UNIVERSITY OF CALIFORNIA
RIVERSIDE

Physics of Viruses: The Role of Genome and Membrane

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Physics

by

Gonca Erdemci Tandogan

August 2016

Dissertation Committee:

Dr. Roya Zandi, Chairperson
Dr. Leonid Pryadko
Dr. Umar Mohideen

The Dissertation of Gonca Erdemci Tandogan is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I would like to express my special thanks to my advisor, Professor Roya Zandi. Without her support and guidance, none of this work would have been possible. Among many other things, Roya has always taught me to think about the fundamental physics no matter how complex the system is. I greatly appreciate that she has been always available to advise and help me. I am very grateful for her motivation and enthusiasm, and constant encouragement in every aspect of my PhD studies.

I would like to thank Dr. Jef Wagner for generously sharing his time and knowledge. Jef was always willing to answer my questions in details. I thank him for taking time to teach me many numerical techniques. I greatly appreciate that. I was very lucky to share an office with him during the early stage of my research.

I would also like to mention my appreciation to my committee members, Prof. Umar Mohideen and Prof. Leonid Pryadko. I would like to recognize our collaborators. I am grateful to Prof. Paul van der Schoot for his contributions to my research. I have learned a lot working with him in many projects. His questions and comments have always made me to think about the problems from different perspectives. I thank Prof. Rudolf Podgornik for sharing his knowledge on polymer physics and mean field theories. I greatly appreciate his contributions to my research. I also thank Prof. A.L.N. Rao and Dr. Venkatesh Sivanandam for the experiments they performed and for many helpful discussions.

I would like to acknowledge the support of the Fulbright Program for my PhD studies. I greatly appreciate their support before and during my studies.

To my friends, thank you for supporting me through these years. Thank you for the dinners, parties, coffee and tea breaks, workout breaks, Friday nights, "special" coffee supplies, endless chats and many more memories that I will never forget.

I thank my mom and dad for their tremendous support and love. I appreciate that they have always respected every decision I made in my life and supported me with their love and understanding for any path I wanted to take. I am very grateful to my sisters Aysegul and Pelin for their endless love and encouragement.

Last, I cannot begin to express my gratefulness to Murat who has been always supportive and proud. He has been by my side at every step of this journey. He has always believed in me. I wouldn't have made it this far without his love, encouragement and generosity. He is my rock and my source of happiness.

Chapter 2 is a version of J. Wagner, G. Erdemci-Tandogan and R. Zandi, "Adsorption of annealed branched polymers on curved surfaces", J. Phys.: Condens. Matter 27, 495101, (2015).

Chapter 3 is a version of G. Erdemci-Tandogan, J. Wagner, P. van der Schoot, R. Podgornik and R. Zandi, "RNA topology remolds electrostatic stabilization of viruses", Phys. Rev. E 89, 032707, (2014).

Chapter 4 is a version of G. Erdemci-Tandogan, J. Wagner, P. van der Schoot, R. Podgornik, and R. Zandi, "Effects of RNA branching on the electrostatic stabilization of viruses", Phys. Rev. E 94, 022408, (2016).

Chapter 5 is a version of V. Sivanandam, D. Mathews, R. Garmann, G. Erdemci-Tandogan, R. Zandi and A.L.N. Rao, "Functional analysis of the N-terminal basic motif

of a eukaryotic satellite RNA virus capsid protein in replication and packaging", Scientific Reports 6, 26328, (2016).

Chapter 6 is a version of G. Erdemci-Tandogan, J. Wagner, P. van der Schoot and R. Zandi, "Role of genome in the formation of conical retroviral shells", J. Phys. Chem. B 120, 6298, (2016).

Chapter 7 is a version of J. Ning*, G. Erdemci-Tandogan*, E. L. Yufenyuy*, J. Wagner, B. A. Himes, G. Zhao, C. Aiken, R. Zandi and P. Zhang, "In vitro Protease Cleavage and Computer Simulations Reveal the HIV-1 Capsid Maturation Pathway", Submitted (2016).

