)}{8\lambda_C\lambda_D}\exp[-\lambda_C(r_1-r_0)-\lambda_D r_0] \\
&+\frac{(Q_C-Q_D)(\lambda_C r_0-1)(\lambda_C-\lambda_D)}{4\lambda_C}\exp[-\lambda_C(r_1-r_0)] \\
&+\frac{(Q_0-Q_D)(\lambda_C+\lambda_D)^2}{8\lambda_C\lambda_D}\exp[\lambda_C(r_1-r_0)+\lambda_D r_0] \\
&-\frac{(Q_0-Q_D)(\lambda_C^2-\lambda_D^2)}{8\lambda_C\lambda_D}\exp[\lambda_C(r_1-r_0)-\lambda_D r_0] \\
&-\frac{(Q_C-Q_D)(\lambda_C r_0+1)(\lambda_C+\lambda_D)}{4\lambda_C}\exp[\lambda_C(r_1-r_0)]+\frac{(Q_C-Q_D)(\lambda_D r_1+1)}{2}\Bigg] \\
&= [(\lambda_D+\gamma_2)R-1]\exp[\lambda_D(R-r_1)] \\
&\times\Bigg[-\frac{(Q_0-Q_D)(\lambda_C^2-2\lambda_C\lambda_D+\lambda_D^2)}{8\lambda_C\lambda_D}\exp[-\lambda_C(r_1-r_0)+\lambda_D r_0] \\
&+\frac{(Q_0-Q_D)(\lambda_C^2-\lambda_D^2)}{8\lambda_C\lambda_D}\exp[-\lambda_C(r_1-r_0)-\lambda_D r_0] \\
&+\frac{(Q_C-Q_D)(\lambda_C r_0-1)(\lambda_C-\lambda_D)}{4\lambda_C}\exp[-\lambda_C(r_1-r_0)] \\
&+\frac{(Q_0-Q_D)(\lambda_C^2+2\lambda_C\lambda_D+\lambda_D^2)}{8\lambda_C\lambda_D}\exp[\lambda_C(r_1-r_0)+\lambda_D r_0] \\
&-\frac{(Q_0-Q_D)(\lambda_C^2-\lambda_D^2)}{8\lambda_C\lambda_D}\exp[\lambda_C(r_1-r_0)-\lambda_D r_0]
\end{aligned}
$$

$$
\begin{aligned}
&-\frac{(Q_C-Q_D)(\lambda_C r_0+1)(\lambda_C+\lambda_D)}{4\lambda_C}\exp[\lambda_C(r_1-r_0)]+\frac{(Q_C-Q_D)(\lambda_D r_1+1)}{2}\Big] \\
&=[(\lambda_D+\gamma_2)R-1]\exp[\lambda_D(R-r_1)] \\
&\times\Big[-\frac{(Q_0-Q_D)}{4\lambda_C\lambda_D}(\lambda_C^2\cosh[\lambda_C(r_1-r_0)-\lambda_D r_0]-\lambda_D^2\sinh[\lambda_C(r_1-r_0)-\lambda_D r_0] \\
&\qquad\qquad -\lambda_C\lambda_D\exp[-\lambda_C(r_1-r_0)+\lambda_D r_0]) \\
&+\frac{(Q_0-Q_D)}{4\lambda_C\lambda_D}(\lambda_C^2\cosh[\lambda_C(r_1-r_0)+\lambda_D r_0]+\lambda_D^2\sinh[\lambda_C(r_1-r_0)+\lambda_D r_0] \\
&\qquad\qquad +\lambda_C\lambda_D\exp[\lambda_C(r_1-r_0)+\lambda_D r_0]) \\
&-\frac{(Q_C-Q_D)}{4\lambda_C}(-(\lambda_C r_0-1)(\lambda_C-\lambda_D)\exp[-\lambda_C(r_1-r_0)] \\
&\qquad\qquad +(\lambda_C r_0+1)(\lambda_C+\lambda_D)\exp[\lambda_C(r_1-r_0)]) \\
&+\frac{(Q_C-Q_D)(\lambda_D r_1+1)}{2}\Big] \\
&=[(\lambda_D+\gamma_2)R-1]\exp[\lambda_D(R-r_1)] \\
&\times\Big[\frac{(Q_0-Q_D)}{4\lambda_C\lambda_D}\left\{\lambda_C^2(\cosh[\lambda_C(r_1-r_0)+\lambda_D r_0]-\cosh[\lambda_C(r_1-r_0)-\lambda_D r_0])\right. \\
&\qquad\qquad +\lambda_D^2(\sinh[\lambda_C(r_1-r_0)+\lambda_D r_0]+\sinh[\lambda_C(r_1-r_0)-\lambda_D r_0]) \\
&\qquad\qquad \left.+2\lambda_C\lambda_D\cosh[\lambda_C(r_1-r_0)]\exp(\lambda_D r_0)\right\} \\
&-\frac{Q_C-Q_D}{2\lambda_C}\left\{\lambda_C(\lambda_C r_0+\frac{\lambda_D}{\lambda_C})\sinh[\lambda_C(r_1-r_0)]+\lambda_C(\lambda_D r_0+1)\cosh[\lambda_C(r_1-r_0)]\right\} \\
&+\frac{(Q_C-Q_D)(\lambda_D r_1+1)}{2}\Big]\ .
\end{aligned}
$$

Using the relation $Q_0-Q_D=\frac{\lambda_C(Q_C-Q_D)}{\lambda_C+\lambda_D}\frac{\lambda_D r_0}{\sinh(\lambda_D r_0)}$, we can finally get

$$
\begin{aligned}
&[(\lambda_D+\gamma_2)R-1]\exp(\lambda_D R)C_{51} \\
&=(Q_C-Q_D)[(\lambda_D+\gamma_2)R-1]\exp[\lambda_D(R-r_1)]\times\Big[\frac{\lambda_C}{\lambda_C+\lambda_D}\frac{\lambda_D r_0}{\sinh(\lambda_D r_0)} \\
&\times\Big\{\frac{\lambda_C}{4\lambda_D}(\cosh[\lambda_C(r_1-r_0)+\lambda_D r_0]-\cosh[\lambda_C(r_1-r_0)-\lambda_D r_0]) \\
&\quad+\frac{\lambda_D}{4\lambda_C}(\sinh[\lambda_C(r_1-r_0)+\lambda_D r_0]+\sinh[\lambda_C(r_1-r_0)-\lambda_D r_0]) \\
&\quad+\frac{1}{2}\cosh[\lambda_C(r_1-r_0)]\exp(\lambda_D r_0)\Big\} \\
&-\frac{1}{2}(\lambda_C r_0+\frac{\lambda_D}{\lambda_C})\sinh[\lambda_C(r_1-r_0)]-\frac{1}{2}(\lambda_D r_0+1)\cosh[\lambda_C(r_1-r_0)]+\frac{(\lambda_D r_1+1)}{2}\Big]\ .
\end{aligned}
\tag{89}
$$

Similarly, the second term in the left hand side of Eq. (86) is calculated as

$$
\begin{aligned}
&-[(\lambda_D-\gamma_2)R+1]\exp(-\lambda_D R)C_{52} = \\
&-\frac{(Q_0-Q_D)(\lambda_C-\lambda_D)^2}{8\lambda_C\lambda_D}[(\lambda_D-\gamma_2)R+1]\exp[\lambda_C(r_1-r_0)-\lambda_D(R+r_0-r_1)] \\
&+\frac{(Q_0-Q_D)(\lambda_C^2-\lambda_D^2)}{8\lambda_C\lambda_D}[(\lambda_D-\gamma_2)R+1]\exp[\lambda_C(r_1-r_0)-\lambda_D(R-r_0-r_1)] \\
&-\frac{(Q_C-Q_D)(\lambda_C r_0+1)(\lambda_C-\lambda_D)}{4\lambda_C}[(\lambda_D-\gamma_2)R+1]\exp[\lambda_C(r_1-r_0)-\lambda_D(R-r_1)] \\
&+\frac{(Q_0-Q_D)(\lambda_C+\lambda_D)^2}{8\lambda_C\lambda_D}[(\lambda_D-\gamma_2)R+1]\exp[-\lambda_C(r_1-r_0)-\lambda_D(R+r_0-r_1)] \\
&-\frac{(Q_0-Q_D)(\lambda_C^2-\lambda_D^2)}{8\lambda_C\lambda_D}[(\lambda_D-\gamma_2)R+1]\exp[-\lambda_C(r_1-r_0)-\lambda_D(R-r_0-r_1)] \\
&+\frac{(Q_C-Q_D)(\lambda_C r_0-1)(\lambda_C+\lambda_D)}{4\lambda_C}[(\lambda_D-\gamma_2)R+1]\exp[-\lambda_C(r_1-r_0)-\lambda_D(R-r_1)] \\
&-(Q_C-Q_D)(\lambda_D r_1-1)[(\lambda_D-\gamma_2)R+1]\exp[-\lambda_D(R-r_1)]/2 \\
&= [(\lambda_D-\gamma_2)R+1]\exp[-\lambda_D(R-r_1)] \\
&\times\left[-\frac{(Q_0-Q_D)(\lambda_C-\lambda_D)^2}{8\lambda_C\lambda_D}\exp[\lambda_C(r_1-r_0)-\lambda_D r_0]\right. \\
&+\frac{(Q_0-Q_D)(\lambda_C^2-\lambda_D^2)}{8\lambda_C\lambda_D}\exp[\lambda_C(r_1-r_0)+\lambda_D r_0] \\
&-\frac{(Q_C-Q_D)(\lambda_C r_0+1)(\lambda_C-\lambda_D)}{4\lambda_C}\exp[\lambda_C(r_1-r_0)] \\
&+\frac{(Q_0-Q_D)(\lambda_C+\lambda_D)^2}{8\lambda_C\lambda_D}\exp[-\lambda_C(r_1-r_0)-\lambda_D r_0] \\
&-\frac{(Q_0-Q_D)(\lambda_C^2-\lambda_D^2)}{8\lambda_C\lambda_D}\exp[-\lambda_C(r_1-r_0)+\lambda_D r_0] \\
&\left.+\frac{(Q_C-Q_D)(\lambda_C r_0-1)(\lambda_C+\lambda_D)}{4\lambda_C}\exp[-\lambda_C(r_1-r_0)]-\frac{(Q_C-Q_D)(\lambda_D r_1-1)}{2}\right] \\
&= [(\lambda_D-\gamma_2)R+1]\exp[-\lambda_D(R-r_1)] \\
&\times\left[-\frac{(Q_0-Q_D)(\lambda_C^2-2\lambda_C\lambda_D+\lambda_D^2)}{8\lambda_C\lambda_D}\exp[\lambda_C(r_1-r_0)-\lambda_D r_0]\right. \\
&+\frac{(Q_0-Q_D)(\lambda_C^2-\lambda_D^2)}{8\lambda_C\lambda_D}\exp[\lambda_C(r_1-r_0)+\lambda_D r_0] \\
&-\frac{(Q_C-Q_D)(\lambda_C r_0+1)(\lambda_C-\lambda_D)}{4\lambda_C}\exp[\lambda_C(r_1-r_0)] \\
&+\frac{(Q_0-Q_D)(\lambda_C^2+2\lambda_C\lambda_D+\lambda_D^2)}{8\lambda_C\lambda_D}\exp[-\lambda_C(r_1-r_0)-\lambda_D r_0] \\
&\left.-\frac{(Q_0-Q_D)(\lambda_C^2-\lambda_D^2)}{8\lambda_C\lambda_D}\exp[-\lambda_C(r_1-r_0)+\lambda_D r_0]\right.
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{(Q_C-Q_D)(\lambda_C r_0-1)(\lambda_C+\lambda_D)}{4\lambda_C}\exp[-\lambda_C(r_1-r_0)]-\frac{(Q_C-Q_D)(\lambda_D r_1-1)}{2}\Bigg]\\
&=[(\lambda_D-\gamma_2)R+1]\exp[-\lambda_D(R-r_1)]\\
&\times\Bigg[-\frac{(Q_0-Q_D)}{4\lambda_C\lambda_D}(\lambda_C^2\cosh[\lambda_C(r_1-r_0)-\lambda_D r_0]+\lambda_D^2\sinh[\lambda_C(r_1-r_0)-\lambda_D r_0]\\
&\qquad\qquad-\lambda_C\lambda_D\exp[\lambda_C(r_1-r_0)-\lambda_D r_0])\\
&+\frac{(Q_0-Q_D)}{4\lambda_C\lambda_D}(\lambda_C^2\cosh[\lambda_C(r_1-r_0)+\lambda_D r_0]-\lambda_D^2\sinh[\lambda_C(r_1-r_0)+\lambda_D r_0]\\
&\qquad\qquad+\lambda_C\lambda_D\exp[-\lambda_C(r_1-r_0)-\lambda_D r_0])\\
&-\frac{(Q_C-Q_D)}{4\lambda_C}((\lambda_C r_0+1)(\lambda_C-\lambda_D)\exp[\lambda_C(r_1-r_0)]\\
&\qquad\qquad-(\lambda_C r_0-1)(\lambda_C+\lambda_D)\exp[-\lambda_C(r_1-r_0)])\\
&-\frac{(Q_C-Q_D)(\lambda_D r_1-1)}{2}\Bigg]\\
&=[(\lambda_D-\gamma_2)R+1]\exp[-\lambda_D(R-r_1)]\\
&\times\Bigg[\frac{(Q_0-Q_D)}{4\lambda_C\lambda_D}\{\lambda_C^2(\cosh[\lambda_C(r_1-r_0)+\lambda_D r_0]-\cosh[\lambda_C(r_1-r_0)-\lambda_D r_0])\\
&\qquad\qquad-\lambda_D^2(\sinh[\lambda_C(r_1-r_0)+\lambda_D r_0]+\sinh[\lambda_C(r_1-r_0)-\lambda_D r_0])\\
&\qquad\qquad+2\lambda_C\lambda_D\cosh[\lambda_C(r_1-r_0)]\exp(-\lambda_D r_0)\}\\
&-\frac{Q_C-Q_D}{2\lambda_C}\left\{\lambda_C(\lambda_C r_0-\frac{\lambda_D}{\lambda_C})\sinh[\lambda_C(r_1-r_0)]-\lambda_C(\lambda_D r_0-1)\cosh[\lambda_C(r_1-r_0)]\right\}\\
&-\frac{(Q_C-Q_D)(\lambda_D r_1-1)}{2}\Bigg]\ .
\end{aligned}
$$

Substituting the relation $Q_0-Q_D=\frac{\lambda_C(Q_C-Q_D)}{\lambda_C+\lambda_D}\frac{\lambda_D r_0}{\sinh(\lambda_D r_0)}$ into the above equation, we can finally get

$$
\begin{aligned}
&-[(\lambda_D-\gamma_2)R+1]\exp(-\lambda_D R)C_{52}\\
&=(Q_C-Q_D)[(\lambda_D-\gamma_2)R+1]\exp[-\lambda_D(R-r_1)]\times\Bigg[\frac{\lambda_C}{\lambda_C+\lambda_D}\frac{\lambda_D r_0}{\sinh(\lambda_D r_0)}\\
&\times\Bigg\{\frac{\lambda_C}{4\lambda_D}(\cosh[\lambda_C(r_1-r_0)+\lambda_D r_0]-\cosh[\lambda_C(r_1-r_0)-\lambda_D r_0])\\
&\quad-\frac{\lambda_D}{4\lambda_C}(\sinh[\lambda_C(r_1-r_0)+\lambda_D r_0]+\sinh[\lambda_C(r_1-r_0)-\lambda_D r_0])\\
&\quad+\frac{1}{2}\cosh[\lambda_C(r_1-r_0)]\exp(-\lambda_D r_0)\Bigg\}-\frac{1}{2}\left(\lambda_C r_0-\frac{\lambda_D}{\lambda_C}\right)\sinh[\lambda_C(r_1-r_0)]\\
&+\frac{1}{2}(\lambda_D r_0-1)\cosh[\lambda_C(r_1-r_0)]-\frac{(\lambda_D r_1-1)}{2}\Bigg]\ .
\end{aligned}
\tag{90}
$$

Plugging Eq. (89) and Eq. (90) into Eq. (86), we ultimately get another equation for $r_0$ and $r_1$:

$$\left(1+\frac{\gamma_2}{\lambda_D}-\frac{1}{\lambda_D R}\right)\left[\frac{\lambda_C}{\lambda_C+\lambda_D}\frac{\lambda_D r_0}{\sinh(\lambda_D r_0)}\left\{\frac{\lambda_C}{2\lambda_D}(\cosh[\lambda_C(r_1-r_0)+\lambda_D r_0]\right.\right.$$
$$-\cosh[\lambda_C(r_1-r_0)-\lambda_D r_0])+\frac{\lambda_D}{2\lambda_C}(\sinh[\lambda_C(r_1-r_0)+\lambda_D r_0]$$
$$\left.+\sinh[\lambda_C(r_1-r_0)-\lambda_D r_0])+\cosh[\lambda_C(r_1-r_0)]\exp[\lambda_D r_0]\right\}$$
$$-\left(\lambda_C r_0+\frac{\lambda_D}{\lambda_C}\right)\sinh[\lambda_C(r_1-r_0)]+(\lambda_D r_1+1)$$
$$\left.-(\lambda_D r_0+1)\cosh[\lambda_C(r_1-r_0)]\right]\exp[\lambda_D(R-r_1)]$$
$$+\left(1-\frac{\gamma_2}{\lambda_D}+\frac{1}{\lambda_D R}\right)\left[\frac{\lambda_C}{\lambda_C+\lambda_D}\frac{\lambda_D r_0}{\sinh(\lambda_D r_0)}\left\{\frac{\lambda_C}{2\lambda_D}(\cosh[\lambda_C(r_1-r_0)+\lambda_D r_0]\right.\right.$$
$$-\cosh[\lambda_C(r_1-r_0)-\lambda_D r_0])-\frac{\lambda_D}{2\lambda_C}(\sinh[\lambda_C(r_1-r_0)+\lambda_D r_0]$$
$$\left.+\sinh[\lambda_C(r_1-r_0)-\lambda_D r_0])+\cosh[\lambda_C(r_1-r_0)]\exp[-\lambda_D r_0]\right\}$$
$$-\left(\lambda_C r_0-\frac{\lambda_D}{\lambda_C}\right)\sinh[\lambda_C(r_1-r_0)]-(\lambda_D r_1-1)$$
$$\left.+(\lambda_D r_0-1)\cosh[\lambda_C(r_1-r_0)]\right]\exp[-\lambda_D(R-r_1)]+\frac{2R(\gamma_1+\gamma_2 Q_D)}{Q_C-Q_D}=0. \quad (91)$$

We get two equations, Eq. (83) and Eq. (91) by applying the boundary condition Eq. (41) and the condition $Q(r_1)=Q_m$. The unknown coefficients $r_0$ and $r_1$ can be obtained by solving Eq. (83) and Eq. (91) together.

*Profile VI.* The sixth possibility is that $Q(r)$ decreases and passes through $Q_m$ at $r=r_0$ $(0<r_0<R)$ such that $Q(r=r_0)=Q_m$ and passes again through $Q_m$ at $r=r_1$ $(r_0<r_1<R)$. $r_1$ can also be considered as the boundary between the dilute and the condensed phases of RNA molecules inside the capsid as $r_0$. A condensed phase changes into a dilute phase at $r_0$ and the dilute phase again changes into a condensed phase at $r_1$.

The solution to the Euler Lagrange Equation (40) can be written by

$$Q(r) = \begin{cases} (Q_0 - Q_C)\frac{\sinh(\lambda_C r)}{\lambda_C r} + Q_C & \text{for} \quad r < r_0 \\ C_{41}\frac{\exp(\lambda_D r)}{\lambda_D r} + C_{42}\frac{\exp(-\lambda_D r)}{\lambda_D r} + Q_D & \text{for} \quad r_0 < r < r_1 \\ C_{61}\frac{\exp(\lambda_C r)}{\lambda_C r} + C_{62}\frac{\exp(-\lambda_C r)}{\lambda_C r} + Q_C & \text{for} \quad r_1 < r < R \end{cases} \tag{92}$$

where $Q_0 = Q(0)$. The coefficients $C_{41}$ and $C_{42}$ is the same as $C_{41}$ and $C_{42}$ in *Profile IV* because we can apply the same continuities of $Q(r)$ and $Q'(r)$ at $r_0$, and follow the process of calculation of *Profile IV* with the same requirements $Q_0 = Q(0)$ and $Q(r_0) = Q_m$. There are still four unknown parameters $C_{61}$, $C_{62}$, $r_0$ and $r_1$. We can obtain these four unknowns with two continuities of the density profile at $r_1$, one requirement $Q(r_1) = Q_m$, and the boundary condition, Eq. (41).