To my parents,

my sisters,

and Murat.

ABSTRACT OF THE DISSERTATION

Physics of Viruses: The Role of Genome and Membrane

by

Gonca Erdemci Tandogan

Doctor of Philosophy, Graduate Program in Physics
University of California, Riverside, August 2016
Dr. Roya Zandi, Chairperson

We study the physics of virus assembly. The dissertation can be separated into two parts. The first part focuses on spherical single stranded (ss) RNA viruses or virus like particles. Using mean-field theory, we explore the role of the secondary structure of RNA on the viral assembly by modeling the RNA as an annealed branched polymer. Our results verify that RNA branchedness maximizes the amount of encapsulated genome into a relatively small capsid and makes assembly more efficient. We furthermore offer an explanation for the phenomena of overcharging observed in viral particles. Chapter 5 demonstrates an application of our theory to satellite tobacco mosaic virus (STMV). Our theory explains, energetically, why a truncated RNA is encapsidated by STMV capsid proteins more favorably.

In the second part, we explore the role of genome on the structure of human immunodeficiency virus (HIV) shells (Chapter 6). Our analytical results show that the free energy of confinement of genome into a conical capsid is less than that for a cylindrical one when the genome does not interact with the capsid as in *in vivo* experiments. This may

explain why the conical structures favor over the cylindrical ones in *in vivo*. In Chapter 7, we carry out coarse-grained simulations to examine the HIV maturation pathways and the role of genome and membrane on the formation of conical shells. We show that the membrane restricts the growth of the otherwise extended incomplete shell and induces local stresses causing formation of the pentamers necessary for the assembly of closed conical or tubular shells. More interestingly, we find that any asymmetry developed in the growing lattice due to interaction with the membrane or genome, or due to the shape of initial immature lattice creates conical capsids, as opposed to cylindrical shells. Furthermore, our work confirms that viruses employ all the accessible pathways to maturation, explaining many aspects of the previous HIV pathway experiments.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Viruses are the simplest physical objects in biology. They are composed of an infectious nucleic acid (genome) either DNA or RNA and a protective protein shell, called the capsid. They amazingly share a characteristic feature of living organisms: they reproduce. They do not have a cell membrane and a host cell is required for reproduction. Outside of the host they act like non-living chemicals by not showing any metabolic activity and they can be crystallized. Besides their beautifully simple contents, spherical viruses, measuring across between 15 to 28 nm, are highly symmetric which many of them adopt icosahedral symmetry [1, 2]. Capsid proteins (CP) of some viruses can also form cylindrical, conical and irregular shells around the genetic materials [3]. Viruses can be categorized according to their capsid shapes, genome types (DNA or RNA) and genome architecture (double stranded (ds), single stranded (ss), etc.) and whether they have a lipid membrane (envelope) around their capsid or not [1].

Remarkably, for many (ss) RNA viruses, multiple of 60 copies of CP spontaneously assemble around the nucleic acid to build the icosahedral protein shell. Cowpea chlorotic mottle virus (CCMV) is the first spherical virus that was reconstituted from its purified components *in vitro* by Bancroft and Hiebert in 1967 [4]. Later Bancroft *et al.* demonstrated CCMV CP can package RNA of other viruses as well as non-viral RNAs suggesting specific nucleotide sequence is not required for the shell formation [5]. The studies of the assembly of viruses over the last decades have established that generic electrostatic interactions between the negative charges on the phosphate backbone of RNA and the positive charges on the arginine-rich motif (ARM) of CP are the driving force for the spontaneous encapsidation. Recent experiments have indeed verified the importance of the "charge-matching" hypothesis for proper genome packaging [6]. In contrast, some studies suggest the "packaging signals" hypothesis that the nucleation of the capsid is initiated by specific interactions between the coat proteins and specific stem-loops in the secondary structure of the viral RNA [7, 8]. Furthermore, experiments between two different competing viruses suggest that there must be a correlation between the nucleic acid structure and the assembly efficiency [9, 10]. For example, when viral RNA1 of BMV (brome mosaic virus) and CCMV (cowpea chlorotic mottle virus) are mixed in solution with the capsid proteins from CCMV, the BMV RNA is packaged three times more efficiently [9]. Although both RNAs are of similar length, BMV RNA was shown to outcompete the CCMV RNA, therefore suggesting that electrostatics alone is not enough for efficient genome encapsidation and that further structural details of RNA, apart from its generic charge, could play a role in the genome encapsidation.