Let's apply two continuity conditions 1) $Q(r_1^-) = Q(r_1^+)$ and 2) $Q'(r_1^-) = Q'(r_1^+)$. The first continuity condition gives

$$C_{41}\frac{\exp(\lambda_D r_1)}{\lambda_D r_1} + C_{42}\frac{\exp(-\lambda_D r_1)}{\lambda_D r_1} + Q_D = C_{61}\frac{\exp(\lambda_C r_1)}{\lambda_C r_1} + C_{62}\frac{\exp(-\lambda_C r_1)}{\lambda_C r_1} + Q_C. \tag{93}$$

The second continuity condition gives

$$\begin{aligned} &C_{41}\left(-\frac{\exp(\lambda_D r_1)}{\lambda_D r_1^2} + \frac{\exp(\lambda_D r_1)}{r_1}\right) + C_{42}\left(-\frac{\exp(-\lambda_D r_1)}{\lambda_D r_1^2} - \frac{\exp(-\lambda_D r_1)}{r_1}\right) = \\ &C_{61}\left(-\frac{\exp(\lambda_C r_1)}{\lambda_C r_1^2} + \frac{\exp(\lambda_C r_1)}{r_1}\right) + C_{62}\left(-\frac{\exp(-\lambda_C r_1)}{\lambda_C r_1^2} - \frac{\exp(-\lambda_C r_1)}{r_1}\right). \end{aligned} \tag{94}$$

Combining Eq. (93) and Eq. (94), we can solve for $C_{61}$ and $C_{62}$:

$$\begin{aligned} C_{61} &= C_{42}\exp[-(\lambda_C + \lambda_D)r_1](\lambda_C/\lambda_D - 1)/2 \\ &\quad + C_{41}\exp[-(\lambda_C - \lambda_D)r_1](\lambda_C/\lambda_D + 1)/2 \\ &\quad - \exp(-\lambda_C r_1)(Q_C - Q_D)(\lambda_C r_1 + 1)/2, \\ C_{62} &= C_{41}\exp[(\lambda_C + \lambda_D)r_1](\lambda_C/\lambda_D - 1)/2 \\ &\quad + C_{42}\exp[(\lambda_C - \lambda_D)r_1](\lambda_C/\lambda_D + 1)/2 \\ &\quad - \exp(-\lambda_C r_1)(Q_C - Q_D)(\lambda_C r_1 - 1)/2, \end{aligned}$$

where

$$
\begin{aligned}
C_{41} &= -\exp[-(\lambda_C+\lambda_D)r_0](Q_0-Q_C)(\lambda_D/\lambda_C-1)/4 \\
&\quad +\exp[(\lambda_C-\lambda_D)r_0](Q_0-Q_C)(\lambda_D/\lambda_C+1)/4 \\
&\quad +\exp(-\lambda_D r_0)(Q_C-Q_D)(\lambda_D r_0+1)/2, \\
C_{42} &= \exp[(\lambda_C+\lambda_D)r_0](Q_0-Q_C)(\lambda_D/\lambda_C-1)/4 \\
&\quad -\exp[-(\lambda_C-\lambda_D)r_0](Q_0-Q_C)(\lambda_D/\lambda_C+1)/4 \\
&\quad +\exp(\lambda_D r_0)(Q_C-Q_D)(\lambda_D r_0-1)/2.
\end{aligned}
$$

There are still two unknowns $r_0$ and $r_1$ in the density profile. The condition $Q(r_1)=Q_m$ gives

$$
Q_m = C_{41}\frac{\exp(\lambda_D r_1)}{\lambda_D r_1} + C_{42}\frac{\exp(-\lambda_D r_1)}{\lambda_D r_1} + Q_D \ . \tag{95}
$$

Using the relation $Q_m - Q_C = -\frac{\lambda_D(Q_C-Q_D)}{\lambda_C+\lambda_D}$, Eq. (95) can be ordered

$$
C_{41}\frac{\exp(\lambda_D r_1)}{\lambda_D r_1} + C_{42}\frac{\exp(-\lambda_D r_1)}{\lambda_D r_1} = \frac{\lambda_C}{\lambda_C+\lambda_D}(Q_C-Q_D) \ . \tag{96}
$$

Plugging $C_{41}$ and $C_{42}$ into the above equation and factorizing it with exponential terms, we can get

$$
\begin{aligned}
&\frac{(Q_0-Q_C)(\lambda_C-\lambda_D)}{4\lambda_C\lambda_D r_1}\exp[\lambda_D(r_1-r_0)-\lambda_C r_0] \\
&+\frac{(Q_0-Q_C)(\lambda_C+\lambda_D)}{4\lambda_C\lambda_D r_1}\exp[\lambda_D(r_1-r_0)+\lambda_C r_0] \\
&+\frac{(Q_C-Q_D)(\lambda_D r_0+1)}{2\lambda_D r_1}\exp[\lambda_D(r_1-r_0)] \\
&-\frac{(Q_0-Q_C)(\lambda_C-\lambda_D)}{4\lambda_C\lambda_D r_1}\exp[-\lambda_D(r_1-r_0)+\lambda_C r_0] \\
&-\frac{(Q_0-Q_C)(\lambda_C+\lambda_D)}{4\lambda_C\lambda_D r_1}\exp[-\lambda_D(r_1-r_0)-\lambda_C r_0] \\
&+\frac{(Q_C-Q_D)(\lambda_D r_0-1)}{2\lambda_D r_1}\exp[-\lambda_D(r_1-r_0)] = \frac{\lambda_C}{\lambda_C+\lambda_D}(Q_C-Q_D) \ .
\end{aligned} \tag{97}
$$

We can write the above equation with trigonometric functions as

$$\begin{aligned}
&\frac{(Q_0-Q_C)(\lambda_C-\lambda_D)}{2\lambda_C\lambda_D r_1}\sinh[\lambda_D(r_1-r_0)-\lambda_C r_0]\\
&+\frac{(Q_0-Q_C)(\lambda_C+\lambda_D)}{2\lambda_C\lambda_D r_1}\sinh[\lambda_D(r_1-r_0)+\lambda_C r_0]\\
&+\frac{(Q_C-Q_D)(\lambda_D r_0)}{\lambda_D r_1}\cosh[\lambda_D(r_1-r_0)]\\
&+\frac{(Q_C-Q_D)}{\lambda_D r_1}\sinh[\lambda_D(r_1-r_0)]=\frac{\lambda_C}{\lambda_C+\lambda_D}(Q_C-Q_D)\ . \qquad (98)
\end{aligned}$$

Using the relation Eq. (67) between $Q_0-Q_C$ and $Q_C-D_D$, we can simplify Eq. (98) as

$$\begin{aligned}
&\frac{(\lambda_C-\lambda_D)\lambda_C r_0}{2(\lambda_C+\lambda_D)\lambda_C r_1\sinh(\lambda_C r_0)}\sinh[\lambda_D(r_1-r_0)-\lambda_C r_0]\\
&+\frac{\lambda_C r_0}{2\lambda_C r_1\sinh(\lambda_C r_0)}\sinh[\lambda_D(r_1-r_0)+\lambda_C r_0]\\
&-\frac{\lambda_D r_0}{\lambda_D r_1}\cosh[\lambda_D(r_1-r_0)]-\frac{1}{\lambda_D r_1}\sinh[\lambda_D(r_1-r_0)]=-\frac{\lambda_C}{\lambda_C+\lambda_D}\ . \qquad (99)
\end{aligned}$$

After some algebraic calculations, we finally arrive at

$$\begin{aligned}
&\frac{1}{2\lambda_C r_1}\frac{\lambda_C r_0}{\sinh(\lambda_C r_0)}\left\{\sinh[\lambda_D(r_1-r_0)+\lambda_C r_0]+\frac{\lambda_C-\lambda_D}{\lambda_C+\lambda_D}\sinh[\lambda_D(r_1-r_0)+\lambda_C r_0]\right\}\\
&-\frac{1}{\lambda_D r_1}\left\{\lambda_D r_0\cosh[\lambda_D(r_1-r_0)]+\sinh[\lambda_D(r_1-r_0)]\right\}+\frac{\lambda_C}{\lambda_C+\lambda_D}=0\ . \qquad (100)
\end{aligned}$$

The boundary condition, Eq. (41) initially gives

$$\begin{aligned}
&C_{61}\left(-\frac{\exp(\lambda_C R)}{\lambda_C R^2}+\frac{\exp(\lambda_C R)}{R}\right)+\\
&C_{62}\left(-\frac{\exp(-\lambda_C R)}{\lambda_C R^2}-\frac{\exp(-\lambda_C R)}{R}\right)=-\gamma_1-\gamma_2 Q(R)\ . \qquad (101)
\end{aligned}$$

Plugging $Q(R)$ obtained from Eq. (92) into the above equation, we can factorize it with $C_{61}$ and $C_{62}$ as

$$\begin{aligned}
&C_{61}\left(-\frac{\exp(\lambda_C R)}{\lambda_C R^2}+\frac{\exp(\lambda_C R)}{R}+\gamma_2\frac{\exp(\lambda_C R)}{\lambda_C R}\right)+\\
&C_{62}\left(-\frac{\exp(-\lambda_C R)}{\lambda_C R^2}-\frac{\exp(-\lambda_C R)}{R}+\gamma_2\frac{\exp(-\lambda_C R)}{\lambda_C R}\right)=-\gamma_1-\gamma_2 Q_C\ . \qquad (102)
\end{aligned}$$

The above equation can be rewritten neatly by

$$[(\lambda_C+\gamma_2)R-1]\exp(\lambda_C R)C_{61}-[(\lambda_C-\gamma_2)R+1]\exp(-\lambda_C R)C_{62}$$
$$=(-\gamma_1-\gamma_2 Q_C)\lambda_C R^2\ . \tag{103}$$

We need to calculate $C_{61}$ and $C_{62}$ to write down the above equation completely. Plugging $C_{41}$ and $C_{42}$ into $C_{61}$, we can obtain

$$\begin{aligned}
C_{61} &= \{\exp[(\lambda_C+\lambda_D)r_0](Q_0-Q_C)(\lambda_D/\lambda_C-1)/4\\
&\quad -\exp[-(\lambda_C-\lambda_D)r_0](Q_0-Q_C)(\lambda_D/\lambda_C+1)/4\\
&\quad +\exp(\lambda_D r_0)(Q_C-Q_D)(\lambda_D r_0-1)/2\}\\
&\quad \times\exp[-(\lambda_C+\lambda_D)r_1](\lambda_C/\lambda_D-1)/2\\
&\quad +\{-\exp[-(\lambda_C+\lambda_D)r_0](Q_0-Q_C)(\lambda_D/\lambda_C-1)/4\\
&\quad +\exp[(\lambda_C-\lambda_D)r_0](Q_0-Q_C)(\lambda_D/\lambda_C+1)/4\\
&\quad +\exp(-\lambda_D r_0)(Q_C-Q_D)(\lambda_D r_0+1)/2\}\\
&\quad \times\exp[-(\lambda_C-\lambda_D)r_1](\lambda_C/\lambda_D+1)/2\\
&\quad -\exp(-\lambda_C r_1)(Q_C-Q_D)(\lambda_C r_1+1)/2,
\end{aligned} \tag{104}$$

$$\begin{aligned}
C_{62} &= \{-\exp[-(\lambda_C+\lambda_D)r_0](Q_0-Q_C)(\lambda_D/\lambda_C-1)/4\\
&\quad +\exp[(\lambda_C-\lambda_D)r_0](Q_0-Q_C)(\lambda_D/\lambda_C+1)/4\\
&\quad +\exp(-\lambda_D r_0)(Q_C-Q_D)(\lambda_D r_0+1)/2\}\\
&\quad \times\exp[(\lambda_C+\lambda_D)r_1](\lambda_C/\lambda_D-1)/2\\
&\quad +\{\exp[(\lambda_C+\lambda_D)r_0](Q_0-Q_C)(\lambda_D/\lambda_C-1)/4\\
&\quad -\exp[-(\lambda_C-\lambda_D)r_0](Q_0-Q_C)(\lambda_D/\lambda_C+1)/4\\
&\quad +\exp(\lambda_D r_0)(Q_C-Q_D)(\lambda_D r_0-1)/2\}\\
&\quad \times\exp[(\lambda_C-\lambda_D)r_1](\lambda_C/\lambda_D+1)/2\\
&\quad -\exp(\lambda_C r_1)(Q_C-Q_D)(\lambda_C r_1-1)/2\ .
\end{aligned} \tag{105}$$

We can express the first term in the left hand side of Eq. (103) as

$$
\begin{aligned}
&[(\lambda_C+\gamma_2)R-1]\exp(\lambda_C R)C_{61} = \\
&-\frac{(Q_0-Q_C)(\lambda_C-\lambda_D)^2}{8\lambda_C\lambda_D}[(\lambda_C+\gamma_2)R-1]\exp[-\lambda_D(r_1-r_0)+\lambda_C(R+r_0-r_1)] \\
&-\frac{(Q_0-Q_C)(\lambda_C^2-\lambda_D^2)}{8\lambda_C\lambda_D}[(\lambda_C+\gamma_2)R-1]\exp[-\lambda_D(r_1-r_0)+\lambda_C(R-r_0-r_1)] \\
&+\frac{(Q_C-Q_D)(\lambda_D r_0-1)(\lambda_C-\lambda_D)}{4\lambda_D}[(\lambda_C+\gamma_2)R-1]\exp[-\lambda_D(r_1-r_0)+\lambda_C(R-r_1)] \\
&+\frac{(Q_0-Q_C)(\lambda_C+\lambda_D)^2}{8\lambda_C\lambda_D}[(\lambda_C+\gamma_2)R-1]\exp[\lambda_D(r_1-r_0)+\lambda_C(R+r_0-r_1)] \\
&+\frac{(Q_0-Q_C)(\lambda_C^2-\lambda_D^2)}{8\lambda_C\lambda_D}[(\lambda_C+\gamma_2)R-1]\exp[\lambda_D(r_1-r_0)+\lambda_C(R-r_0-r_1)] \\
&+\frac{(Q_C-Q_D)(\lambda_D r_0+1)(\lambda_C+\lambda_D)}{4\lambda_D}[(\lambda_C+\gamma_2)R-1]\exp[\lambda_D(r_1-r_0)+\lambda_C(R-r_1)] \\
&-(Q_C-Q_D)(\lambda_C r_1+1)[(\lambda_C+\gamma_2)R-1]\exp[\lambda_C(R-r_1)]/2 \\
&= [(\lambda_C+\gamma_2)R-1]\exp[\lambda_C(R-r_1)] \\
&\times\left[-\frac{(Q_0-Q_C)(\lambda_C-\lambda_D)^2}{8\lambda_C\lambda_D}\exp[-\lambda_D(r_1-r_0)+\lambda_C r_0]\right. \\
&-\frac{(Q_0-Q_C)(\lambda_C^2-\lambda_D^2)}{8\lambda_C\lambda_D}\exp[-\lambda_D(r_1-r_0)-\lambda_C r_0] \\
&+\frac{(Q_C-Q_D)(\lambda_D r_0-1)(\lambda_C-\lambda_D)}{4\lambda_D}\exp[-\lambda_D(r_1-r_0)] \\
&+\frac{(Q_0-Q_C)(\lambda_C+\lambda_D)^2}{8\lambda_C\lambda_D}\exp[\lambda_D(r_1-r_0)+\lambda_C r_0] \\
&+\frac{(Q_0-Q_C)(\lambda_C^2-\lambda_D^2)}{8\lambda_C\lambda_D}\exp[\lambda_D(r_1-r_0)-\lambda_C r_0] \\
&\left.+\frac{(Q_C-Q_D)(\lambda_D r_0+1)(\lambda_C+\lambda_D)}{4\lambda_D}\exp[\lambda_D(r_1-r_0)]-\frac{(Q_C-Q_D)(\lambda_C r_1+1)}{2}\right] \\
&= [(\lambda_C+\gamma_2)R-1]\exp[\lambda_C(R-r_1)] \\
&\times\left[-\frac{(Q_0-Q_C)(\lambda_C^2-2\lambda_C\lambda_D+\lambda_D^2)}{8\lambda_C\lambda_D}\exp[-\lambda_D(r_1-r_0)+\lambda_C r_0]\right. \\
&-\frac{(Q_0-Q_C)(\lambda_C^2-\lambda_D^2)}{8\lambda_C\lambda_D}\exp[-\lambda_D(r_1-r_0)-\lambda_C r_0] \\
&+\frac{(Q_C-Q_D)(\lambda_D r_0-1)(\lambda_C-\lambda_D)}{4\lambda_D}\exp[-\lambda_D(r_1-r_0)] \\
&+\frac{(Q_0-Q_C)(\lambda_C^2+2\lambda_C\lambda_D+\lambda_D^2)}{8\lambda_C\lambda_D}\exp[\lambda_D(r_1-r_0)+\lambda_C r_0] \\
&\left.+\frac{(Q_0-Q_C)(\lambda_C^2-\lambda_D^2)}{8\lambda_C\lambda_D}\exp[\lambda_D(r_1-r_0)-\lambda_C r_0]\right.
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{(Q_C-Q_D)(\lambda_D r_0+1)(\lambda_C+\lambda_D)}{4\lambda_D}\exp[\lambda_D(r_1-r_0)]-\frac{(Q_C-Q_D)(\lambda_C r_1+1)}{2}\Bigg]\\
&=[(\lambda_C+\gamma_2)R-1]\exp[\lambda_C(R-r_1)]\\
&\times\Bigg[-\frac{(Q_0-Q_C)}{4\lambda_C\lambda_D}(\lambda_D^2\cosh[\lambda_D(r_1-r_0)-\lambda_C r_0]-\lambda_C^2\sinh[\lambda_D(r_1-r_0)-\lambda_C r_0]\\
&\qquad\qquad-\lambda_C\lambda_D\exp[-\lambda_D(r_1-r_0)+\lambda_C r_0])\\
&+\frac{(Q_0-Q_C)}{4\lambda_C\lambda_D}(\lambda_D^2\cosh[\lambda_D(r_1-r_0)+\lambda_C r_0]+\lambda_C^2\sinh[\lambda_D(r_1-r_0)+\lambda_C r_0]\\
&\qquad\qquad+\lambda_C\lambda_D\exp[\lambda_D(r_1-r_0)+\lambda_C r_0])\\
&+\frac{(Q_C-Q_D)}{4\lambda_D}((\lambda_D r_0-1)(\lambda_C-\lambda_D)\exp[-\lambda_D(r_1-r_0)]\\
&\qquad\qquad+(\lambda_D r_0+1)(\lambda_C+\lambda_D)\exp[\lambda_D(r_1-r_0)])\\
&-\frac{(Q_C-Q_D)(\lambda_C r_1+1)}{2}\Bigg]\\
&=[(\lambda_C+\gamma_2)R-1]\exp[\lambda_C(R-r_1)]\\
&\times\Bigg[\frac{(Q_0-Q_C)}{4\lambda_C\lambda_D}\left\{\lambda_D^2(\cosh[\lambda_D(r_1-r_0)+\lambda_C r_0]-\cosh[\lambda_D(r_1-r_0)-\lambda_C r_0])\right.\\
&\qquad\qquad+\lambda_C^2(\sinh[\lambda_D(r_1-r_0)+\lambda_C r_0]+\sinh[\lambda_D(r_1-r_0)-\lambda_C r_0])\\
&\qquad\qquad\left.+2\lambda_C\lambda_D\cosh[\lambda_D(r_1-r_0)]\exp(\lambda_C r_0)\right\}\\
&+\frac{Q_C-Q_D}{2\lambda_D}\left\{\lambda_D(\lambda_D r_0+\frac{\lambda_C}{\lambda_D})\sinh[\lambda_D(r_1-r_0)]+\lambda_D(\lambda_C r_0+1)\cosh[\lambda_D(r_1-r_0)]\right\}\\
&-\frac{(Q_C-Q_D)(\lambda_C r_1+1)}{2}\Bigg]\ .
\end{aligned}
$$

Using the relation $Q_0-Q_C=\frac{-\lambda_D(Q_C-Q_D)}{\lambda_C+\lambda_D}\frac{\lambda_C r_0}{\sinh(\lambda_C r_0)}$, we can finally get

$$
\begin{aligned}
&[(\lambda_C+\gamma_2)R-1]\exp(\lambda_C R)C_{61}\\
&=-(Q_C-Q_D)[(\lambda_C+\gamma_2)R-1]\exp[\lambda_C(R-r_1)]\times\Bigg[\frac{\lambda_D}{\lambda_C+\lambda_D}\frac{\lambda_C r_0}{\sinh(\lambda_C r_0)}\\
&\times\left\{\frac{\lambda_D}{4\lambda_C}(\cosh[\lambda_D(r_1-r_0)+\lambda_C r_0]-\cosh[\lambda_D(r_1-r_0)-\lambda_C r_0])\right.\\
&\quad+\frac{\lambda_C}{4\lambda_D}(\sinh[\lambda_D(r_1-r_0)+\lambda_C r_0]+\sinh[\lambda_D(r_1-r_0)-\lambda_C r_0])\\
&\quad\left.+\frac{1}{2}\cosh[\lambda_D(r_1-r_0)]\exp(\lambda_C r_0)\right\}-\frac{1}{2}(\lambda_D r_0+\frac{\lambda_C}{\lambda_D})\sinh[\lambda_D(r_1-r_0)]\\
&-\frac{1}{2}(\lambda_C r_0+1)\cosh[\lambda_D(r_1-r_0)]+\frac{(\lambda_C r_1+1)}{2}\Bigg]\ .
\end{aligned}
\tag{106}
$$

Similarly, the second term in the left hand side of Eq. (103) is calculated as

$$
\begin{aligned}
&-[(\lambda_C-\gamma_2)R+1]\exp(-\lambda_C R)C_{62} = \\
&-\frac{(Q_0-Q_C)(\lambda_C-\lambda_D)^2}{8\lambda_C\lambda_D}[(\lambda_C-\gamma_2)R+1]\exp[\lambda_D(r_1-r_0)-\lambda_C(R+r_0-r_1)] \\
&-\frac{(Q_0-Q_C)(\lambda_C^2-\lambda_D^2)}{8\lambda_C\lambda_D}[(\lambda_C-\gamma_2)R+1]\exp[\lambda_D(r_1-r_0)-\lambda_C(R-r_0-r_1)] \\
&-\frac{(Q_C-Q_D)(\lambda_D r_0+1)(\lambda_C-\lambda_D)}{4\lambda_D}[(\lambda_C-\gamma_2)R+1]\exp[\lambda_D(r_1-r_0)-\lambda_C(R-r_1)] \\
&+\frac{(Q_0-Q_C)(\lambda_C+\lambda_D)^2}{8\lambda_C\lambda_D}[(\lambda_C-\gamma_2)R+1]\exp[-\lambda_D(r_1-r_0)-\lambda_C(R+r_0-r_1)] \\
&+\frac{(Q_0-Q_C)(\lambda_C^2-\lambda_D^2)}{8\lambda_C\lambda_D}[(\lambda_C-\gamma_2)R+1]\exp[-\lambda_D(r_1-r_0)-\lambda_C(R-r_0-r_1)] \\
&-\frac{(Q_C-Q_D)(\lambda_D r_0-1)(\lambda_C+\lambda_D)}{4\lambda_D}[(\lambda_C-\gamma_2)R+1]\exp[-\lambda_D(r_1-r_0)-\lambda_C(R-r_1)] \\
&+(Q_C-Q_D)(\lambda_C r_1-1)[(\lambda_C-\gamma_2)R+1]\exp[-\lambda_C(R-r_1)]/2 \\
&= [(\lambda_C-\gamma_2)R+1]\exp[-\lambda_C(R-r_1)] \\
&\times\left[-\frac{(Q_0-Q_C)(\lambda_C-\lambda_D)^2}{8\lambda_C\lambda_D}\exp[\lambda_D(r_1-r_0)-\lambda_C r_0]\right. \\
&-\frac{(Q_0-Q_C)(\lambda_C^2-\lambda_D^2)}{8\lambda_C\lambda_D}\exp[\lambda_D(r_1-r_0)+\lambda_C r_0] \\
&-\frac{(Q_C-Q_C)(\lambda_D r_0+1)(\lambda_C-\lambda_D)}{4\lambda_D}\exp[\lambda_D(r_1-r_0)] \\
&+\frac{(Q_0-Q_C)(\lambda_C+\lambda_D)^2}{8\lambda_C\lambda_D}\exp[-\lambda_D(r_1-r_0)-\lambda_C r_0] \\
&+\frac{(Q_0-Q_C)(\lambda_C^2-\lambda_D^2)}{8\lambda_C\lambda_D}\exp[-\lambda_D(r_1-r_0)+\lambda_C r_0] \\
&\left.-\frac{(Q_C-Q_D)(\lambda_D r_0-1)(\lambda_C+\lambda_D)}{4\lambda_D}\exp[-\lambda_D(r_1-r_0)]+\frac{(Q_C-Q_D)(\lambda_C r_1-1)}{2}\right] \\
&= [(\lambda_C-\gamma_2)R+1]\exp[-\lambda_C(R-r_1)] \\
&\times\left[-\frac{(Q_0-Q_C)(\lambda_C^2-2\lambda_C\lambda_D+\lambda_D^2)}{8\lambda_C\lambda_D}\exp[\lambda_D(r_1-r_0)-\lambda_C r_0]\right. \\
&-\frac{(Q_0-Q_C)(\lambda_C^2-\lambda_D^2)}{8\lambda_C\lambda_D}\exp[\lambda_D(r_1-r_0)+\lambda_C r_0] \\
&-\frac{(Q_C-Q_D)(\lambda_D r_0+1)(\lambda_C-\lambda_D)}{4\lambda_D}\exp[\lambda_D(r_1-r_0)] \\
&+\frac{(Q_0-Q_C)(\lambda_C^2+2\lambda_C\lambda_D+\lambda_D^2)}{8\lambda_C\lambda_D}\exp[-\lambda_D(r_1-r_0)-\lambda_C r_0] \\
&\left.+\frac{(Q_0-Q_C)(\lambda_C^2-\lambda_D^2)}{8\lambda_C\lambda_D}\exp[-\lambda_D(r_1-r_0)+\lambda_C r_0]\right.
\end{aligned}
$$

$$
\begin{aligned}
&-\frac{(Q_C-Q_D)(\lambda_D r_0-1)(\lambda_C+\lambda_D)}{4\lambda_D}\exp[-\lambda_D(r_1-r_0)]+\frac{(Q_C-Q_D)(\lambda_C r_1-1)}{2}\Big] \\
&=[(\lambda_C-\gamma_2)R+1]\exp[-\lambda_C(R-r_1)] \\
&\times\Big[-\frac{(Q_0-Q_C)}{4\lambda_C\lambda_D}(\lambda_D^2\cosh[\lambda_D(r_1-r_0)-\lambda_C r_0]+\lambda_C^2\sinh[\lambda_D(r_1-r_0)-\lambda_C r_0] \\
&\qquad\qquad -\lambda_C\lambda_D\exp[\lambda_D(r_1-r_0)-\lambda_C r_0]) \\
&+\frac{(Q_0-Q_C)}{4\lambda_C\lambda_D}(\lambda_D^2\cosh[\lambda_D(r_1-r_0)+\lambda_C r_0]-\lambda_C^2\sinh[\lambda_D(r_1-r_0)+\lambda_C r_0] \\
&\qquad\qquad +\lambda_C\lambda_D\exp[-\lambda_D(r_1-r_0)-\lambda_C r_0]) \\
&+\frac{(Q_C-Q_D)}{4\lambda_D}(-(\lambda_D r_0+1)(\lambda_C-\lambda_D)\exp[\lambda_D(r_1-r_0)] \\
&\qquad\qquad -(\lambda_D r_0-1)(\lambda_C+\lambda_D)\exp[-\lambda_D(r_1-r_0)]) \\
&+\frac{(Q_C-Q_D)(\lambda_C r_1-1)}{2}\Big] \\
&=[(\lambda_C-\gamma_2)R+1]\exp[-\lambda_C(R-r_1)] \\
&\times\Big[\frac{(Q_0-Q_C)}{4\lambda_C\lambda_D}\{\lambda_D^2(\cosh[\lambda_D(r_1-r_0)+\lambda_C r_0]-\cosh[\lambda_D(r_1-r_0)-\lambda_C r_0]) \\
&\qquad\qquad -\lambda_C^2(\sinh[\lambda_D(r_1-r_0)+\lambda_C r_0]+\sinh[\lambda_D(r_1-r_0)-\lambda_C r_0]) \\
&\qquad\qquad +2\lambda_C\lambda_D\cosh[\lambda_D(r_1-r_0)]\exp(-\lambda_C r_0)\} \\
&+\frac{Q_C-Q_D}{2\lambda_D}\left\{\lambda_D(\lambda_D r_0-\frac{\lambda_C}{\lambda_D})\sinh[\lambda_D(r_1-r_0)]-\lambda_D(\lambda_C r_0-1)\cosh[\lambda_D(r_1-r_0)]\right\} \\
&+\frac{(Q_C-Q_D)(\lambda_C r_1-1)}{2}\Big]\ .
\end{aligned}
$$

Substituting the relation $Q_0-Q_C=\frac{-\lambda_D(Q_C-Q_D)}{\lambda_C+\lambda_D}\frac{\lambda_C r_0}{\sinh(\lambda_C r_0)}$ into the above equation, we can finally get

$$
\begin{aligned}
&-[(\lambda_C-\gamma_2)R+1]\exp(-\lambda_C R)C_{62} \\
&=-(Q_C-Q_D)[(\lambda_C-\gamma_2)R+1]\exp[-\lambda_C(R-r_1)]\times\Big[\frac{\lambda_D}{\lambda_C+\lambda_D}\frac{\lambda_C r_0}{\sinh(\lambda_C r_0)} \\
&\times\Big\{\frac{\lambda_D}{4\lambda_C}(\cosh[\lambda_D(r_1-r_0)+\lambda_C r_0]-\cosh[\lambda_D(r_1-r_0)-\lambda_C r_0]) \\
&\quad -\frac{\lambda_C}{4\lambda_D}(\sinh[\lambda_D(r_1-r_0)+\lambda_C r_0]+\sinh[\lambda_D(r_1-r_0)-\lambda_C r_0]) \\
&\quad +\frac{1}{2}\cosh[\lambda_D(r_1-r_0)]\exp(-\lambda_C r_0)\Big\}-\frac{1}{2}\left(\lambda_D r_0-\frac{\lambda_C}{\lambda_D}\right)\sinh[\lambda_D(r_1-r_0)] \\
&+\frac{1}{2}(\lambda_C r_0-1)\cosh[\lambda_D(r_1-r_0)]-\frac{(\lambda_C r_1-1)}{2}\Big]\ . \qquad (107)
\end{aligned}
$$

Plugging Eq. (106) and Eq. (107) into Eq. (103), we ultimately get another equation for $r_0$ and $r_1$:

$$\begin{aligned}
&\left(1+\frac{\gamma_2}{\lambda_C}-\frac{1}{\lambda_C R}\right)\left[\frac{\lambda_D}{\lambda_C+\lambda_D}\frac{\lambda_C r_0}{\sinh(\lambda_C r_0)}\left\{\frac{\lambda_D}{2\lambda_C}(\cosh[\lambda_D(r_1-r_0)+\lambda_C r_0]\right.\right.\\
&-\cosh[\lambda_D(r_1-r_0)-\lambda_C r_0])+\frac{\lambda_C}{2\lambda_D}(\sinh[\lambda_D(r_1-r_0)+\lambda_C r_0]\\
&\left.+\sinh[\lambda_D(r_1-r_0)-\lambda_C r_0])+\cosh[\lambda_D(r_1-r_0)]\exp[\lambda_C r_0]\right\}\\
&-\left(\lambda_D r_0+\frac{\lambda_C}{\lambda_D}\right)\sinh[\lambda_D(r_1-r_0)]+(\lambda_C r_1+1)\\
&\left.-(\lambda_C r_0+1)\cosh[\lambda_D(r_1-r_0)]\right]\exp[\lambda_C(R-r_1)]\\
&+\left(1-\frac{\gamma_2}{\lambda_C}+\frac{1}{\lambda_C R}\right)\left[\frac{\lambda_D}{\lambda_C+\lambda_D}\frac{\lambda_C r_0}{\sinh(\lambda_C r_0)}\left\{\frac{\lambda_D}{2\lambda_C}(\cosh[\lambda_D(r_1-r_0)+\lambda_C r_0]\right.\right.\\
&-\cosh[\lambda_D(r_1-r_0)-\lambda_C r_0])-\frac{\lambda_C}{2\lambda_D}(\sinh[\lambda_D(r_1-r_0)+\lambda_C r_0]\\
&\left.+\sinh[\lambda_D(r_1-r_0)-\lambda_C r_0])+\cosh[\lambda_D(r_1-r_0)]\exp[-\lambda_C r_0]\right\}\\
&-\left(\lambda_D r_0-\frac{\lambda_C}{\lambda_D}\right)\sinh[\lambda_D(r_1-r_0)]-(\lambda_C r_1-1)\\
&\left.+(\lambda_C r_0-1)\cosh[\lambda_D(r_1-r_0)]\right]\exp[-\lambda_C(R-r_1)]-\frac{2R(\gamma_1+\gamma_2 Q_C)}{Q_C-Q_D}=0.
\end{aligned} \tag{108}$$

We get two equations, Eq. (100) and Eq. (108) by applying the boundary condition Eq. (41) and the condition $Q(r_1)=Q_m$. The unknown coefficients $r_0$ and $r_1$ can be obtained by solving Eq. (100) and Eq. (108) together.

## *5.4 Discussion*

Obtaining an analytical solution for $r_0$ from Eq. (60) is a highly non-trivial task and numerical solution is generally needed. Nevertheless, we can understand important qualitative features of the RNA concentration profile by solving for $r_0$ in the limit of strong capsid RNA adsorption ($\lambda_s R \gg 1$) and small correlation length of RNA concentrated phase ($\lambda_C R \gg 1$). In this limit, the first two terms in Eq. (60) are the

two most dominant ones. Balancing them, we get $u \simeq 2\lambda_s R$, or

$$r_0 \simeq R - {\lambda_C}^{-1} \ln(2\lambda_s R). \tag{109}$$

As we mentioned above, $r_0$ can be considered as the boundary between a dense RNA phase near the capsid and a dilute RNA phase at the capsid center. The quantity $d = R - r_0$, therefore, can be considered the thickness of this dense RNA layer. According Eq. (109), $d \propto \ln R$ which is *parametrically smaller* than the capsid radius, $R$ [1]. In other words, our RNA concentration profile shows a dense RNA layer condensed on the inner capsid with thickness which varies *very slowly* with its radius. Consequently, the amount of RNA packaged inside the virus is proportional to the capsid *area* (or the number of capsid proteins) instead of its volume. In recent works [14, 72], a similar dependence is observed when positively charged amino acids of capsid proteins are located in their long flexible peptide arms. In their works, the thickness of RNA molecules (treated as *linear* polymers) layer depends on the length of these arms. On the other hand, for the class of viruses we study in this chapter where the basic amino acids are located at the inner capsid surface instead of peptide arms, the competition between the *branching* degree of freedom of the secondary structure of RNA molecules and the attraction of capsid proteins is responsible for the layer structure and the thickness scales as $\ln R$. Another interesting feature of RNA concentration profile III is the fact that it does not peak at the inner capsid radius $R$ but at some smaller radius. This is the direct consequence of the boundary condition, Eq. (41) which forces the RNA concentration to decrease in the vicinity of the capsid.

In Figure 19, we plot examples of the two profiles, Eq. (48) and Eq. (51), fitted to the experimental data for two viruses, the Dengue virus and bacteriophage $MS_2$. The

[1] The $\ln R$ dependence of $d$ is also obtained for the wetting layer on the surface of a colloid [70]. This is to be expected because Eq. (37) can be mapped onto the Cahn theory of wetting transition[112]. Here we show that the negative curvature of the inner viral capsid apparently does not significantly affect this logarithmic dependency.

data for the Dengue virus was obtained using cryoelectron microscopy [85]. The data for bacteriophage $MS_2$ was obtained using small angle neutron scattering measurements [75]. Both viruses have most of their basic amino acids located on the surface of inner capsid, therefore our model capsid can be used. Both theoretical profiles show reasonable agreement with experiment results.

So far, when solving the Euler-Lagrange equation for RNA density profile, we assume $Q(r)$ crosses the value $Q_m$ at most two times. Certainly, there is a possibility that $Q(r)$ can cross $Q_m$ multiple times as $r$ increases from zero to $R$. This results in an oscillating RNA concentration profile. One could easily extend our calculation presented in this chapter to such a case by adding more piecewise solution to the ansatz, Eq. (51), and requiring $Q(r)$ and its derivative to be continuous at the crossing points. *Profile V* and *Profile VI* are the extensions of *Profile III* and *Profile IV* respectively. Such extension could offer insights, for e.g., into the oscillating radial profile of RNA molecules packaged inside Turnip Yellow Mosaic Virus (TYMV) [75]. Nevertheless, these cases are relatively uncommon and the calculations would go beyond the scope of this chapter. We will address these cases in more detail in future study.

Naturally, one wants to know which RNA concentration profile is the most thermodynamically stable. To answer this question, one needs to substitute these profiles (Eq. (48) and Eq. (51)) into the original expression for the capsid excess free energy, Eq. (38), and compare the resulting energies. This is a tedious task. Numerically, it is found that for small adsorption strength of viral capsid, the second profile would be thermodynamically stable and RNA concentration is maximum at the capsid center. For stronger surface adsorption, the third profile is lower in energy. In this case, RNA molecules form a dense layer at the capsid and the RNA concentration is maximum at a finite radius smaller than $R$.

It is known [112] that the mean-field theory, Eq. (38), breaks down when the critical point is approached and the first order transition between dilute and condensed phases of RNA solution becomes of second order. Once this happens, a physical picture similar to that of a solution of branched polymer with *frozen* branching arrangement emerges [31]. In this case, the RNA molecules become unscreened and non-overlapped. For viruses with several packaged RNA molecules, each of them would adsorb independently onto the capsid and the layer thickness of each molecule scales as square root of its molecular weight. Conversely, if such separation between constituent viral genomes is observed, it would signal the breakdown of mean-field theory.

In conclusion, in this chapter we found two different nucleotide concentration profiles of viral RNA molecules packaged inside spherical viruses. The theory applies to a class of viruses where capsid-RNA interaction occurs at the capsid surface only. For small interaction strength, the RNA monomer concentration is maximized at the center of the capsid to maximize their configurational entropy [164]. For higher interaction strength, RNA forms a dense layer near the capsid surface. The thickness of this layer is a slowly varying (logarithmic) function of the inner capsid radius. In this case, the amount of packaged RNA would be proportional to capsid area (or number of capsid proteins) instead of its volume. The profiles describe reasonably well the experimental profiles for various viruses.

# CHAPTER VI

# RNA CONDENSATION IN THE PRESENCE OF A SINGLE NUCLEOCAPSID PROTEIN

## *6.1 Introduction*

In case of HIV-1, reverse transcription, the process that ssRNA transcribes double stranded proviral DNA, which is integrated into the host chromosome at later steps, has become a drug target for curing AIDS because this process is an essential step in the viral life cycle and there exist several possible ways to block reverse transcription. NC protein plays many important roles in reverse transcription so that there are extensive studies of the role of NC protein of retroviruses [92, 101]. The NC proteins of retroviruses which have been known to be highly basic nucleic acid-binding polypeptide containing less than 100 amino acids are produced as a result of proteolytic cleaveage of the Gag polyprotein, which plays an important role in virus assembly [63, 149, 157]. In addition, NC protein has one or two weakly interacting zinc fingers of the form CCHC metal binding motif, flanked by flexible N-terminal and C-terminal regions and , for example, NC protein of HIV-1 has only 55 amino acid residues and two zinc fingers [16, 55, 32, 126, 94]. Despite their small size, NC protein has been known to have essential functions in many steps during retroviral life cycle, such as genomic RNA packaging [125, 17], reverse transcription [94, 30, 126, 150], tRNA primer annealing [25, 26, 129, 45, 61, 62, 121], viral assembly [3, 20, 21, 110], and integration [146, 147, 148].

Reverse transcriptase (RT) tends to pause at specific sites such as stem loops when DNA is replicated from the viral genomic RNA *in vitro* [102, 40, 36, 82, 7, 163]. Adding NC to reverse transcription *in vitro* reduces the pausing of reverse

transcription presumably by destabilizing the secondary structures of RNA [163, 77, 143] and greatly increases the efficiency of strand transfer reactions by HIV-1 NC RT [33, 5, 116, 80, 130, 131]. The extensive studies of the role of NC protein in reverse transcription also show that NC has nucleic acid chaperone activity which catalyzes the rearrangement of a nucleic acid molecule into conformation that is thermodynamically more stable [126, 94, 32, 150, 10, 64, 97] i.e the maximum number of base pairs by lowering the energy barrier for breaking and reforming the base pairs. The two main functions of NC protein responsible for the nucleic acid chaperone activity are the aggregation of nucleic acids [23, 103, 142, 158] and destabilization of nucleic acid helixes [8, 12, 13, 18, 79, 154, 160, 161] associated with the cationic N-terminal domain and zinc fingers, respectively.

Different shapes of NC proteins from several retroviruses have the different binding strength of NC leading to the different pattern of RNA aggregation quantitatively and qualitatively( specially HTLV-1 NC ) [141]. The aim of this chapter is to get a profile of how RNA is radially distributed around a single NC protein using a polymer physics as an initial foundation to understand NC's aggregating ability to nucleic acids, one of the chaperone functions. There are extensive efforts in experiments, but to our best knowledge, there is no theoretical work to describe RNA - NC protein interaction depending on the binding strength of NC protein. We employed a mean field approximation for theory of RNA condensation suggested before [112, 90]. Our results demonstrated that RNA has a screening effect for the strong binding strength of NC.