With the inevitable importance of RNA on the virus life cycle, in the first part of this dissertation, we focus on the role of RNA on the assembly and efficiency of spherical RNA viruses. Since viral RNAs are very compact inside their shells and highly branched due to the base-pairing between the nucleotides, we model RNA as a branched polymer.

In Chapter 2, we introduce the mean-field theories describing branched polymers and explore the adsorption of RNA onto gold-nanoparticles, a very important problem, due to its potential application in drug delivery or gene therapy. In addition to nanoparticles, we numerically and analytically study the behavior of annealed branched polymers next to adsorbing flat walls, spheres and sinusoidal walls. We calculate the surface tension, surface excess and polymer density as a function of the degree of branching which could find applications in material science and nanotechnology.

In Chapters 3 and 4, we examine the impact of RNA base-pairing (secondary structure) on the stability of virus or virus like particles. We not only show that the presence of secondary structure makes the assembly process more efficient through the calculation of relevant free energies but offer an explanation for the phenomena of overcharging observed in viral particles. We explore how the secondary structure of the RNA affects the osmotic pressure of spherical RNA viruses. These findings provide insights into the mechanism of virion packaging which could advance the development of antiviral drugs.

In Chapter 5, we apply our theory to explain the experimental observations of Satellite Tobacco Mosaic Virus (STMV). We study the role in the virion self-assembly of the positive charges sitting on the N-terminal domain (ARM) of STMV CP. In the experiments, the N-terminal domain was mutated to remove one of the positive charges at a time to

investigate the importance of position of charges along the N-terminal tails. This created defects in packaging of the full-length of STMV RNA, but in contrast a truncated RNA was encapsulated more efficiently. Our theory explains, energetically, why the truncated RNA is encapsidated more favorably.

For the rest of the dissertation, we study the physics of assembly of human immunodeficiency virus (HIV). HIV is one of the most studied viruses due to its connection to the AIDS pandemic. Understanding the structure and details of the life cycle of HIV is a challenging yet very intriguing subject. Immature HIV is an incomplete hexagonal spherical shell made of Gag proteins [11, 12]. During maturation, the viral protease cleaves Gag into three sections, the matrix (MA), capsid (CA), and nucleocapsid (NC), resulting into significant structural changes and the formation of a mature conical shell [11, 12]. Mature HIV self-assembles predominantly into conical capsids *in vivo* [13]. However, *in vitro* self-assemble studies reveal that the cylindrical capsids form more than the conical ones [14].

In Chapter 6, we seek to determine the factors that contribute to the selection of the unique shape of HIV capsids. We numerically and analytically investigate the free energy of genome inside cylindrical and conical capsids by taking into account the genome-capsid interactions. We find that the confinement free energy of the same length genome into a conical capsid is smaller than that into a cylindrical capsid when the genome does not interact with the capsid as in *in vivo* experiments. This may explain why the conical structures predominates over the cylindrical ones in *in vivo*. In the case in which the genome

interacts with the capsid, as in *in vitro* studies, we show that cylindrical capsids favor over the conical ones having less confinement energy for the same genome length.