## *6.2 RNA condensation around a single NC protein*

It is known that the secondary structure of ssRNA molecules is generated by the different possible complementary pairing arrangements of the bases [67]. It is assumed that ssRNA molecules are highly flexible branching polymers and fluctuate freely over

all possible branching configurations. Different branching configurations are described in the schematic way shown in Figure 20, characterized by fugacities for "bifunctional units", "trifunctional units", and "end points". We assume good solvent conditions with repulsive interactions between the different units and with no "tertiary" pairing. To investigate the interaction between RNA molecules and NC protein, one can start with the generalized Hilhorst model [68, 69] which was studied by Lubensky and Isaacson [98]. The Hamiltonian for annealed, branched polymers in this model is an extension of the $n$-vector model and is defined based on a field theoretic formalism in terms of a field variable $\psi_{ij}$ as

$$H = \int d^3r \left\{ \frac{1}{2}(\nabla\psi)^2 + \frac{\epsilon}{2}\psi^2 + u\psi^4 - \frac{w}{6}\sum_j \psi_{1j}^3 - h\sum_j \psi_{1j} \right\} , \quad (110)$$

where $\psi^2 = \sum_{i=1}^{n}\sum_{j=1}^{m}\psi_{ij}^2$ with $i$ = 1,..., $n$ and $j$ = 1,..., $m$. In the $n \to 0$ limit, the Hamiltonian is related to the partition function $Z = \lim_{n\to 0} \mathrm{Tr}\exp(-H)$. Here, $\epsilon, w, h$, and $m$ are the fugacity of monomers (linear sequences), trifunctional units (branching points), the end points (hairpins or stem loops), and a complete polymer, respectively. The coefficient $u$ corresponds the effect of repulsive interactions between monomers and is positive assuming that RNA is present in a good solvent. Using a mean field approximation, one set $\psi_{ij}(\vec{r}) = \delta_{i,1}Q(\vec{r})$ where $\vec{r}$ is a radial vector in a spherical coordinate. The free energy density of RNA branching polymers in solution by a mean field approximation [112] based on the expression of the annealed, branched polymer of the generalized Hilhorst model [98] can be written as

$$\frac{W[Q(\vec{r})]}{m} = \frac{\epsilon}{2}Q(\vec{r})^2 - \frac{w}{6}Q(\vec{r})^3 + muQ(\vec{r})^4 - hQ(\vec{r}) \ . \quad (111)$$

Here, $Q(\vec{r})$ is the order parameter of the field theory used to analyze the amount RNA branching polymers in solution. $Q(\vec{r})$ corresponds to the concentration of the end points, $Q(\vec{r})^2$ to the concentration of the monomers, and $Q(\vec{r})^3$ to the concentration of the branching points. $Q(\vec{r})^4$ represents the interactions between two monomers. The expression for the branching polymer Eq. (111) is reduced to the well known

expression for the free energy of linear polymer when branching points are disappeared ($w = 0$) [35]. The mean field expression Eq. (111) is used to study the behavior of RNA condensation in solution in the presence of a single NC protein. As a simplified model, we think of a single NC protein as an infinite well located at the origin to trap RNA molecules and write down the interaction term $H_0$ between a single NC protein and RNA polymers as

$$H_0 = -\gamma Q(0). \tag{112}$$

$H_0$ comes from the direct contact between RNA and a NC protein at origin. $\gamma$ is the adsorption strength (binding strength) and is positive. The negative $H_0$ tells that RNA - NC protein interaction is attractive. $Q(0)$ is proportional to end points density at the origin and corresponds to an infinite well with $\gamma$ being positive. It is assumed that RNA is condensed radially around the NC protein located at the origin so that $Q(\vec{r}) \equiv Q(r)$ where $r$ is the radial distance from the origin. The mean field excess free energy of a whole RNA branching polymer in solution can be written as

$$H_{MF} = H_0 + \int_0^\infty 4\pi r^2 dr \left\{ \frac{m}{2} \left( \frac{dQ}{dr} \right)^2 + \Delta W[Q] \right\}, \tag{113}$$

where $\Delta W[Q(r)] = W[Q(r)] - W[Q_{bulk}]$. This expression is a gradient square functional of the order parameter. The integral formula in the right hand side of Eq. (113) comes from the distortions of RNA polymer distribution. The free energy density $W(Q)$ has two minima at $Q_D$ and $Q_C$ due to the cubic term of $Q(\vec{r})$ in Eq. (111). $Q_D$ and $Q_C$ are the mean field order parameter of a dilute bulk RNA solution and a condensed bulk RNA solution respectively. A first order transition from a dilute to a condensed solution happens when $W(Q_D)$ is equal to $W(Q_C)$. It is assumed that RNA bulk solution lies between these two phases so that a dilute phase and a condensed phase of RNA solution have a similar energy and $Q_{bulk} = Q_D$. The minimization of $H_{MF}$ Eq. (113) respect to $Q(r)$ leads to the Euler-Lagrange equation

$$\frac{d^2Q}{dr^2} + \frac{2}{r}\frac{dQ}{dr} - \frac{1}{m}\frac{\partial \Delta W}{\partial Q} = 0, \tag{114}$$

together with two boundary conditions,

$$\left.\frac{dQ}{dr}\right|_{r=0} = \frac{-\gamma}{4\pi r^2 m}, \tag{115}$$

$$\left.\frac{dQ}{dr}\right|_{r\to\infty} = 0. \tag{116}$$

The detail derivation for Euler Lagrange equation and boundary condition is shown in Appendix C.2. $\Delta W(Q)$ can be approximated as two double parabola potential form [70] because $W(Q)$ has two minima at $Q_D$ and $Q_C$ :

$$\Delta W(Q) = \begin{cases} \frac{1}{2}m\lambda_D^2(Q-Q_D)^2 & \text{for } Q < Q_m \\ \frac{1}{2}m\lambda_C^2(Q-Q_C)^2 & \text{for } Q > Q_m \end{cases} \tag{117}$$

where $Q_m = (\lambda_D Q_D + \lambda_C Q_C)/(\lambda_D + \lambda_C)$ is the point where the two parabolas cross each other forming a cusp. The two coefficients $\lambda_D^2$, $\lambda_C^2$ are the stiffness of the free energy density of a dilute and a condensed RNA solution, respectively. A dilute RNA solution is easier to be compressed than a condensed one so that $\lambda_C^2 > \lambda_D^2$. This approximate double parabolic potential form is not valid near critical temperature where the first order transition becomes second order. This approximation also breaks down when the fugacity of branching points $w$ approaches to 0 where a branching polymer becomes a linear polymer. However, the mean field approximation for the free energy density Eq. (111) becomes invalid before this limit is approached. Keeping this limit of mean field theory, this is a good approximation. This double parabola approximation was also mentioned in the Cahn wetting theory [34].

## *6.3 Results and Discussion*

One can solve Euler-Lagrange Equation (114) easily with the approximate potential form $\Delta W(Q)$ of Eq. (117) because Eq. (114) becomes linear. The general solution to Eq. (114) is a linear combination of $exp(\pm\lambda_{D,C}r)/r$ and we have only one case of RNA concentration profile. $Q(r)$ starts from a condensed RNA solution at the origin and

passes through $Q_m$ at some distance $r = r_0$ $(0 < r < \infty)$ such that $Q(r = r_0) = Q_m$. $r_0$ indicates the boundary distance between a condensed and a dilute phase of RNA solution. The density profile can be set up as a linear combination of $exp(\pm\lambda_{D,C}r)/r$:

$$Q(r) = \begin{cases} C_1\frac{e^{\lambda_C r}}{\lambda_C r} + C_2\frac{e^{-\lambda_C r}}{\lambda_C r} + Q_C & \text{for } r < r_0 \\ C_3\frac{e^{\lambda_D r}}{\lambda_D r} + C_4\frac{e^{-\lambda_D r}}{\lambda_D r} + Q_D & \text{for } r_0 < r \end{cases} \tag{118}$$

To find five unknowns ($C_1$, $C_2$, $C_3$, $C_4$ and $r_0$) in Eq. (118), let's apply the boundary conditions one by one. From a boundary condition Eq. (116), we know $Q(r = \infty)$ should be finite, leading to $C_3 = 0$. $Q'(r)$ in a condensed phase of RNA molecules can be derived from $Q(r)$ of Eq. (118).

$$\frac{dQ}{dr} = C_1\left(-\frac{e^{\lambda_C r}}{\lambda_C r^2} + \lambda_C\frac{e^{\lambda_C r}}{\lambda_C r}\right) + C_2\left(-\frac{e^{-\lambda_C r}}{\lambda_C r^2} - \lambda_C\frac{e^{-\lambda_C r}}{\lambda_C r}\right) \tag{119}$$

The boundary condition Eq. (115) gives the following equation :

$$\left.\frac{dQ}{dr}\right|_{r\to 0} = \lim_{r\to 0}\left(\frac{-C_1 - C_2}{\lambda_C r^2} + \lambda_C\frac{C_1 - C_2}{\lambda_C r}\right) = -\frac{\gamma}{4\pi m r^2} \tag{120}$$

Here, $\frac{-C_1-C_2}{\lambda_C r^2}$ is the most dominant term as $r$ approaches to 0. Therefore, we can get $C_2 = \frac{\gamma\lambda_C}{4\pi m} - C_1$. The RNA density profile can be written by

$$Q(r) = \begin{cases} C_1\frac{\exp(\lambda_C r)}{\lambda_C r} + (\frac{\gamma\lambda_C}{4\pi m} - C_1)\frac{\exp(-\lambda_C r)}{\lambda_C r} + Q_C & \text{for } \; r < r_0 \\ C_4\frac{\exp(-\lambda_D r)}{\lambda_D r} + Q_D & \text{for } \; r_0 < r \end{cases} \tag{121}$$

The boundary condition between a condensed and a dilute phase, $Q(r = r_0) = Q_m$ , where $Q_m = \frac{\lambda_C Q_C + \lambda_D Q_D}{\lambda_C + \lambda_D}$ , gives

$$C_4 = \lambda_D r_0 e^{\lambda_D r_0}\frac{\lambda_C}{\lambda_C + \lambda_D}(Q_C - Q_D) \tag{122}$$

Using two continuity conditions for $Q(r)$ and $Q'(r)$ at the boundary $r = r_0$ with two boundary conditions Eq. (115) and Eq. (116), we can obtain the following equations:

$$C_1\frac{\exp(\lambda_C r_0)}{\lambda_C r_0} + \left(\frac{\gamma}{4\pi m}\lambda_C - C_1\right)\frac{\exp(-\lambda_C r_0)}{\lambda_C r_0} + Q_C = C_4\frac{\exp(-\lambda_D r_0)}{\lambda_D r_0} + Q_D \tag{123}$$

and

$$C_1\left(-\frac{\exp(\lambda_C r_0)}{\lambda_C r_0^2}+\frac{\exp(\lambda_C r_0)}{r_0}\right)+\left(\frac{\gamma}{4\pi m}\lambda_C-C_1\right)\left(-\frac{\exp(-\lambda_C r_0)}{\lambda_C r_0^2}-\frac{\exp(-\lambda_C r_0)}{r_0}\right)$$
$$=C_4\left(-\frac{\exp(-\lambda_D r_0)}{\lambda_D r_0^2}-\frac{\exp(-\lambda_D r_0)}{r_0}\right)\ . \tag{124}$$

Combining Eq. (122) and Eq. (123), we can get

$$C_1 \;=\; \frac{1}{2\sinh(\lambda_C r_0)}\left(-\frac{\gamma\lambda_C}{4\pi m}\exp(-\lambda_C r_0)-\frac{\lambda_C}{\lambda_C+\lambda_D}\lambda_D r_0(Q_C-Q_D)\right)\ . \tag{125}$$

Plugging Eq. (122) and Eq. (125) into Eq. (124), we arrive at the equation for the last unknown $r_0$:

$$\frac{-\gamma\lambda_C}{4\pi m\sinh(\lambda_C r_0)}=(Q_C-Q_D)\left\{-1+\lambda_D r_0\left(1+\frac{\lambda_D}{\lambda_C}\right)^{-1}\frac{\exp(-\lambda_C r_0)}{\sinh(\lambda_C r_0)}\right\}. \tag{126}$$

The equation for $r_0$ from Eq. (126) cannot be solved analytically and $r_0$ can be obtained numerically. Eq. (126) can be rewritten by

$$\begin{aligned} r_0 \;&=\; \frac{1}{\lambda_C}\left\{\ln\frac{\gamma\lambda_C}{4\pi m(Q_C-Q_D)}\right. \\ &+\; \left.\ln\left(1+\sqrt{1+\left(2\lambda_D r_0\left(1+\frac{\lambda_D}{\lambda_C}\right)^{-1}+1\right)\left(\frac{4\pi m(Q_C-Q_D)}{\gamma\lambda_C}\right)^2}\right)\right\} \end{aligned} \tag{127}$$

To understand the qualitative behavior of RNA condensation by NC protein from Eq. (127), we can consider two limiting cases.

*Case 1 (weak adsorption strength)* : When $4\pi m(Q_C-Q_D) >> \gamma\lambda_C$, $\lambda_D r_0$ is much less than 1. From Eq. (127), we obtain

$$r_0 \simeq \frac{\lambda_C+\lambda_D}{\lambda_C}\frac{\gamma}{4\pi m(Q_C-Q_D)}. \tag{128}$$

In Eq. (128), the boundary distance of RNA condensation $r_0$ is also a small value because of the weak adsorption strength of NC protein so that the RNA condensation region is confined in a small region. $r_0$ is linearly proportional to $\gamma$ due to this small condensation volume.

*Case 2 (strong adsorption strength)* : When $4\pi m(Q_C - Q_D) << \gamma\lambda_C$, we obtain

$$r_0 \simeq \frac{1}{\lambda_C} \ln\left(\frac{2\gamma\lambda_C}{4\pi m(Q_C - Q_D)}\right). \tag{129}$$

The radius of RNA condensation from the origin changes very slowly with $\gamma$. Condensed RNAs screen the NC protein so that the single NC protein at the origin interacts with condensed RNA in the vicinity of the origin more strongly than condensed RNA located farther from the origin. The logarithmic dependence of Eq. (129) can be interpreted as a result of screening effect of condensed RNA molecules.

The effect of different retroviral NC proteins on mini-TAR RNA aggregation has been studied [141]. They compared the chaperone activities of NC proteins from four different retroviruses : HIV-1, Moloney murine leukemia virus (MLV), Rous sarcoma virus (RSV), and human T-cell lymphotropic virus type 1 (HTLV-1). The results showed that all NC proteins, except for HTLV-1 NC, aggregate nucleic acid upon saturated binding at $\sim 1\mu M$ of NC. RSV NC's aggregating ability to nucleic acids is more effective than HIV-1 NC. MLV NC appears to be slightly more effective at aggregating nucleic acids than HIV-1 NC at concentrations of $\geq 1\mu M$ of NC. A bulk NC's case shows some saturation levels of $\sim 1\mu M$. It has been known that NC binds to $6 \pm 1$ nucleotides at saturation [165, 32, 126, 150, 154, 38, 87, 46, 153]. Therefore, RNA aggregation saturates at some concentration of NC. In this chapter, we have not calculated the aggregation behavior of several NC proteins with nucleic acids and will work on this problem in the future. However, RNA condensation theory explains a single NC's aggregation ability to nucleic acids as the basics to understand the interactions between many NC proteins and RNA molecules depending on the binding strengths of different NC proteins. Interestingly, HTLV-1 NC was not able to show its aggregating ability unlike HIV-1 NC, RSV NC, and MLV NC. HTLV-1 NC is neutral at physiological condition (pH $\sim$ 7). As I mentioned above, the adsorption strength $\gamma$ is positive so that the interaction between RNA and NC is attractive. In Eq. (126), there is no solution if $\gamma$ is 0 or negative. Eq. (126) tells that RNA

condensation does not happen if a single NC protein is neutral or shows repulsive interaction with a negatively charged RNA. This behavior corresponds to the bulk non aggregating behavior of HTLV-1 NC.

To approach to the solution to the bulk aggregating ability of different NC proteins, one can use the solution to RNA condensation profile from a single NC protein obtained from this chapter. The spherical symmetry was used to study RNA aggregation from a single NC protein and was broken if we added another NC protein into our system to study the interaction between two NC proteins and RNA. Therefore, the spherical symmetry does not validate any more and electrostatic interaction between NC proteins appears to be significantly important. The problem became a little more difficult, but it can be solvable. In this way, the aggregating ability of NC proteins to nucleic acids can be studied in near future.

It is known that RNA condensation theory tends to predict the RNA condensation profiles in spherical viruses [112, 90] well, consistent with experiments [85, 75]. In the previous paper [90], we showed that the thickness of RNA condensation layer from inside capsid of RNA spherical viruses is slowly varying (logarithmic) function of the inner capsid radius and is parametrically smaller than the capsid radius for strong adsorption strength of the inner capsid. This phenomenon results from the competition between the branching of RNA secondary structure and its adsorption to the inner capsid. Similarly, in the single NC problem, the thickness of the RNA condensation layer from the origin is also logarithmic function of $\frac{\gamma}{4\pi m(Q_C - Q_D)}$ for strong adsorption case (Case 2). Note that $\frac{\gamma}{4\pi m(Q_C - Q_D)}$ has a dimension of length and is much larger than 1 with a fixed constant $\lambda_C$. This means that the thickness of RNA condensation layer is also parametrically smaller than the length defined by $\frac{\gamma}{4\pi m(Q_C - Q_D)}$ for Case 2. The difference is that the thickness of RNA condensation layer is measured from the origin for the single NC problem and the thickness of that is measured from the inner capsid surface for the spherical virus problem.

It is known [112] that the mean field theory, Eq (111), breaks down when the critical point is approached and the first-order transition between dilute and condensed phases of RNA solution becomes of second order. Once this happens, a physical picture similar to that of a solution of branched polymer with frozen branching arrangement emerges [31] and the RNA molecules become unscreened and nonoverlapped. In a single NC problem, each RNA molecule adsorb independently onto the NC at the origin and the layer thickness of each molecule scales as a square root of its molecular weight. In other words, if such separation between RNA molecules is noticed, the mean field approximation breaks down.

In summary, we have found the RNA concentration profile interacting with a single NC protein. RNA condensation theory based on field theoretic formalism is applied to deal with the interaction between a single NC protein and RNA for weak and strong adsorption cases. For weak adsorption strength of NC protein, the radial RNA condensation distance is linearly proportional to the adsorption strength $\gamma$. For strong adsorption strength, the RNA condensation distance is varying very slowly (logarithmically) with $\gamma$. This results from the screening of NC protein due to condensed RNA molecules. In addition, the RNA condensation theory is consistent with the inability to condense RNA in the presence of neutral protein like HTLV-1 NC protein. Our work is the fundamental progress to study from a single NC protein - RNA interaction to several NC proteins - RNA interaction.

# CHAPTER VII

# CONCLUSIONS

In this thesis we study five problems of bacteriophage and HIV. A lot of research about bacteriophages has been done since they were discovered because of their simple structure and scientific curiosity. Studying HIV, the causative agent of AIDS, also draws attention because of a strong desire to cure AIDS.

The first topic (in Chapter 2) is the DNA genome translocation problem of bacteriophage in the presence of $MgSO_4$ salt. DNA enters into the cell when the receptor of a cell membrane recognizes the tail of the bacteriophage *in vivo*. What drives ejection of DNA is the big pressure difference between inside and outside the capsid. *In vitro*, many experiments on force-balancing have been done. Among those, we search for interesting physical phenomena in the experiment of DNA ejection from phage in the presence of a divalent counterion such as $Mg^{+2}$. Experimental studies showed that a monovalent salt has a negligible effect on the ejection of DNA from bacteriophage [43]. However, a divalent counterion ($MgSO_4$) was shown to exert a strong effect leading to inhibition of DNA ejection non-monotonically [42]. Nguyen et al [57, 113] confirmed that negative DNA can be overcharged with positive multivalent counterions so that the net charge of DNA changes sign from negative and to positive, so called *charge inversion*. To predict the ejection percentage of DNA from the phage, we included the electrostatic energy of DNA in solution, derived by charge inversion theory, in the total system. We finally obtained several results. 1) We find the neutral concentration of DNA to be 64 mM, where the least DNA is ejected. 2) The mediated attractive DNA-DNA interaction energy per persistence length is about $k_BT$ so that DNA condensation does not happen in free solution. 3)

The non monotonic effect comes from charge inversion of DNA and a short range attraction between DNA inside the capsid. 4) The maximum inhibition shows 80% of the total genome length experimentally and 70% at 64 mM in our work theoretically. 5) Negative or positive DNA gives more pulling force from the capsid of the phage by reducing its electrostatic self-energy. We get these results by setting up DNA-DNA mediated interactions by divalent counterions. Indeed, DNA condensation by divalent ions has been observed in 2 dimensional system with DNA on bilipid layers [84]. Actually, the radius of gyration is much bigger than the bacteriophage capsid size, so DNA is highly confined inside capsid, and we point out that the dimension of DNA configuration inside the capsid is less than 2 dimension where DNA-DNA interaction is attractive. We also observed the value of DNA-DNA interaction energy mediated by $Mg^{+2}$ counterions. Even though DNA-DNA interaction energy is obtained by fitting using our phenomenological theory, it does give a good starting point for future research of DNA condensation in the presence of divalent counterions.

The second problem we discuss (in Chapter 3) is how divalent counterions affect DNA-DNA electrostatic interaction in a hexagonally oriented DNA bundle. In Chapter 2, we suggest a phenomenological theory to fit the experimental data of DNA ejection from bacteriophage in the presence of divalent counterions such as $Mg^{+2}$. One interesting feature is its non-monotonic influence on DNA ejection. In simulation, we model the hexagonally-packed DNA bundle as a number of DNA molecules arranged in parallel, and the strength of DNA-DNA interaction is calculated with a periodic boundary condition. We obtain DNA-DNA effective electrostatic interaction by calculating osmotic pressure of a DNA bundle using the Expanded Ensemble method [99, 59]. We show that divalent counterions can induce DNA reentrant condensation similar to that caused by tri- or tetra-valent counterions when DNA configuration entropy is restricted. This phenomenon of DNA reentrant condensation [113, 133, 117] is expected to have the same physical origin as the non-monotonic dependence of DNA

ejection. The DNA packaging free energy is calculated to show the non-monotonic effect which shows that DNA-DNA interaction is highly repulsive at small or large concentration of divalent counterions and is negligible or slightly attractive for a concentration in between, 90 mM, at which like-charge attraction among DNA molecules mediated by the counterions [111, 52] is dominant. This non-monotonic effect can be interpreted by *charge inversion* of DNA. It also gives the strength of the short range attraction of DNA-DNA molecules, -0.001$k_BT$/base where DNA-DNA attraction can be easily attacked by thermal agitation. Although the obtained value of neutral concentration $c_{Z,0}$ differs slightly from the fitted value from our phenomenological theory in Chapter 2, it supports the finding that the net charge of DNA can be inverted from negative to positive with increasing counterion concentration and confirms that DNA-DNA interaction is attractive inside a DNA bundle at $c_{Z,0}$. We also see the saturation effect of the "effective force" between DNA molecules at around -4 atm, which can be understood in terms of the overall entropy of the solution and the charge neutrality condition inside the DNA bundle.

The third problem we dealt with (in Chapter 4) is understanding how the mature HIV-1 capsid exhibits various shapes. Upon budding, HIV-1 capsid shape changes from spherical to conical or other shapes. To investigate this diversity, we limit HIV-1 capsid shapes only to sphere, cylindrical, and conical. Nguyen [114] et al paved the way to study conical and cylindrical shapes of HIV capsid by generalizing the icosahedral virus that Caspar and Klug developed in 1960s. This is called extended isometric construction of viral capsids. Previous studies [114] show that cylindrical capsids are always thermodynamically stable and conical capsid shape is not. However, the conical capsid shape with volume constraint can be an optimal shape [115]. To study the diversity of the HIV-1 capsid, we dealt with this problem by giving the length constraint by HIV-1 membrane with the assumption that spherical membrane changes shape from spherical to ellipsoidal during capsid maturation, leading to the conclusion

that HIV membrane energy is always lowest when membrane forms a spherical shape. Using an elastic shell model for the retroviral capsids, the energies of various shapes are calculated both numerically (for cylindrical and 7-5 conical capsid shape) and analytically (for cylindrical and conical shape with different cone angles). These two methods lead to similar conclusions: the cylindrical shape is lowest in energy (thus most thermodynamically stable) if the surface tension of the membrane is low and there is no length constraint. However, for high membrane surface tension, we found that cylindrical and conical shapes are very similar in energies (within the thermal energy $k_BT$ from each other). It is shown that the high surface tension constraint applies to the envelop membrane of retroviruses. This study reaches the conclusion that conical and cylindrical capsid geometries have similar energies and similar probability to appear. Our results explain the experimental data well qualitatively.

The fourth problem we study (in Chapter 5) is obtaining the radial distribution of RNA genomes inside a spherical virus. HIV is a retrovirus including RNA genomes, and this study applies to the RNA distribution inside an immature HIV capsid. RNA in HIV is a key factor in reverse transcription as it replicates itself. RNA has a secondary structure composed of linear sections, branch points, and end-points. Nguyen et al [112] develop RNA condensation theory based on the branching polymer model that Lubensky and Isaacson created [98], and relate this theory to the wetting model that Cahn [35] developed. This problem is solved using a mean field approximation and a spherical symmetry. We model a spherical capsid as a hollow sphere only including RNA molecules nonspecifically interacting with the inner capsid surface. We matched our theoretical data with two experimental data sets. One profile is applied to the weak adsorption case (Dengue virus) [85], where RNA concentration is maximum at the center of the capsid to maximize their configurational entropy. The other case is the strong adsorption case (Bacteriophage $MS_2$) [75]. For this case,

the competition between the branching of RNA secondary structure and its adsorption to the inner capsid results in the formation of a dense layer of RNA near the capsid surface. The layer thickness is a slowly varying function of the capsid inner radius. Consequently, the amount of RNA packaged is proportional to the capsid area( or the number of proteins) instead of its volume. These numerical profiles describe reasonably well the experimental RNA profiles of spherical viruses.

The last problem (in Chapter 6) is about RNA condensation by a single NC protein. We also approach this problem with RNA condensation theory introduced in Chapter 5. RNA-NC protein interaction is critical for reverse transcription in retroviruses. NC proteins also play an important role in many replication processes [147]. One of the important functions of the NC protein in reverse transcription is a chaperone function which catalyzes the rearrangement of a nucleic acid molecule into a conformation that is thermodynamically more stable [126, 32, 94, 150, 10, 64, 97]. The chaperone function is based on the interaction between NC proteins and RNA molecules during reverse transcription. We investigated the interaction between a single NC protein and RNA molecules, hoping to understand the interaction between several NC proteins and RNA molecules. RNA-NC protein interaction is assumed to be non specific and we model a single NC protein as an infinite well at the origin describing the attraction between RNA and a NC protein. For weak adsorption of the NC protein, only a small portion of RNA is condensed near the origin, and the boundary distance $r_0$ between a dilute and condensed phase is linearly proportional to the adsorption strength. For strong adsorption of the NC protein, it is shown that $r_0$ is a slowly varying function with the adsorption strength leading to the conclusion that condensed RNA screens the NC protein. The fact [141] that HTLV-1 NC protein does not condense RNA agrees well with the solution derived.

# APPENDIX A

# DETERMINATION OF THE EQUILIBRIUM LENGTH OF DNA EJECTED

The energy of the DNA segment inside the bacteriophage capsid, Eq. (7) reads

$$E_{in}(L_i, d) = E_{bend}(L_i, d) + E_{int}(L_i, d).$$

To obtain the the optimal DNA-DNA interaxial distance for a given DNA ejected length, $d^*(L_i)$, the energy $E_{in}(L_i, d)$ of DNA segment inside the capsid is minimized with respect to d:

$$\left.\frac{\partial E_{in}(L_i,d)}{\partial d}\right|_{L_i} = \left.\frac{\partial E_{bend}(L_i,d)}{\partial d}\right|_{L_i} + \left.\frac{\partial E_{int}(L_i,d)}{\partial d}\right|_{L_i} = 0\ . \tag{130}$$

Differentiating Eq. (8) and Eq. (10) with respect $d$ with $L_i$ fixed, it gives

$$\left.\frac{\partial E_{bend}(L_i,d)}{\partial d}\right|_{L_i} = a\left\{\frac{4}{3}bL_i^{\frac{1}{3}}d^{\frac{-7}{3}}\left[1+\frac{R^2}{2\left(R^2-b^2L^{\frac{2}{3}}d^{\frac{4}{3}}\right)}\right] + Rd^{-3}\ln\frac{R^2-b^2L^{\frac{2}{3}}d^{\frac{4}{3}}}{\left(R+bL_i^{\frac{1}{3}}d^{\frac{2}{3}}\right)^2}\right\} \tag{131}$$

and

$$\left.\frac{\partial E_{int}(L_i,d)}{\partial d}\right|_{L_i} = -\sqrt{3}F_0L_id\left[\exp\left(\frac{d_0-d}{c}\right)-1\right], \tag{132}$$

where $a = \frac{4\pi l_p k_B T}{\sqrt{3}}$ and $b = \left(\frac{3\sqrt{3}}{8\pi}\right)^{\frac{1}{3}}$. Plugging Eq. (131) and Eq. (132) into Eq. (130), the following equation can be obtained as

$$a\left\{\frac{4}{3}bL_i^{\frac{-2}{3}}d^{\frac{-10}{3}}\left[1+\frac{R^2}{2\left(R^2-b^2L^{\frac{2}{3}}d^{\frac{4}{3}}\right)}\right] + RL_i^{-1}d^{-4}\ln\frac{R^2-b^2L^{\frac{2}{3}}d^{\frac{4}{3}}}{\left(R+bL_i^{\frac{1}{3}}d^{\frac{2}{3}}\right)^2}\right\}$$
$$-\sqrt{3}F_0\left[\exp\left(\frac{d_0-d}{c}\right)-1\right] = 0. \tag{133}$$

By solving Eq. (133), the optimal length of DNA ejected $d$ can be obtained as a function of the packaged length $L_i$.

Using Eqs. (1, 2 , 7, 8, 10), $G(L_i)$ is given by

$$G(L_i, d) = E_{bend}(L_i, d) + E_{int}(L_i, d) + E_{out}(L - L_i) + \Pi_{\rm osm}(L - L_i)\pi a^2. \quad (134)$$

Differentiation $G(L_i, d)$ with respect to $L_i$ is given by

$$\begin{aligned}
\frac{\partial G(L_i,d)}{\partial L_i} &= \left.\frac{\partial G(L_i,d)}{\partial L_i}\right|_d + \left.\frac{\partial G(L_i,d)}{\partial d}\right|_{L_i}\frac{\partial d}{\partial L_i} \\
&= \left.\frac{\partial E_{bend}(L_i,d)}{\partial L_i}\right|_d + \left.\frac{\partial E_{bend}(L_i,d)}{\partial d}\right|_{L_i}\frac{\partial d}{\partial L_i} \\
&+ \left.\frac{\partial E_{int}(L_i,d)}{\partial L_i}\right|_d + \left.\frac{\partial E_{int}(L_i,d)}{\partial d}\right|_{L_i}\frac{\partial d}{\partial L_i} \\
&+ \frac{\partial E_{out}(L - L_i)}{\partial L_i} - \Pi_{\rm osm}\pi a^2 \ . \qquad (135)
\end{aligned}$$

Differentiating Eq. (2), Eq. (8), and Eq. (10) with respect $L_i$ with $d$ fixed, it gives

$$\frac{\partial E_{out}(L - L_i)}{\partial L_i} = \frac{k_B T}{4Z^2 l_B}\left(\ln\frac{N_0}{N}\right)^2, \quad (136)$$

$$\left.\frac{\partial E_{bend}(L_i,d)}{\partial L_i}\right|_d = \frac{abL_i^{\frac{-2}{3}}d^{\frac{-4}{3}}}{3}\left\{-1 + \frac{R^2}{R^2 - b^2 L_i^{\frac{2}{3}} d^{\frac{4}{3}}}\right\}, \quad (137)$$

and

$$\left.\frac{\partial E_{int}(L_i,d)}{\partial L_i}\right|_d = \sqrt{3}F_0\left[(c^2 + cd)\exp\left(\frac{d_0 - d}{c}\right) - (c^2 + cd_0) - \frac{1}{2}(d_0^2 - d^2)\right] - \epsilon \quad (138)$$

where $Z = 2$ for divalent ions. To obtain $\frac{\partial d}{\partial L_i}$, we need to use the balance equation (130) between bending and interaction energy inside capsid . Let's set the left hand side of Eq. (133) to a function $f(d(L_i), L_i)$ and we have $f(d(L_i), L_i) = 0$. Differentiation of $f(d(L_i), L_i)$ with respect to $L_i$ gives the relation $\frac{\partial d}{\partial L_i} = -\left(\frac{\partial f}{\partial L_i}\right)_d \Big/ \left(\frac{\partial f}{\partial d}\right)_{L_i}$.

After some calculations, we can get

$$\begin{aligned}\frac{\partial d}{\partial L_i} &= \left(\frac{2}{3}l_pk_BT\frac{b^2L_i^{\frac{-1}{3}}d^{\frac{10}{3}}}{\left(R^2d^2-b^2L_i^{\frac{2}{3}}d^{\frac{10}{3}}\right)^2}+\frac{3l_pk_BT}{b^2}\left[-\frac{2}{3}L_i^{\frac{-5}{3}}d^{\frac{-10}{3}}\left\{1+\frac{R^2}{2\left(R^2-b^2L_i^{\frac{2}{3}}d^{\frac{4}{3}}\right)}\right\}\right.\right.\\ &\left.\left.-\frac{RL_i^{-2}d^{-4}}{b}\ln\frac{R-bL_i^{\frac{1}{3}}d^{\frac{2}{3}}}{\left(R^2-b^2L_i^{\frac{2}{3}}d^{\frac{4}{3}}\right)^{\frac{1}{2}}}\right]\right)\\ &\div\left(\frac{2l_pk_BT\left(3R^2d-5b^2L_i^{\frac{2}{3}}d^{\frac{7}{3}}\right)}{3\left(R^2d^2-b^2L_i^{\frac{2}{3}}d^{\frac{10}{3}}\right)^2}+\frac{3l_pk_BT}{b^2}\left[\frac{10}{3}L_i^{\frac{-2}{3}}d^{\frac{-13}{3}}\left\{1+\frac{R^2}{5\left(R^2-b^2L_i^{\frac{2}{3}}d^{\frac{4}{3}}\right)}\right\}\right.\right.\\ &\left.\left.+\frac{4RL_i^{-1}d^{-5}}{b}\ln\frac{R-bL_i^{\frac{1}{3}}d^{\frac{2}{3}}}{\left(R^2-b^2L_i^{\frac{2}{3}}d^{\frac{4}{3}}\right)^{\frac{1}{2}}}\right]-\frac{\sqrt{3}F_0}{c}\exp\left(\frac{d_0-d}{c}\right)\right)\quad . \end{aligned}\tag{139}$$

By setting $\frac{\partial G(L_i,d)}{\partial L_i}=0$ together with Eqs. (131, 132, 136,137, 138 and 139), Eq. (135) gives the following equation :

$$\begin{aligned}&\sqrt{3}F_0\left(\exp\left(\frac{d_0-d}{c}\right)(c^2+cd)-(c^2+cd)-\frac{d_0^2-d^2}{2}\right)-\epsilon+\frac{k_BT}{16l_B}\left(\log\frac{N_0}{N}\right)^2\\ &+\frac{4\pi l_pk_BT}{\sqrt{3}}\frac{b}{3}L_i^{\frac{-2}{3}}d^{\frac{-4}{3}}\left(-1+\frac{R^2}{R^2-b^2L_i^{\frac{2}{3}}d^{\frac{4}{3}}}\right)\\ &+\left(\frac{4\pi l_pk_BT}{\sqrt{3}}\left(\frac{4}{3}bL_i^{\frac{1}{3}}d^{\frac{-7}{3}}\left(1+\frac{R^2}{2\left(R^2-b^2L_i^{\frac{2}{3}}d^{\frac{4}{3}}\right)}\right)\right.\right.\\ &\left.\left.+Rd^{-3}\ln\frac{R^2-\left(bL_i^{\frac{1}{3}}d^{\frac{2}{3}}\right)^2}{\left(R+bL_i^{\frac{1}{3}}d^{\frac{2}{3}}\right)^2}\right)-\sqrt{3}F_oL_id\left(\exp\left(\frac{d_0-d}{c}\right)-1\right)\right)\\ &\times\left(\frac{2}{3}l_pk_BT\frac{b^2L_i^{\frac{-1}{3}}d^{\frac{10}{3}}}{\left(R^2d^2-b^2L_i^{\frac{2}{3}}d^{\frac{10}{3}}\right)^2}+\frac{3l_pk_BT}{b^2}\left[-\frac{2}{3}L_i^{\frac{-5}{3}}d^{\frac{-10}{3}}\left\{1+\frac{R^2}{2\left(R^2-b^2L_i^{\frac{2}{3}}d^{\frac{4}{3}}\right)}\right\}\right.\right.\\ &\left.\left.-\frac{RL_i^{-2}d^{-4}}{b}\ln\frac{R-bL_i^{\frac{1}{3}}d^{\frac{2}{3}}}{\left(R^2-b^2L_i^{\frac{2}{3}}d^{\frac{4}{3}}\right)^{\frac{1}{2}}}\right]\right)\end{aligned}$$

$$\div \left( \frac{2l_p k_B T \left(3R^2 d - 5b^2 L_i^{\frac{2}{3}} d^{\frac{7}{3}}\right)}{3\left(R^2 d^2 - b^2 L_i^{\frac{2}{3}} d^{\frac{10}{3}}\right)^2} + \frac{3l_p k_B T}{b^2} \left[ \frac{10}{3} L_i^{\frac{-2}{3}} d^{\frac{-13}{3}} \left\{ 1 + \frac{R^2}{5\left(R^2 - b^2 L_i^{\frac{2}{3}} d^{\frac{4}{3}}\right)} \right\} \right. \right.$$

$$\left. \left. + \frac{4RL_i^{-1} d^{-5}}{b} \ln \frac{R - bL_i^{\frac{1}{3}} d^{\frac{2}{3}}}{\left(R^2 - b^2 L_i^{\frac{2}{3}} d^{\frac{4}{3}}\right)^{\frac{1}{2}}} \right] - \frac{\sqrt{3} F_0}{c} \exp\left(\frac{d_0 - d}{c}\right) \right)$$

$$-\Pi_{\text{osm}} \pi a^2 = 0 \tag{140}$$

where $l_B \left(= \frac{e^2}{4\pi\epsilon_0 D k_B T}\right)$ is the Bjerrum length. By solving Eq. (140), we can obtain the equilibrium DNA packaged length $L_i^*(\Pi_{\text{osm}}, N)$. The equilibrium DNA ejected length can be easily obtained by the relation $L_o^*(\Pi_{\text{osm}}, N) = L - L_i^*(\Pi_{\text{osm}}, N)$ where $L$ is total genome length.