In Chapter 7, we explore the HIV kinetic pathways to maturation. Two models have been proposed for the maturation: *de novo* (disassembly/reassembly) and displacive (non-diffusive) models[3, 15, 16, 17, 18, 19]. For the first model, the immature lattice is fully disassembled after proteolysis followed by *de novo* reassembly of a mature capsid lattice. For the latter, immature lattice is transformed into the mature capsid directly by non-diffusional remodeling of the immature lattice to the mature capsid lattice. We develop a coarse-grained simulation and monitor the viral capsid growth. We not only elucidate the impact of membrane and genome on the structural choice of the virion but also tackle the maturation pathway. Our work confirms that viruses employ all the accessible pathways to maturation, explaining many aspects of the previous HIV pathway experiments. While our model can show that *de novo* assembly is possible, our combined experimental and numerical modeling show that a combination of displacive and *de novo* models also exists. Understanding the factors involved with the selection of unique shapes of HIV or its maturation pathway is of vital importance as a step toward on the development of antiviral strategies against the infectivity of the virus.

Chapter 2

Adsorption of annealed branched polymers on curved surfaces

2.1 Introduction

Branched polymers play important roles in many biological and industrial systems, notable among them single stranded RNA (ssRNA) that in solution takes on a branched secondary structure [20, 21, 22, 23, 24, 25, 26, 27, 28, 29]. Recent experiments on viruses show that some viral RNAs, in particular, assume highly branched structures [30, 31]. The physics of polymer adsorption on different kinds of interfaces has, specifically, attracted a lot of interest for over half a century [32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43]. In particular, it has been shown that polymer topology can effect the thermodynamic behavior of polymers near surfaces [44]. More recently the adsorption of RNA onto spherical gold

nano-particles has been the focus of intense research because of its potential application in drug delivery or gene therapy [45, 46, 47, 48, 49, 50, 51, 52, 53].

RNA is considered as an *annealed* branched polymer mainly due to the fact that the base-pair binding in RNA is often weak enough that the branching structure can change due to thermal fluctuations [21, 54]. Beyond the adsorption of RNA onto nano-particles, the behavior of annealed branched polymers next to surfaces of complex geometries is intriguing. Despite the presence of several excellent books and review articles, the impact of branching on adsorption of biopolymers at planar or rough substrates is yet not well-studied.

Several experiments compare the efficiency–directly connected to the free energy– of encapsidation of linear polymers and viral RNAs by virus capsid proteins[9, 55, 56]. Field theoretic models have been used extensively to calculate the free energy of linear polymers [57, 58, 59, 60, 61, 62, 63, 64, 65]. In a 1972 seminal paper de Gennes noted an equivalence between the statistics of a self-avoiding polymer and the $n \to 0$ limit of an $\mathcal{O}(n)$ model of a magnet [66], see Appendix A for a review of $\mathcal{O}(n)$ model. Using this observation, which relates a mathematically interesting but unphysical limit for the model of a magnet to the statistics of polymers, it became possible to use the tools of statistical field theory to describe the physical properties of a polymer solution. Later, de Gennes field theory was expanded to describe the statistics of *annealed* branched polymers [67, 68].

In this paper we use a mean field theory to study the adsorption of annealed branched polymers on different types of surfaces from a semi-dilute polymer solution in a good solvent. We study the effect of curvature by examining the adsorption onto the exterior and interior of a spherical surface, and investigate the impact of roughness by examining

the adsorption onto a sinusoidal surface. Instead of considering a random roughness, we employ a grating geometry because of its enormous mathematical simplifications and the fact that it has been shown that qualitatively the essential features of the results are the same [42, 60].

By numerically solving the relevant nonlinear equations we find that compared to the adsorption to a flat wall, branching density, surface tension, and the monomer density all increase if the polymer is adsorbed onto the interior wall of a spherical cavity but decrease if adsorbed on the exterior surface of the sphere. While our results show that surface tension always decreases as branching density increases independent of the geometry of the wall, we find the interplay of curvature and branching density conspires to further lower the surface tension when the wall curves toward the polymers but lessens the amount of decrease in the surface tension when the wall curves away from the polymer. In the limit of large spheres, we solve the nonlinear equations perturbatively, which match the numerical results. Furthermore, in case of sinusoidal surfaces, we find inhomogeneity in the branching density as it increases in the valleys but decreases in the peaks.