# APPENDIX B

# ANALYTICAL DERIVATION TO CAPSID ELASTIC ENERGY

Each term of Eq. (33) can be calculated as follows. Two dimensional non-spherical capsid is shown in Figure 18. Using the following the relations

$$\tan\theta_m = \frac{R_s'}{z_1} = \frac{R_l'}{z_2} = \frac{R_s\cos\theta_m}{z_1} = \frac{R_l\cos\theta_m}{z_2}, \tag{141}$$

the length of capsid, $H$, is given by

$$\begin{aligned} H &= (z_2 - z_1) + R_l(1+\sin\theta_m) + R_s(1-\sin\theta_m) \\ &= R_l\left(\frac{\cos^2\theta_m}{\sin\theta_m} + 1 + \sin\theta_m\right) + R_s\left(-\frac{\cos^2\theta_m}{\sin\theta_m} + 1 - \sin\theta_m\right), \end{aligned}$$

where $\sin\theta_m = \frac{m}{6}$ and $\frac{\cos^2\theta}{\sin\theta} = \frac{6}{m} - \frac{m}{6}$. Noting that $m$ is equal to $6-M$, $H$ can be finally expressed by

$$H = \frac{12-M}{6-M}R_l - \frac{M}{6-M}R_s. \tag{142}$$

Eq. (20) can be rewritten with Eq. (21) and Eq. (22) by

$$\begin{aligned} \frac{E_0(\gamma,\alpha)}{\kappa} &= 6B\left[1+\ln\left(\frac{\gamma}{\gamma_B}\right)\right] + 6B\left[-\frac{2\alpha}{\sqrt{\pi}}\left(\sqrt{F(\gamma)} - \sqrt{\frac{\gamma_B}{\gamma}}\right) + \frac{\alpha^2}{4\pi}\left(F(\gamma) - \frac{\gamma_B}{\gamma}\right)\right] \\ &\quad + 8\pi - 4\sqrt{\pi}\alpha\sqrt{\frac{\gamma_B}{\gamma}} + \frac{\alpha^2}{2}\frac{\gamma_B}{\gamma}. \end{aligned}$$

Using the relation $F(\gamma) = \frac{1-\gamma_B/\gamma(1-3\cos\theta_1/\tan\theta_1)}{3\cos\theta_1/\tan\theta_1} = \frac{10}{11}\left(1+\frac{1}{10}\frac{\gamma_B}{\gamma}\right)$, the above equation can be written by

$$\begin{aligned} \frac{E_0(\gamma,\alpha)}{\kappa} &= 6B\left[1+\ln\left(\frac{\gamma}{\gamma_B}\right)\right] + 6B\left[-\frac{2\alpha}{\sqrt{\pi}}\left(\sqrt{\frac{10}{11}\left(1+\frac{1}{10}\frac{\gamma_B}{\gamma}\right)} - \sqrt{\frac{\gamma_B}{\gamma}}\right)\right. \\ &\quad \left. + \frac{5\alpha^2}{22\pi}\left(1-\frac{\gamma_B}{\gamma}\right)\right] + 8\pi - 4\sqrt{\pi}\alpha\sqrt{\frac{\gamma_B}{\gamma}} + \frac{\alpha^2}{2}\frac{\gamma_B}{\gamma}. \end{aligned} \tag{143}$$

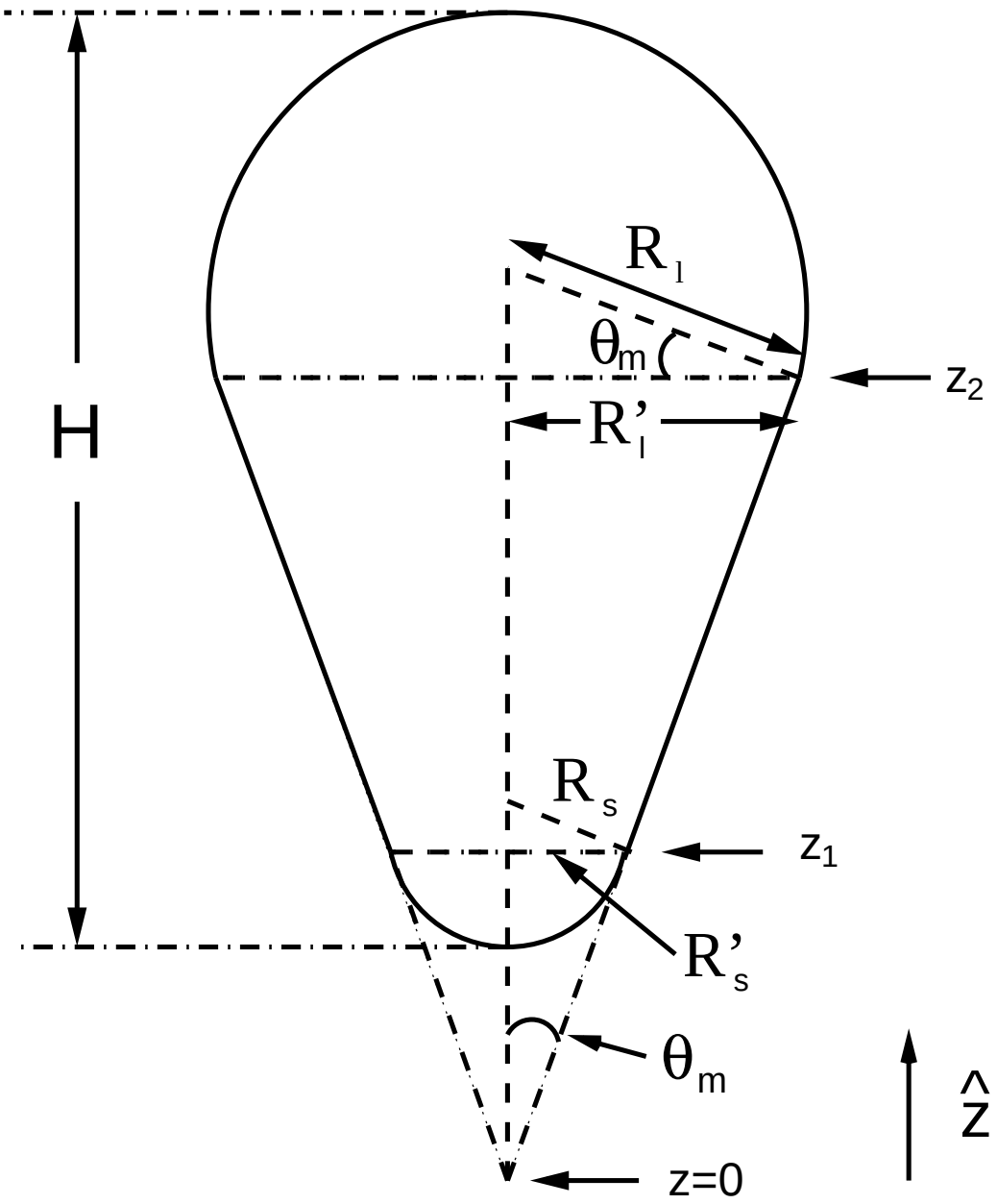


**Figure 21:** Two dimensional approximate shape of non-spherical capsid. This figure is an extension to Figure 13 (A) to support the analytic calculation to elastic capsid energy.

Noting that $\sin\theta_M$ is equal to $1-\frac{M}{6}$, Eq. (24) can be rewritten by

$$\begin{aligned} D_M(\gamma_l,\gamma_s,\alpha) &= \frac{\cos\theta_M}{2\tan\theta_M}\left[\pi\ln\frac{\gamma_l}{\gamma_s}-2\sqrt{\pi}\alpha\left(\sqrt{\frac{\gamma_l}{\gamma}}-\sqrt{\frac{\gamma_s}{\gamma}}\right)+\frac{\alpha^2}{4}\left(\frac{\gamma_l}{\gamma}-\frac{\gamma_s}{\gamma}\right)\right] \\ &= \frac{M(12-M)}{12(6-M)}\left[\pi\ln\frac{\gamma_l}{\gamma_s}-2\sqrt{\pi}\alpha\left(\sqrt{\frac{\gamma_l}{\gamma}}-\sqrt{\frac{\gamma_s}{\gamma}}\right)+\frac{\alpha^2}{4}\left(\frac{\gamma_l}{\gamma}-\frac{\gamma_s}{\gamma}\right)\right]. \end{aligned} \tag{144}$$

Together with Eq. (143) and Eq. (144), Eq. (23) can be rewritten by

$$\begin{aligned} \frac{E_M(\gamma_l,\gamma_s,\alpha)}{\kappa} &= \frac{12-M}{12}\left\{6B\left[1+\ln\left(\frac{\gamma_l}{\gamma_B}\right)\right]+6B\left[-\frac{2\alpha_l}{\sqrt{\pi}}\left(\sqrt{\frac{10}{11}\left(1+\frac{1}{10}\frac{\gamma_B}{\gamma_l}\right)}-\sqrt{\frac{\gamma_B}{\gamma_l}}\right)\right.\right. \\ &\quad \left.\left.+\frac{5\alpha_l^2}{22\pi}\left(1-\frac{\gamma_B}{\gamma_l}\right)\right]+8\pi-4\sqrt{\pi}\alpha_l\sqrt{\frac{\gamma_B}{\gamma_l}}+\frac{\alpha_l^2}{2}\frac{\gamma_B}{\gamma_l}\right\} \\ &\quad +\frac{M}{12}\left\{6B\left[1+\ln\left(\frac{\gamma_s}{\gamma_B}\right)\right]+6B\left[-\frac{2\alpha_s}{\sqrt{\pi}}\left(\sqrt{\frac{10}{11}\left(1+\frac{1}{10}\frac{\gamma_B}{\gamma_s}\right)}-\sqrt{\frac{\gamma_B}{\gamma_s}}\right)\right.\right. \\ &\quad \left.\left.+\frac{5\alpha_s^2}{22\pi}\left(1-\frac{\gamma_B}{\gamma_s}\right)\right]+8\pi-4\sqrt{\pi}\alpha_s\sqrt{\frac{\gamma_B}{\gamma_s}}+\frac{\alpha_s^2}{2}\frac{\gamma_B}{\gamma_s}\right\} \\ &\quad +\frac{M(12-M)}{12(6-M)}\left[\pi\ln\frac{\gamma_l}{\gamma_s}-2\sqrt{\pi}\alpha\left(\sqrt{\frac{\gamma_l}{\gamma}}-\sqrt{\frac{\gamma_s}{\gamma}}\right)+\frac{\alpha^2}{4}\left(\frac{\gamma_l}{\gamma}-\frac{\gamma_s}{\gamma}\right)\right]. \end{aligned} \tag{145}$$

Using the relation $\frac{\alpha_l^2}{\gamma_l}=\frac{\alpha_s^2}{\gamma_s}=\frac{\alpha^2}{\gamma}$, the above equation can be written by

$$\begin{aligned} \frac{E_M(\gamma_l,\gamma_s,\alpha)}{\kappa} &= \frac{12-M}{12}\left\{6B\left[1+\ln\left(\frac{\gamma_l}{\gamma_B}\right)\right]+6B\left[-\frac{2\alpha_l}{\sqrt{\pi}}\left(\sqrt{\frac{10}{11}\left(1+\frac{1}{10}\frac{\gamma_B}{\gamma_l}\right)}-\sqrt{\frac{\gamma_B}{\gamma_l}}\right)\right.\right. \\ &\quad \left.\left.+\frac{5\alpha_l^2}{22\pi}\left(1-\frac{\gamma_B}{\gamma_l}\right)\right]\right\}+\frac{M}{12}\left\{6B\left[1+\ln\left(\frac{\gamma_s}{\gamma_B}\right)\right]\right. \\ &\quad \left.+6B\left[-\frac{2\alpha_s}{\sqrt{\pi}}\left(\sqrt{\frac{10}{11}\left(1+\frac{1}{10}\frac{\gamma_B}{\gamma_s}\right)}-\sqrt{\frac{\gamma_B}{\gamma_s}}\right)+\frac{5\alpha_s^2}{22\pi}\left(1-\frac{\gamma_B}{\gamma_s}\right)\right]\right\} \\ &\quad +8\pi-4\sqrt{\pi}\alpha\sqrt{\frac{\gamma_B}{\gamma}}+\frac{\alpha^2}{2}\frac{\gamma_B}{\gamma} \\ &\quad +\frac{M(12-M)}{12(6-M)}\left[\pi\ln\frac{\gamma_l}{\gamma_s}-2\sqrt{\pi}\alpha\left(\sqrt{\frac{\gamma_l}{\gamma}}-\sqrt{\frac{\gamma_s}{\gamma}}\right)+\frac{\alpha^2}{4}\left(\frac{\gamma_l}{\gamma}-\frac{\gamma_s}{\gamma}\right)\right]. \end{aligned} \tag{146}$$

$\alpha_l,\alpha_s,\gamma_l$ and $\gamma_s$ should be calculated as a function of the small parameter$\epsilon$ to expand Eq. (146). It is assumed that the reference area of the sphere is equal to non spherical capsid area to compare each other. The body of the non spherical capsid can be

calculated by

$$\begin{aligned} S_{body} &= \int_{z_1}^{z_2} dz \int_0^{2\pi} d\phi \; z\frac{\tan\theta_m}{\cos\theta_m} = \pi(z_2^2 - z_1^2)\frac{\tan\theta_m}{\cos\theta_m} = \frac{\pi(R_l'^2 - R_s'^2)}{\sin\theta_m} \\ &= \pi\left(\frac{6}{m} - \frac{m}{6}\right)(R_l^2 - R_s^2). \end{aligned} \tag{147}$$

The cap area is given by

$$S_l = 2\pi R_l^2(1 + \sin\theta_m) = 2\pi R_l^2\frac{6+m}{6}, \tag{148}$$

$$S_s = 2\pi R_s^2(1 - \sin\theta_m) = 2\pi R_s^2\frac{6-m}{6}. \tag{149}$$

The total area of the non-spherical capsid is

$$S_{tot} = S_{body} + S_l + S_l = \pi R_l^2\left(\frac{12m + m^2 + 36}{6m}\right) + \pi R_s^2\left(\frac{12m - m^2 - 36}{6m}\right) \; . \tag{150}$$

By rewriting $S_{tot}$ in terms of $M$ and applying the same area assumption, one equation can be obtained by

$$\frac{(12-M)^2}{6(6-M)}\pi R_l^2 - \frac{M^2}{6(6-M)}\pi R_s^2 = 4\pi R_0^2, \tag{151}$$

where $4\pi R_0^2$ is the reference area of the original spherical capsid. The relation between $H$ and $\epsilon$ is $\frac{H}{H_0} = 1 + \epsilon$ where $H_0 = 2R_0$ is the diameter of the spherical capsid. Using Eq. (142), the other equation can be given by

$$2R_0(1+\epsilon) = \frac{12-M}{6-M}R_l - \frac{M}{6-M}R_s. \tag{152}$$

Combining Eq. (151) and Eq. (152), the following relations can be obtained:

$$\frac{R_l}{R_0} = \frac{1}{12-M}\frac{6 + (6-M)(1+\epsilon)^2}{1+\epsilon}, \tag{153}$$

$$\frac{R_s}{R_0} = \frac{1}{M}\frac{6 - (6-M)(1+\epsilon)^2}{1+\epsilon}. \tag{154}$$

Using Eq. (153) and Eq. (154), $\gamma_l, \gamma_s, \alpha_l$ and $\alpha_s$ can be calculated by

$$\gamma_l = \gamma \left(\frac{R_l}{R_0}\right)^2 = \gamma \left[\frac{6+(6-M)(1+\epsilon)^2}{(12-M)(1+\epsilon)}\right]^2, \tag{155}$$

$$\gamma_s = \gamma \left(\frac{R_s}{R_0}\right)^2 = \gamma \left[\frac{6-(6-M)(1+\epsilon)^2}{M(1+\epsilon)}\right]^2, \tag{156}$$

$$\alpha_l = \alpha \frac{R_l}{R_0} = \alpha \left[\frac{6+(6-M)(1+\epsilon)^2}{(12-M)(1+\epsilon)}\right], \tag{157}$$

$$\alpha_s = \alpha \frac{R_s}{R_0} = \alpha \left[\frac{6-(6-M)(1+\epsilon)^2}{M(1+\epsilon)}\right]. \tag{158}$$

Inserting Eqs. (155 - 158) into Eq. (146) and expanding Eq. (146) in terms of $\epsilon$, the analytic solution Eq. (36) is obtained:

$$\begin{aligned}
\frac{E_M(\gamma,\alpha,\epsilon)}{\kappa} &\simeq \frac{E_M(\gamma,\alpha,0)}{\kappa} + \epsilon\frac{E_M^{(1)}(\gamma,\alpha,\epsilon)}{\kappa} + \epsilon^2\frac{E_M^{(2)}(\gamma,\alpha,\epsilon)}{\kappa} \\
&= 6B\left[1+\ln\left(\frac{\gamma}{\gamma_B}\right)\right] + 8\pi + \alpha\left[\frac{12B}{\sqrt{\pi}}\sqrt{\frac{\gamma_B}{\gamma}} - 4\sqrt{\pi}\sqrt{\frac{\gamma_B}{\gamma}}\right. \\
&\quad \left. - 12B\sqrt{\frac{10}{11\pi}\left(1+\frac{\gamma_B}{10\gamma}\right)}\right] + \alpha^2\left[\frac{15}{11\pi}B + \frac{\gamma_B}{2\gamma} - \frac{15B}{11\pi}\frac{\gamma_B}{\gamma}\right] \\
&+ \epsilon\left\{-12B + 4\pi + \alpha\left[-4\sqrt{\pi} + \frac{6B}{\sqrt{11\pi}}\frac{20}{\sqrt{10+\gamma_B/\gamma}}\right] + \alpha^2\left(1 - \frac{30B}{11\pi}\right)\right\} \\
&+ \epsilon^2\left\{\left[\frac{-132(5M^2-60M+144)}{22(12-M)M}B + \frac{6(M^2-12M+48)}{(12-M)M}\pi\right]\right. \\
&+ \alpha\left[2\sqrt{\pi} - 6B11\pi\left(200 + \frac{10(M^2-12M+144)}{(12-M)M}\frac{\gamma_B}{\gamma}\right)\Big/\left(10+\frac{\gamma_B}{\gamma}\right)^{\frac{3}{2}}\right] \\
&+ \left.\alpha^2\left[\frac{-(M^2-12M+144)}{2(12-M)M} + \frac{30(M^2-12M+144)}{22(12-M)M}\frac{B}{\pi}\right]\right\}
\end{aligned} \tag{159}$$

# APPENDIX C

# CALCULATION OF FUNCTIONAL DERIVATIVES

## *C.1 Chapter IV*

The Euler-Lagrangian equation Eq. (40) can be obtained by setting the function derivative $\delta H_{MF}/\delta Q$ to zero. Let's write the mean field free energy, Eq. (38):

$$H_{\mathrm{MF}} = H_s + \int_0^R 4\pi r^2 dr \left\{ \frac{m}{2} \left( \frac{dQ}{dr} \right)^2 + \Delta W[Q] \right\}.$$

Setting $U(Q, \dot{Q}) = \frac{m}{2} \left( \frac{dQ}{dr} \right)^2 + \Delta W[Q]$,

$$\begin{aligned}
\delta H_{MF} &= \left. \frac{\partial H_s}{\partial Q} \right|_s \delta Q + \int_0^R dr 4\pi r^2 \left\{ \frac{\partial U}{\partial \dot{Q}} \delta \dot{Q} + \frac{\partial U}{\partial Q} \delta Q \right\} \\
&= \left. \frac{\partial H_s}{\partial Q} \right|_s \delta Q + 4\pi \int_0^R r^2 \frac{\partial U}{\partial \dot{Q}} \delta \dot{Q} dr + 4\pi \int_0^R r^2 \frac{\partial U}{\partial Q} \delta Q dr \\
&= \left. \frac{\partial H_s}{\partial Q} \right|_s \delta Q + 4\pi \left[ r^2 \frac{\partial U}{\partial \dot{Q}} \delta Q \right|_0^R - \int_0^R \frac{d}{dr} \left( r^2 \frac{\partial U}{\partial \dot{Q}} \right) \delta Q dr \Bigg] + 4\pi \int_0^R r^2 \frac{\partial U}{\partial Q} \delta Q dr \\
&= \left. \frac{\partial H_s}{\partial Q} \right|_s \delta Q + 4\pi \left. r^2 \frac{\partial U}{\partial \dot{Q}} \delta Q \right|_0^R + 4\pi \int_0^R \left( -2r \frac{\partial U}{\partial \dot{Q}} - r^2 \frac{d}{dr} \frac{\partial U}{\partial \dot{Q}} + r^2 \frac{\partial U}{\partial Q} \right) \delta Q dr \\
&= 0, \qquad (160)
\end{aligned}$$

where $\frac{\partial U}{\partial \dot{Q}} = m \frac{dQ}{dr}$ and $\frac{\partial U}{\partial Q} = \frac{\partial \Delta W(Q)}{\partial Q}$. From Eq. (160), the Euler-Lagrange equation (40)

$$\frac{d^2 Q}{dr^2} + \frac{2}{r} \frac{dQ}{dr} - \frac{1}{m} \frac{\partial \Delta W}{\partial Q} = 0$$

and a boundary condition Eq. (41)

$$\left. \frac{dQ}{dr} \right|_{r=R} = \frac{H_s{}'[Q(R)]}{4\pi R^2 m} = -\gamma_1 - \gamma_2 Q(R)$$

are derived.

## *C.2 Chapter V*

In this case, the integration ranges from 0 to $\infty$. Let's write down the mean field energy :

$$H_{MF} = H_0 + \int_0^\infty 4\pi r^2 dr \left\{ \frac{m}{2}\left(\frac{dQ}{dr}\right)^2 + \Delta W[Q] \right\} \tag{161}$$

Following the step shown in Appendix C.1, it arrives at

$$\begin{aligned} \delta H_{MF} &= \left.\frac{\partial H_0}{\partial Q}\right|_{r=0} \delta Q + 4\pi r^2 \frac{\partial U}{\partial \dot{Q}} \delta Q \Big|_0^\infty \\ &+ 4\pi \int_0^\infty \left( -2r\frac{\partial U}{\partial \dot{Q}} - r^2 \frac{d}{dr}\frac{\partial U}{\partial \dot{Q}} + r^2 \frac{\partial U}{\partial Q} \right) \delta Q dr. \\ &= 0. \end{aligned} \tag{162}$$

Eq. (162) gives the Euler Lagrange equation and two boundary conditions

$$\left.\frac{dQ}{dr}\right|_{r=0} = \frac{-\gamma}{4\pi r^2 m}$$

and

$$\left.\frac{dQ}{dr}\right|_{r\to\infty} = 0.$$

# REFERENCES


[1] "HIV and its transmission," Tech. Rep. http://www.cdc.gov/HIV/resources/factsheets/, Centers for Disease Control and Prevention, Atlanta, Georgia, July 1999.

[2] "2007 aids epidemic update," Tech. Rep. UNAIDS/07.27E/JC1322E, UNAIDS, World Health Organization, UNAIDS, Switzerland, Dec. 2007.

[3] Ako-Adjei, D., Johnson, M. C., and Vogt, V. M., "The retroviral capsid domain dictates virion size, morphology, and coassembly of gag into virus-like particles," *J. Virol.*, vol. 79, pp. 13463–13472, 2005.

[4] Aldovini, A. and Young, R. A., "Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus," *J. Virol.*, vol. 64, pp. 1920–1926, 1990.

[5] Allain, B., Lapadat-Tapolsky, M., Berlioz, C., and Darlix, J.-L., "Transactivation of the minus-strand DNA transfer by nucleocapsid protein during reverse transcription of the retroviral genome," *EMBO J.*, vol. 13, pp. 973–981, 1994.

[6] Arsuaga, J., Tan, R.-Z., Vazquez, M., Sumners, D. W., and Harvey, S. C., "Investigation of viral dna packaging using molecular mechanics models," *Biophys. Chem*, vol. 101-102, pp. 475–484, 2002.

[7] Arts, E. J., Li, Z., and Wainberg, M. A., "Analysis of primer extension and the first template switch during human immunodeficiency virus reverse transcription," *J. Biomed. Sci.*, vol. 2, pp. 314–321, 1995.

[8] Azoulay, J., Clamme, J. P., Darlix, J. L., Roques, B. P., and Mely, Y., "Destabilization of the HIV-1 complementary sequence of tar by the nucleocapsid protein through activation of conformational fluctuations," *J. Mol. Biol.*, vol. 326, pp. 691–700, 2003.

[9] Baker, T. S., Olson, N. H., and Fuller, S. D., "Adding the third dimension to virus life cycles: Three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs," *Microbiol. Mol. Biol. Rev.*, vol. 63, no. 4, p. 862, 1999.

[10] Bampi, C., Jacquenet, S., Lener, D., Decimo, D., and Darlix, J. L., "The chaperoning and assistance roles of the HIV-1 nucleocapsid protein in proviral DNA synthesis and maintenance," *Curr. HIV Res.*, vol. 2, pp. 79–92, 2004.

[11] Bell, C., Devarajan, S., and Gersbach, H., "The long-run economic costs of aids: Theory and an application to south africa," Tech. Rep. The World Bank Policy Research Working Paper No. 3152, The World Bank, 2003.

[12] Beltz, H., Azoulay, J., Bernacchi, S., Clamme, J. P., Ficheux, D., and et al., B. R., "Impact of the terminal bulges of HIV-1 cTAR DNA on its stability and the destabilizing activity of the nucleocapsid protein ncp7," *J. Mol. Biol.*, vol. 328, pp. 95–108, 2003.

[13] Beltz, H., Piemont, E., Schaub, E., Ficheux, D., Roques, B., Darlix, J. L., and Mely, Y., "Role of the structure of the top half of HIV-1 cTAR DNA on the nucleic acid destabilizing activity of the nucleocapsid protein NCp7," *J. Mol. Biol.*, vol. 338, pp. 711–723, 2004.

[14] Belyi, V. A. and Muthukumar, M., "Electrostatic origin of the genome packing in viruses," *Proc. Natl. Acad. Sci. USA*, vol. 103, no. 46, pp. 17174–17178, 2006.

[15] Benjamin, J., Ganser-Pornillos, B. K., Tivol, W. F., Sundquist, W. I., and Jensen, G. J., "Three-dimensional structure of HIV-1 virus-like particles by electron cryotomography," *J. Mol. Biol.*, vol. 346, p. 577, 2005.

[16] Berg, J. M., "Potential metal-binding domains in nucleic acid binding proteins," *Science*, vol. 232, pp. 485–487, 1986.

[17] Berkowitz, R., Fisher, J., and Goff, S. P., "RNA packaging," *Curr. Top. Microbiol. Immunol.*, vol. 214, pp. 177–218, 1996.

[18] Bernacchi, S., Stoylov, S., Piemont, E., Ficheux, D., Roques, B. P., Darlix, J. L., and Mely, Y., "HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence," *J. Mol. Biol.*, vol. 317, pp. 385–399, 2002.

[19] Black, L. W., "DNA packaging in dsDNA bacteriophages," *Annu. Rev. Microbiol.*, vol. 43, pp. 267–292, 1989.

[20] Briggs, J. A., Johnson, M. C., Simon, M. N., and andV. M. Vogt, S. D. F., "Cryo-electron microscopy reveals conserved and divergent features of gag packing in immature particles of rous sarcoma virus and human immunodeficiency virus," *J. Mol. Biol.*, vol. 355, pp. 157–168, 2006.

[21] Briggs, J. A., Simon, M. N., Gross, I., Krausslich, H. G., Fuller, S. D., Vogt, V. M., and Johnson, M. C., "The stoichiometry of gag protein in HIV-1," *Nat. Struct. Mol. Biol.*, vol. 11, pp. 672–675, 2004.

[22] Buchbinder, S. P., Katz, M. H., Hessol, N. A., O'Malley, P. M., and Holmberg, S. D., "Long-term HIV-1 infection without immunologic progression," *AIDS*, vol. 8, no. 8, pp. 1123–1128, 1994.

[23] Cam, E. L., Coulaud, D., Delain, E., Petitjean, P., Roques, B. P., and et al, D. G., "Properties and growth mechanism of the ordered aggregation of a model RNA by the hiv-1 nucleocapsid protein: an electron microscopy investigation," *Biopolymers*, vol. 45, pp. 217–229, 1998.

[24] Castelnovo, M., Bowles, R. K., Reiss, H., and Gelbart, W. M., "Osmotic force resisting chain insertion in a colloidal suspension," *Eur. Phys. J. E*, vol. 10, no. 2, pp. 191–197, 2003.

[25] Chan, B. and Musier-Forsyth, K., "The nucleocapsid protein specifically anneals tRNA Lys-3 onto a noncomplementary primer binding site within the HIV-1 RNA genome in vitro," *Proc. Natl. Acad. Sci. USA*, vol. 94, pp. 13530–13535, 1997.

[26] Chan, B., Weidemaier, K., Yip, W. T., Barbara, P. F., and Musier-Forsyth, K., "Intra-tRNA distance measurements for nucleocapsid protein-dependent tRNA unwinding during priming of HIV reverse transcription," *Proc. Natl. Acad. Sci. USA*, vol. 96, pp. 459–464, 1999.

[27] Coffin, J. M., Hughes, S. H., and Varmus, H. E., *Retroviruses*. New York: Cold Spring Harbor Laboratory Press, 1997.

[28] Comolli, L. R., Spakowitz, A. J., Siegerist, C. E., Jardine, P. J., Grimes, S., Anderson, D. L., Bustamante, C., and Downing, K. H., "Three-dimensional architecture of the bacteriophage $\phi 29$ packaged genome and elucidation of its packaging process," *Virology*, vol. 371, pp. 267–277, 2008.

[29] Covey, S. N., "Amino acid sequence homology in gag region of reverse transcribing elements and the coat protein gene of cauliflower mosaic virus," *Nucleic Acids Res.*, vol. 14, no. 2, pp. 623–633, 1986.

[30] Cristofari, G. and Darlix, J. L., "The ubiquitous nature of RNA chaperone proteins," *Prog. Nucleic. Acid Res. Mol. Biol.*, vol. 72, pp. 223–268, 2002.

[31] Daoud, M. and Joanny, J., "Conformation of branched polymers," *Journal de physique*, vol. 42, pp. 1359–1371, 1981.

[32] Darlix, J. L., Lapadat-Tapolsky, M., de Rocquigny, H., and Roques, B. P., "First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses," *J. Mol. Biol.*, vol. 254, pp. 523–537, 1995.

[33] Darlix, J. L., Vincent, A., Gabus, C., de Rocquigny, H., and Roques, B., "Transactivation of the 5' to 3' viral DNA strand transfer by nucleocapsid protein during reverse transcription of HIV-1 RNA," *C. R. Acad. Sci (Ser. III)*, vol. 316, pp. 763–771, 1993.

[34] de Gennes, P. G., "Wetting: statics and dynamics," *Rev. Mod. Phys.*, vol. 57, no. 3, pp. 827–863, 1985.

[35] de Gennes, P. G., *Scaling concepts in Polymer Physics*. Ithaca and London: Cornell University Press, 1988.

[36] DeStefano, J. J., Buiser, R. G., Mallaber, L. M., Fay, P. J., and Bambara, R. A., "Parameters that influence processive synthesis and site-specific termination by human immunodeficiency virus reverse transcriptase on RNA and DNA templates," *Biochim. Biophys. Acta*, vol. 1131, pp. 270–280, 1992.

[37] D'Herelle, F., "Sur un microbe invisible antagoniste des bacilles dysenterigues," *Acad Sci Ser D*, vol. 165, p. 373, 1917.

[38] Dib-Hajj, F., Khan, R., and Giedroc, D. P., "Retroviral nucleocapsid proteins possess potent nucleic acid strand renaturation activity," *Protein Sci.*, vol. 2, pp. 231–243, 1993.

[39] Dorfman, T., Luban, J., Goff, S. P., Haseltine, W. A., and Gottlinger, H. G., "Mapping of functionally important residues of a cysteine-histidine box in the human immunodeficiency virus type 1 nucleocapsid protein," *J. Virol.*, vol. 67, pp. 6159–6169, 1993.

[40] Dudding, L. R., Nkabinde, N. C., and Mizrahi, V., "Analysis of RNA- and DNA-dependent DNA polymerase activities of point mutants of HIV-1 reverse transcriptase lacking ribonuclease h activity," *Biochemistry*, vol. 30, pp. 10498–10506, 1991.

[41] Evilevitch, A., Castelnovo, M., Knobler, C. M., and Gelbart, W. M., "Measuring the force ejecting DNA from phage," *J. Phys. Chem. B*, vol. 108, no. 21, pp. 6838–6843, 2004.

[42] Evilevitch, A., Fang, L. T., Yoffe, A. M., Castelnovo, M., Rau, D. C., Parsegian, V. A., Gelbart, W. M., and Knobler, C. M., "Effects of salt concentrations and bending energy on the extent of ejection of phage genomes," *Biophys. J.*, vol. 94, no. 3, pp. 1110–1120, 2008.

[43] Evilevitch, A., Lavelle, L., Knobler, C. M., Raspaud, E., and Gelbart, W. M., "Osmotic pressure inhibition of DNA ejection from phage," *Proc. Nat. Acad. Sci. USA*, vol. 100, p. 9292, 2003.

[44] Fang, L. T. and Gelbart, W. M. *private communications*, 2009.

[45] Feng, Y. X., Campbell, S., Harvin, D., Ehresmann, B., Ehresmann, C., and Rein, A., "The human immunodeficiency virus type 1 gag polyprotein has nucleic acid chaperone activity: possible role in dimerization of genomic RNA and palcement of tRNA on the primer binding site," *J. virol.*, vol. 73, pp. 4251–4256, 1999.

[46] Fisher, R. J., Rein, A., Fivash, M., Urbaneja, M. A., Casas-Finet, J. R., Medaglia, M., and Henderson, L. E., "Sequence-specific binding of human immunodeficiency virus type 1 nucleocapsid protein to short oligonucleotides," *J. Virol.*, vol. 72, pp. 1902–1909, 1998.

[47] Forrey, C. and Muthukumar, M., "Langevin dynamics simulations of genome packing in bacteriophage," *Biophys. J.*, vol. 91, pp. 25–41, 2006.

[48] Fuller, D. N., Raymer, D. M., Kottediel, V. I., Rao, V. B., and Smith, D. E., "Single phage t4 dna packaging motors exhibit large force generation, hight velocity, and dynamic variability," *Porc. Natl. Acad. Sci. USA*, vol. 104, pp. 16868–16873, 2007.

[49] Fuller, D. N., Raymer, D. M., Rickgauer, J. P., Robertson, R. M., Catalano, C. E., Anderson, D. L., Grimes, S., and Smith, D. E., "Measurements of single dna molecule packaging dynamics in bacteriophage lambda reveal high forces, high motor processivity, and capsid transformations," *J. Mol. Biol.*, vol. 373, pp. 1113–1122, 2007.

[50] Fuller, D. N., Rickgauer, J. P., Jardine, P. J., Grimes, S., Anderson, D. L., and Smith, D. E., "Ionic effects on viral dna packaging and portal motor function in bacteriophage $\phi$29," *Porc. Natl. Acad. Sci. USA*, vol. 104, pp. 11245–11250, 2007.

[51] Ganser, B. K., Li, S., Klishko, V. Y., Finch, J. T., and Sundquist, W. I., "Assembly and analysis of conical models for the HIV-1 core," *Science*, vol. 283, p. 80, 1999.

[52] Gelbart, W. M., Bruinsma, R. F., Pincus, P. A., and Parsegian, A. V., "DNA-inspired electrostatics," *Phys. Today*, vol. 53, pp. 38–44, 2000.

[53] Gorelick, R. J., Nigida Jr., S. M., Bess Jr., J. W., Arthur, L. O., Henderson, L. E., and Rein, A., "Noninfectious human immunodeficiency virus type 1 mutants deficient in genomic RNA," *J. Virol.*, vol. 64, pp. 3207–3211, 1990.

[54] Grayson, P., Evilevitch, A., Inamdar, M. M., Purohit, P. K., and Gelbart, W. M., "The effect of genome length on ejection forces in bacteriophage lambda," *Virology*, vol. 348, pp. 430–432, 2006.

[55] Green, L. M. and Berg, J. M., "Retroviral nucleocapsid protein-metal ion interactions," *Proc. Natl Acad. Sci. USA*, vol. 87, pp. 6403–6407, 1990.

[56] Grigorov, B., decimo, D., Smagulova, F., Mougel, M., Muriaux, D., and Darlix, J. L., "Intracelluar HIV-1 gag localization is impaired by mutations in the nucleocapsid zinc fingers," *Retrovirology*, vol. 4, p. 54, 2007.

[57] Grosberg, A. Y., Nguyen, T. T., and Shklovskii, B., "Colloquium: The physics of charge inversion in chemical and biological systems," *Rev. Mod. Phys.*, vol. 74, pp. 329–345, 2002.

[58] Gr$\phi$nbech-Jensen, N., Mashl, R. J., Bruinsma, R. F., and Gelbart, W. M., "Counterion-induced attraction between rigid polyelectrolytes," *Phys. Rev. Lett.*, vol. 78, pp. 2477–2480, 1997.

[59] Guldbrand, L., Nilsson, L. G., and Nordenskiöld, L., "A Monte Carlo simulation study of electrostatic forces between hexagonally packed DNA double helices," *J. Chem. Phys.*, vol. 85, p. 6686, 1986.

[60] Guo, J., Wu, T., Anderson, J., Kane, B. F., Johnson, D. G., Gorelick, R. J., Henderson, L. E., and Levin, J. G., "Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate efficient minus- and plus-strand transfer," *J. Virol.*, vol. 74, no. 19, pp. 8980–8988, 2000.

[61] Hargittai, M. R., Gorelick, R. J., Rouzina, I., and Musier-Forsyth, K., "Mechanistic insights into the kinetics of HIV-1 nucleocapsid protein-facilitated tRNA annealing to the primer binding site," *J. Mol. Biol.*, vol. 337, pp. 951–968, 2004.

[62] Hargittai, M. R., mangla, A. T., Gorelick, R. J., and Musier-Forsyth, K., "HIV-1 nucleocapsid protein zinc finger structures induce tRNA(Lys,3) structural changes but are not critical for primer/template annealing," *J. Mol. Biol.*, vol. 312, pp. 985–997, 2001.

[63] Henderson, L. E., Bowers, M. A., II, R. C. S., Serabyn, S. A., Johnson, D. G., Jr., J. W. B., Arthur, L. O., Bryant, D. K., and Fenselau, C., "Gag proteins of the highly replicative mn strain of human immunodeficiency virus type 1: posttranslational modifications, proteolytic processings and complete amino acid sequences," *J. Virol.*, vol. 66, pp. 1856–1865, 1992.

[64] Herschlag, D., "RNA chaperones and the RNA folding problem," *J. Biol. Chem.*, vol. 270, pp. 20871–20874, 1995.

[65] Hershey, A. D. and Chase, M. M., "Independent functions of viral protein and nucleic acid in growth of bacteriophage," *J. Gen. Physiol.*, vol. 36, pp. 39–56, 1952.

[66] Hicks, S. D. and Henley, C. L., "Irreversible growth model for virus capsid assembly," *Phys. Rev. E*, vol. 74, p. 031912, 2006.

[67] Higgs, P. G., "RNA secondary structure: physical and computational aspects," *Q. Rev. Biophys.*, vol. 33, p. 199, August 2000.

[68] Hilhorst, H. J., "Renormalization of the self-avoiding walk on a lattice," *Phys. Lett.*, vol. 56A, no. 3, pp. 153–154, 1976.

[69] Hilhorst, H. J., "Real-space renormalization of the self-avoiding walk by a linear transformation," *Phys. Rev. B*, vol. 16, no. 3, pp. 1253–1265, 1977.

[70] Holyst, R. and Poniewierski, A., "Wetting on a spherical surface," *Phys. Rev. B.*, vol. 36, pp. 5628–5630, 1987.

[71] Hu, T. and Shklovskii, B. I., "Kinetics of viral self-assembly: Role of the single-stranded RNA antenna," *Phys. Rev. E*, vol. 75, p. 051901, 2007.

[72] Hu, T., Zhang, R., and Shklovskii, B. I., "Electrostatic theory of viral self-assembly," *Physica A*, vol. 387, pp. 3059–3064, 2008.

[73] Hud, N. V., "Double-stranded dna organization in bacteriophage heads: an alternative toroid-based model," *Biophys. J.*, vol. 69, pp. 1355–62, 1995.

[74] Hud, N. V. and Downing, K. H., "Cryoelectron microscopy of phage DNA condensates in vitreous ice: The fine structure of DNA toroids," *Proc. Nat. Acad. Sci. USA*, vol. 98, no. 26, pp. 14925–14930, 2001.

[75] Jacrot, B., Chauvin, C., and Witz, J., "Comparative neutron small-angle scattering study of small spherical RNA viruses," *Nature*, vol. 266, pp. 417–421, 1977.

[76] Jentoft, J. E., Smith, L. M., Fu, X. D., Johnson, M., and Leis, J., "Conserved cycteine and histidine residues of the avian myeloblastosis virus nucleocapsid protein are essential for viral replication but are not zinc-binding fingers," *Proc. natl. Acad. Sci. USA*, vol. 85, pp. 7094–7098, 1988.

[77] Ji, X., Klarmann, G. J., and Preston, B. D., "Effect of human immunodeficiency virus type 1 (HIV-1) nucleocapsid protein on HIV-1 reverse transcriptase activity in vitro," *Biochemistry*, vol. 35, pp. 132–143, 1996.

[78] Johnson, J. and Rueckert, R., *in Structural Biology of Viruses.* , edited by Chiu, W., Burnett, R. M. , and Garcea, R. L., Oxford University Press, New York, 1997.

[79] Kankia, B. I., Barany, G., and Musier-Forsyth, K., "Unfolding of DNA quadruplexes induced by HIV-1 nucleocapsid protein," *Nucleic Acids Res.*, vol. 33, pp. 4395–4403, 2005.

[80] Kim, J. K., Palaniappan, C., Wu, W., Fay, P. J., and Bambara, R. A., "Evidence for a unique mechanism of strand transfer from the transactivation response region of HIV-1," *J. Biol. chem.*, vol. 272, pp. 16769–16777, 1997.

[81] Kindt, J., Tzlil, S., Ben-Shaul, A., and Gelbart, W. M., "DNA packaging and ejection forces in bacteriophage," *Proc. Nat. Acad. Sci. USA*, vol. 98, no. 24, pp. 13671–13674, 2001.

[82] Klarmann, G. J., Schauber, C. A., and Preston, B. D., "Template-directed pausing of DNA synthesis by HIV-1 reverse transcriptase during polymerization of HIV-1 sequences in vitro," *J. Biol. Chem*, vol. 268, pp. 9793–9802, May 1993.

[83] Knobler, C. M. and Gelbart, W. M., "Physical chemistry of DNA viruses," *Annu. Rev. Phys. Chem.*, vol. 60, pp. 367–383, 2009.

[84] Koltover, I., Wagner, K., and Safinya, C. R., "DNA condensation in two dimensions," *Proc. Nat. Acad. Sci. USA*, vol. 97, no. 26, pp. 14046–14051, 2000.

[85] Kuhn, R. J., Zhang, W., Rossmann, M. G., Pletnev, S. V., Corver, J., Lenches, E., Jones, C. T., Mukhopadhyay, S., Chipman, P. R., Strauss, E. G., Baker, T. S., and Strauss, J. H., "Structure of dengue virus: Implications for flavivirus organization, maturation, and fusion," *Cell*, vol. 108, pp. 717–725, 2002.