The remainder of this paper is organized as follows. In the first section we present our mean field approach and in the following section we use this method to investigate the impact of branching combined with surface curvature on polymer adsorption. In particular we will examine what effect the branching structure has on the adsorption of polymers to nano-spheres and sinusoidal surfaces. We will finish with a brief summary and present our main conclusions.

In the appendix, for completeness and pedagogical reasons, we derive a simple field theoretical model for a branched polymer by revisiting the field theory developed by Isaacson and Lubensky[67] for branched polymers and will spell out in detail the equivalence between the polymer statistics and the $n \to 0$ limit of the $\mathcal{O}(n)$ model.

2.2 Mean Field Approximation

To describe a branched polymer on a lattice, we assume the polymer system consists of branch-points and end-points lying on the lattice sites, and bonds that join neighboring lattice sites. We treat the system as an annealed branched polymer, so the structure of the branched polymer is not fixed. For simplicity, we assume that all branch-points are exactly of order three because all higher order branch-points can be considered as many order three branch-points in close proximity to each other. For example, two order three branch-points close together will behave very similarly to an order four branch-point. The quantities that describe such a polymer system are: (i) N_p, the number of polymers; (ii) N_b, the number of bonds; (iii) N_1, the number of endpoints; (iv) N_3, the number of branch-points; and (v) N_l, the number of loops. There is a constraint [67] relating most of these quantities such that

$$2(N_p - N_l) = N_1 - N_3. \tag{2.1}$$

In this paper, we consider a system of branched polymers with no loops and set $N_l = 0$.

The primary statistical quantity of interest is the multiplicity $\Omega(N_b, N_1, N_3; V)$, defined as the number of ways to arrange a polymer system of N_b bonds, N_1 end-points, and N_3 branch-points on a lattice that occupies a volume V. This quantity is equivalent to

the number of microstates for the microcanonical ensemble. From the multiplicty, we can form the grand canonical partition function

$$\Xi(K, f_1, f_3; V) = \sum_{N_b, N_1, N_3} K^{N_b} f_1^{N_1} f_3^{N_3} \Omega(N_b, N_1, N_3; V), \tag{2.2}$$

where K, f_1, and f_3 are the fugacities for the bonds, end-points, and branch-points respectively. Note that we use Eq. (2.1) along with the assumption that $N_l = 0$ to eliminate the dependence upon N_p. From this definition it is simple to derive expressions for the number of bonds, end-points, and branch-points as derivatives of the logarithm of the grandpartition function

$$N_b = K \frac{\partial}{\partial K} \ln \Xi(K, f_1, f_3), \tag{2.3}$$

$$N_1 = f_1 \frac{\partial}{\partial f_1} \ln \Xi(K, f_1, f_3), \tag{2.4}$$

$$N_3 = f_3 \frac{\partial}{\partial f_3} \ln \Xi(K, f_1, f_3). \tag{2.5}$$

Following the idea of de'Gennes [66], and using the methods of Lubensky and Isaacson [67], we equate the grand partition function for the branched polymers system with the $n \to 0$ limit of the partition function of an $\mathcal{O}(n)$ model of a magnet

$$\Xi(K, f_1, f_3; V) \approx \mathcal{Z}(K, f_1, f_3; V). \tag{2.6}$$

The partition for the $\mathcal{O}(n)$ model of a magnet can be written as a function integral over a continuous field $\psi(x)$

$$\mathcal{Z}(K, f_1, f_3; V) = \int \mathcal{D}\psi \, \exp\big(- \beta \mathcal{H}(\psi, K, f_1, f_3; V) \big). \tag{2.7}$$