[86] LaMarque, J. C., Le, T. V., and Harvey, S. C., "Packaging double-helical dna into viral capsids," *Biopolymers*, vol. 73, pp. 348–355, 2004.

[87] Lapadat-Tapolsky, M., Rocquigny, H. D., Gent, D. V., Roques, B., Plasterk, R., and Darlix, J. L., "Interactions between HIV-1 nucleocapsid protein and viral DNA may have important functions in the viral life cyle," *Nucleic Acids Res.*, vol. 21, pp. 831–839, 1993.

[88] Lee, S., Tran, C. V., and Nguyen, T. T., "Inhibition of DNA ejection from bacteriophage by $mg^{+2}$ counterions," *Phys. Rev. Lett. (submitted)*, pp. arXiv:cond–mat/0811.1296.

[89] Lee, S., Tran, C. V., and Nguyen, T. T., "Inhibition of DNA ejection from bacteriophage by $Mg^{+2}$ counterions," *Abstract in American Society Meeting*, no. D40.00008, Session D40, 2009.

[90] Lee, S. I. and Nguyen, T. T., "Radial distribution of RNA genomes packaged inside spherical viruses," *Phys. Rev. Lett.*, vol. 100, p. 198102, 2008.

[91] Lee, S. I. and Nguyen, T. T., "Diversity of *in-vivo* assembled HIV-1 capsids," *Abstract in American Society Meeting*, no. X40.00005, Session X40, 2008.

[92] Leis, J., Baltimore, D., Bishop, J. M., Coffin, J., Fleissner, E., Goff, S. P., Oroszlan, S., Robinson, H., Skalka, A. M., Temin, H. M., and Vogt, V., "Standardized and simplified nomenclature for proteins common to all retroviruses," *J. Virol.*, vol. 62, pp. 1808–1809, 1988.

[93] Letellier, L., Boulanger, P., Plancon, L., Jacquot, P., and Santamaria, M., "Main features on tailed phage, host recognition and DNA uptake," *Front. Biosci*, vol. 9, pp. 1228–1339, 2004.

[94] Levin, J. G., Guo, J., Rouzina, I., and Musier-Forsyth, K., "Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism," *Prog. Nucleic Acid Res. Mol. Biol.*, vol. 80, pp. 217–286, 2005.

[95] Lidmar, J., Mirny, L., and Nelson, D. R., "Virus shapes and buckling transitions in spherical shell," *Phys. Rev. E*, vol. 68, p. 051910, 2003.

[96] Locker, R. C. and Harvey, S. C., "A model for viral genome packing," *Mutiscale Model. Simul.*, vol. 5, pp. 1264–1279, 2006.

[97] Lorsch, J. R., "RNA chaperones exist and dead box proteins get a life," *Cell*, vol. 109, pp. 797–800, 2002.

[98] Lubensky, T. C. and Isaacson, J., "Statistics of lattice animals and dilute branched polymers," *Phys. Rev. A*, vol. 20, pp. 2130–2146, 1979.

[99] Lyubartsev, A. P. and Nordenskiöld, L., "Monte Carlo simulation study of ion distribution and osmotic pressure in hexagonally oriented DNA," *J. Phys. Chem*, vol. 99, no. 25, pp. 10373–10382, 1995.

[100] Lyubartsev, A. P., Tang, J. X., Janmey, P. A., and Nordenskiöld, L., "Electrostatically induced polyelectrolyte association of rodlike virus particles," *Phys. Rev. Lett.*, vol. 81, pp. 5465–5468, 1998.

[101] M. H. V. van Regenmortel, C. M. F., Bishop, D. H. L., Carstens, E. B., Estes, M. K., Lemon, S. M., Maniloff, J., Mayo, M. A., McGeoch, D. J., Pringle, C. R., and Wickner, R. B., *Virus Taxonomy Classification and Nomenclature of Viruses: Seventh Report of the International Committee on Taxonomy of Viruses*. San Diego: Academic Press, 2000.

[102] Messer, L. I., currey, K. M., O'Neill, B. J., Jr., J. V. M., Levin, J. G., and Gerwin, B. I., "Functional analysis of reverse transcription by a frameshift pol mutant of murine leukemia virus," *Virology*, vol. 146, pp. 146–152, 1985.

[103] Mirambeau, G., Lyonnais, S., Coulaud, D., Hameau, L., Lafosse, S., and et al., J. J., "Transmission electron microscopy reveals an optimal HIV-1 nucleocapsid aggregation with single-stranded nucleic acids and the mature HIV-1 nucleocapsid protein," *J. Mol. Biol.*, vol. 364, pp. 496–511, 2006.

[104] Mizuno, A., Ido, E., Goto, T., Kuwata, T., Nakai, M., and Hayami, M., "Mutational analysis of two zinc finger motifs in HIV type 1 nucleocapsid proteins: effects on proteolytic processing of gag precursors and particle formation," *AIDS Res. Hum. Retroviruses*, vol. 12, pp. 793–800, 1996.

[105] Moreira, A. G. and Netz, R. R., "Simulations of counterions at charged plates," *Eur. Phys. J. E*, vol. 8, no. 1, pp. 33–58, 2002.

[106] Morgan, D., Mahe, C., Mayanja, B., Okongo, J. M., Lubega, R., and Whitworth, J. A., "HIV-1 infection in rural africa: is there a difference in median time to aids and survival compared with that in industrialized countries?," *AIDS*, vol. 16, no. 4, pp. 597–603, 2002.

[107] Morris, C. E. and Homman, U., "Cell surface area regulation and membrane tension," *J. Membrane Biol.*, vol. 179, pp. 79–102, 2001.

[108] Murialdo, H., "Bacteriophage lambda DNA maturation and packaging," *Annu. Rev. Biochem.*, vol. 60, pp. 125–153, 1991.

[109] Muriaux, D., Darlix, J. L., and Cimarelli, A., "Targetting HIV-1 budding," *Curr Pharmacol Design*, vol. 10, pp. 3725–3739, 2004.

[110] Muriaux, D., Mirro, J., Harvin, D., and Rein, A., "RNA is a structural element in retrovirus particles," *Proc. Natl. Acad. Sci.*, vol. 98, pp. 5246–5251, 2001.

[111] Naji, A., Arnold, A., Holm, C., and Netz, R. R., "Attraction and unbinding of like-charged rods," *Europhy. Lett.*, vol. 67, no. 1, pp. 130–136, 2004.

[112] Nguyen, T. T. and Bruinsma, R. F., "RNA condensation and the wetting transition," *Phys. Rev. Lett.*, vol. 97, p. 108102, 2006.

[113] Nguyen, T. T., Rouzina, I., and Shklovskii, B. I., "Reentrant condensation of DNA induced by multivalent counterions," *J. Chem. Phys.*, vol. 112, no. 5, pp. 2562–2568, 2000.

[114] Nguyen, T. T. and Bruinsma, R. F., "Elasticity theory and shape transitions of viral shells," *Phys. Rev. E*, vol. 72, p. 051923, 2005.

[115] Nguyen, T. T. and Bruinsma, R. F., "Continuum theory of retroviral capsids," *Phys. Rev. Lett.*, vol. 96, p. 078102, 2006.

[116] Peliska, J. A., Balasubramanian, S., Giedroc, D. P., and Benkovic, S. J., "Recombinant HIV-1 nucleocapsid protein accelerates HIV-1 reverse transcriptase catalyzed DNA strand transfer reactions and modulates RNase H activity," *Biochemistry*, vol. 33, pp. 13817–13823, 1994.

[117] Pelta, J., Durand, D., Doucet, J., and Livolant, F., "DNA mesophases induced by spermidine: Structural properties and biological implications," *Biophys. J.*, vol. 71, pp. 48–63, 1996.

[118] Petrov, A. S., Lim-Hing, K., and Harvey, S. C., "Packaging of DNA by bacteriophage epsilon15: Structure, forces, and thermodynamics," *Structure*, vol. 15, p. 807, 2007.

[119] Poon, D. T., Wu., J., and Aldovini, A., "Charged amino acid residues of human immunodeficiency virus type 1 nucleocapsid p7 protein involved in RNA packaging and infectivity," *J. Virol.*, vol. 70, pp. 6607–6616, 1996.

[120] Pornillos, O., Ganser-Pornillos, B. K., Kelly, B. N., Hua, Y., Whitby, F. G., Stout, C. D., Sundquist, W. I., Hill, C. P., , and Yeager, M., "X-ray structures of the hexameric building block of the HIV capsid," *Cell*, vol. 137, pp. 1282–1292, 2009.

[121] Prats, A. C., Sarih, L., Gabus, C., Litvak, S., Keith, G., and Darlix, J. L., "Small finger protein of avian and murine retroviruses has nucleic acid annealing activity and positions the replication primer tRNA onto genomic RNA," *EMBO J.*, vol. 7, pp. 1777–1783, 1988.

[122] Purohit, P. K., Inamdar, M. M., Grayson, P. D., Squires, T. M., Knodev, J., and Phillips, R., "Forces during bacteriophage dna packaging and ejection," *Biophys. J.*, vol. 88, pp. 851–866, 2005.

[123] Rau, D. C., Lee, B., and Parsegian, V. A., "Measurement of the repulsive force between polyelectrolyte molecules in ionic solution: hydration forces between parallel dna double helices," *Porc. natl. Acad. Sci. USA*, vol. 81, pp. 2621–25, 1984.

[124] Rau, D. C. and Parsegian, V. A., "Direct measurement of the intermolecular forces between counterion-condensed DNA double helices. Evidence for long range attractive hydration forces," *Biophys. J.*, vol. 61, pp. 246–259, 1992.

[125] Rein, A., "Retroviral RNA packaging: a review," *Arch. Virol. Suppl.*, vol. 9, pp. 513–522, 1994.

[126] Rein, A., Henderson, L. E., and Levin, J. G., "Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication," *Trends Biochem. Sci.*, vol. 23, pp. 297–301, 1998.

[127] Rickgauer, J. P., Fuller, D. N., Grimes, S., jardine, J., Anderson, D. L., and Smith, D. E., "Portal motor velocity and internal force resisting viral dna packaging in bacteriophage $\phi$29," *biophys. J.*, vol. 94, pp. 159–167, 2008.

[128] Riemer, S. C. and Bloomfield, V. A., "Packaging of dna in bacteriophage heads: Some considerations on energetics," *Biopolymers*, vol. 17, pp. 785–794, 1978.

[129] Rocquigny, H. D., Gabus, H. C., Vincent, A., Fournie-Zaluski, M. C., Roques, B., and Darlix, J. L., "Viral RNA annealing activities of human immunodeficiency virus type 1 nucleocapsid protein require only peptide domains outside the zinc fingers," *Proc. Natl. Acad. Sci. USA*, vol. 89, pp. 6472–6476, 1992.

[130] Roda, R. H., Balakrishnan, M., Hanson, M. N., dot)hrl, B. M. W., Grice, S. F. J. L., Roques, B. P., Gorelick, R. J., and Bambara, R. A., "Role of the reverse transcriptase, nucleocapsid protein, and tempalte

structure in the two-step transfer mechanism in retroviral recombination," *J. Biol. Chem.*, vol. 278, no. 34, pp. 31536–2003, 2003.

[131] Rodriguez-Rodriguez, L., Tsuchihashi, Z., Fuentes, G. M., Bambara, R. A., and Fay, P. J., "Influence of human immunodeficiency virus nucleocapsid protein on synthesis and strand transfer by the reverse transcriptase in vitro," *J. Biol. Chem.*, vol. 270, no. 25, pp. 15005–15011, 1995.

[132] Rudnick, J. and Bruinsma, R. F., "Icosahedral packing of RNA viral genomes," *Phys. Rev. Lett.*, vol. 94, p. 038101, 2005.

[133] Saminathan, M., Antony, T., Shirahata, A., Sigal, L. H., Thomas, T., and Thomas, T. J., "Ionic and structural specificity effects of natural and synthetic polyamines on the aggregation and resolubilization of single-, double-, and triple-stranded DNA," *Biochem.*, vol. 38, pp. 3821–3830, 1999.

[134] Schnelder, M. F., Gange, S. J., Williams, C. M., Anastos, K., Greenblatt, R. M., Kingsley, L., Detels, R., and Munoz, A., "Patterns of the hazard of death after aids through the evolution of antiretroviral therapy: 1984-2004," *AIDS*, vol. 19, no. 17, pp. 2009–2018, 2005.

[135] Schwedler, U. K., Stemmler, T. L., Klishko, V. Y., Li, S., Albertine, K. H., Davis, D. R., and Sundquist, W. I., "Proteolytic refolding of the HIV-1 capsid protein amino-terminus facilitates viral core assembly," *The EMBO J.*, vol. 17, no. 6, pp. 1555–1568, 1998.

[136] Shklovskii, B. I., "Screening of a macroion by multivalent ions: Correlation-induced inversion of charge," *Phys. Rev. E*, vol. 60, p. 5802, 1999.

[137] Shklovskii, B. I., "Screening of a macroion by multivalent ions: Correlation-induced inversion of charge," *Phys. Rev. E*, vol. 60, pp. 5802–5811, 1999.

[138] Skinner, B. and Shklovskii, B. I., "Non-monotonic swelling of a macroion due to correlation-induced charge inversion," *Physica A*, vol. 388, pp. 1–8, 2009.

[139] Smith, D. E., Trans, S. J., Smith, S. B., Grimes, S., Anderson, D. L., and Bustamante, C., "The bacteriophage $\phi 29$ portal motor can package DNA against a large internal force," *Nature*, vol. 413, pp. 748–752, 2001.

[140] Spakowitz, A. J. and Wang, Z.-G., "Dna packaging in bacteriophage: Is twist important?," *Biophys. J.*, vol. 88, pp. 3912–3923, 2004.

[141] Stewart-Maynard, K. M., Cruceanu, M., Wang, F., Vo, M. N., Gorelick, R. J., Williams, M. C., Rouzina, I., and musier Forsyth, K., "Retroviral nucleocapsid proteins display noneuivalent levels of nucleic acid chaperone activity," *J. Virol.*, vol. 82, no. 20, pp. 10129–10142, 2008.

[142] Stoylov, S. P., Vuilleumier, C., Stoylova, E., Rocquigny, H. D., Roques, B. P., Gerard, D., and Mely, Y., "Ordered aggregation of ribonucleic acids by the human immunodeficiency virus type 1 nucleocapsid protein," *Biopolymers*, vol. 41, pp. 301–312, 1997.

[143] Tanchou, V., Gabus, C., Rogemond, V., and Darlix, J. L., "Formation of stable and functional HIV-1 nucleoprotein complexes in vitro," *J. Mol. Biol.*, vol. 252, pp. 563–571, 1995.

[144] Tang, L. and Johnson, J., "Structural biology of viruses by the combination of electron cryomicroscopy and x-ray crystallography," *Biochem.*, vol. 41, pp. 11517–24, 2002.

[145] Tang, L., Johnson, K. N., Ball, L. A., Lin, T., Yeager, M., and Johnson, J. E., "The structure of Pariacoto virus reveals a dodecahedral cage of duplex RNA," *Nat. Struct. Biol.*, vol. 8, pp. 77–83, 2001.

[146] Thomas, J. A., Gagliardi, T. D., Alvord, W. G., Lubomirski, M., Bosche, W. J., and Gorelick, R. J., "Human immunodeficiency virus type 1 nucleocapsid zinc-finger mutations cause defects in reverse transcription and integration," *Virology*, vol. 353, pp. 41–51, 2006.

[147] Thomas, J. A. and Gorelick, J. J., "Nucleocapsid protein function in early infection processes," *Virus Res.*, vol. 134, pp. 39–63, 2008.

[148] Thomas, J. A., Shulenin, S., Coren, L. V., Bosche, W. J., Gagliardi, T. D., Gorelick, R. J., and Oroszlan, S., "Characterization of human immunodeficiency virus type 1 (HIV-1) containing mutations in the nucleocapsid protein at a putative HIV-1 protease cleavage site," *Virology*, vol. 354, pp. 261–270, 2006.

[149] Tritch, R. J., Cheng, Y. E., Yin, F. H., and Erickson-viitanen, S., "Mutagenesis of protease cleavage sites in the human immunodeficiency virus type 1 gag polyprotein," *J. Virol.*, vol. 65, pp. 922–930, 1991.

[150] Tsuchihashi, Z. and Brown, P. O., "DNA strand exchange and selective DNA annealing promoted by the human immunodeficiency virus type 1 nucleocapsid protein," *J. Virol.*, vol. 68, pp. 5863–5870, 1994.

[151] Tsuruta, H., Reddy, V. S., Wikoff, W. R., and Johnson, J. E., "Imaging RNA and dynamic protein segments with low-resolution virus crystallography: experimental design, data processing and implications of electron density maps," *J. Mol. Biol.*, vol. 284, pp. 1439–1452, 1998.

[152] Twort, F. and Lond, L., "An investigation on the nature of ultra-microscopic viruses," *The Lancet*, vol. 2, pp. 1241–1243, 1915.

[153] Urbaneja, M. A., Kane, B. P., Johnson, D. G., gorelick, R. J., Henderson, L. E., and Casas-Finet, J. R., "Binding properties of the human immunodeficiency virus type 1 nucleocapsid protein p7 to a model RNA: elucidation of the structural determinants for function," *J. Mol. Biol.*, vol. 287, pp. 59–75, 1999.

[154] Urbaneja, M. A., Wu, M., Casas-Finet, J. R., and Karpel, R. L., "HIV-1 nucleocapsid protein as a nucleic acid chaperone: spectroscopic study of its helix-destabilizing properties, structural binding specificity, and annealing activity," *J. Mol. Biol.*, vol. 318, pp. 749–764, 2002.

[155] Valleau, J. P. and Cohen, L. K., "Primitive model electrolytes. I. Grand cononical Monte Carlo computations," *J. Chem. Phys.*, vol. 72, pp. 5935–5941, 1980.

[156] van der Schoot, P. and Bruinsma, R. F., "Electrostatics and the assembly of an rna virus," *Phys. Rev. E*, vol. 71, p. 061928, 2005.

[157] Veronese, F. D., Rahman, R., Copeland, T. D., Oroszlan, S., Gallo, R. C., and Sarngadharan, M. G., "Immunological and chemical analysis of p6, the carboxyl-terminal fragment of HIV p15," *AIDS Res. Hum. Retroviruses*, vol. 3, pp. 253–264, 1987.

[158] Vo, M. N., Barany, G., Rouzina, I., and Musier-Forsyth, K., "Mechanistic studies of mini-TAR RNA/DNA annealing in the absence and presence of HIV-1 nucleocapsid protein," *J. Mol. Biol.*, vol. 363, pp. 244–261, 2006.

[159] Wang, C. T. and Barklis, E., "Assembly, processing, and infectivity of human immunodeficiency virus type 1 gag mutants," *J. Virol.*, vol. 67, pp. 4264–4273, 1993.

[160] Williams, M. C., Gorelick, J. J., and Musier-Forsyth, K., "Specific zinc-finger architecture required for HIV-1 nucleocapsid protein's nucleic acid chaperone function," *Proc. Natl Acad. Sci. USA*, vol. 99, pp. 8614–8619, 2002.

[161] Williams, M. C., Rouzina, I., Wenner, J. R., Gorelick, R. J., Musier-Forsyth, K., and Bloomfield, V. A., "Mechanism for nucleic acid chaperone activity of HIV-1 nucleocapsid protein revealed by single molecule stretching," *Proc. Natl. Acad. Sci. USA*, vol. 98, pp. 6121–6126, 2001.

[162] Winterhalter, M. and Helfrich, W., "Effect of surface charge on the curvature elasticity of membranes," *J. Phys. Chem.*, vol. 92, pp. 6865–6867, 1988.

[163] Wu, W., Henderson, L. E., Copeland, T. D., Gorelick, R. J., Bosche, W. J., Rein, A., and Levin, J. G., "Human immunodeficiency virus type 1 nucleocapsid protein reduces reverse transcriptase pausing at a secondary structure near the murine leukemia virus polypurine tract," *J. Virol.*, vol. 70, pp. 7132–7142, 1996.

[164] Yoffe, A. M., Prinsen, P., andC. M. Knobler, A. G., Gelbart, W. M., and Ben-Shaul, A., "Predicting the sizes of large rna molecules," *Proc. Natl. Acad. Sci. USA*, vol. 105, no. 42, pp. 16153–16158, 2008.

[165] You, J. C. and McHenry, C. S., "Human immunodeficiency virus nucleocapsid protein accelerates strand transfer of the terminally redundant sequences involved in reverse transcription," *J. Biol. Chem*, vol. 269, pp. 31491–31495, 1994.

[166] Zhang, R. and Nguyen, T. T., "A model of HIV budding and self-assembly, role of cell membrane," *Phys. Rev. E*, vol. 78, p. 051903, 2008.

[167] Zhang, Y. and Barklis, E., "Nucleocapsid protein effects on the specificity of retrovirus RNA encapsidation," *J. Virol.*, vol. 69, pp. 5716–5722, 1995.