where $\mathcal{H}$ is the effective Hamiltonian

$$\mathcal{H}(\psi, K, f_1, f_3; V) = \sum_x \left(\frac{1}{2}\boldsymbol{\psi}(\boldsymbol{x}) \sum_{x'} \delta^{-1}_{\langle x,x' \rangle} \cdot \boldsymbol{\psi}(\boldsymbol{x}') \right.$$
$$- \frac{K}{2}|\boldsymbol{\psi}(\boldsymbol{x})|^2 + \frac{K^2}{8}|\boldsymbol{\psi}(\boldsymbol{x})|^4$$
$$\left. - f_1\sqrt{K}\psi_1(\boldsymbol{x}) - \frac{f_3 K^{\frac{3}{2}}}{6}\psi_1^3(\boldsymbol{x}) \right). \quad (2.8)$$

We emphasize here that the parameters K, f_1 and f_3 take on different meanings for the $\mathcal{O}(n)$ model of the magnet (see Appendix A). For example, the K parameter represents the coupling constant between the nearest neighbors spins in the $\mathcal{O}(n)$ model of a magnet and the $\boldsymbol{\psi}$ field is the average magnetization in a small region for the magnet. However, the $\boldsymbol{\psi}$ field in Eq. (2.8) is proportional to the monomer density for the branched polymer. The first term in Eq. (2.8) is an entropic term that smoothes out the $\boldsymbol{\psi}(x)$ field. The $\delta^{-1}_{\langle x,x' \rangle}$ in the first term is the inverse of the nearest neighbor operator $\delta_{\langle x,x' \rangle}$, defined as 1 if x and x' are neighboring sites on a lattice and 0 otherwise. The second term in Eq. (2.8) proportional to $|\boldsymbol{\psi}|^2$ is due to the nearest neighbor attraction, which tends to increase the $\boldsymbol{\psi}(x)$ field. The third term proportional to $|\boldsymbol{\psi}|^4$ is repulsive representing the self-avoiding nature of the polymer, and leads to a decrease in the $\boldsymbol{\psi}(x)$ field. The fourth and fifth terms proportional to $\psi_1(x)$ and $\psi_1^3(x)$ are both attractive, and show that both branch-points and endpoints serve to increase the $\psi(x)$ field. It is the balance of these attractive and repulsive terms that creates a well-defined finite $\psi(x)$ field in equilibrium. For completeness, as well as pedagogical reasons, we show the derivation of the equivalence between the grand canonical partition function, Eq. 2.2, for the polymer system and the partition function for

the $n \to 0$ limit of the $\mathcal{O}(n)$ model and all relevant approximations for finding the effective Hamiltonian, Eq. 2.8, in Appendix A.

We now make a mean field approximation and assume that the value of the field $\psi(x)$ is uniform and is well approximated by its average value ψ_0. Thus the sum over the nearest neighbors simply becomes

$$\sum_{x}^{V} \delta_{\langle xx' \rangle} \psi(x') = z\psi_0, \tag{2.9}$$

with z the number of nearest neighbors or coordination number. The inverse of the nearest neighbor delta function is then simply the reciprocal of the coordination number $\sum_x^V \delta^{-1}_{\langle xx' \rangle} \psi(\boldsymbol{x}) = \frac{1}{z}\psi_0$. So in the mean field theory, the grand canonical partition function is the exponential of the effective Hamiltonian evaluated at its minimum

$$\Xi(K, f_1, f_3; V) \approx e^{-\mathcal{H}(\psi_0, K, f_1, f_3; V)}, \tag{2.10}$$

with ψ_0 found by minimizing Eq. (2.8), $\delta\mathcal{H}/\delta\psi|_{\psi_0} = 0$,

$$\left(\frac{1}{z} - K\right)\psi_0 - f_1\sqrt{K} - \frac{f_3 K^{\frac{3}{2}}}{2}\psi_0^2 + \frac{K^2}{2}\psi_0^3 = 0. \tag{2.11}$$

From stastical mechanics we can identify the grand potential, $\beta\Phi_G(K, f_1, f_3; V) = -\,ln[\Xi(K, f_1, f_3; V)]$ using Eqs. (2.8), (2.9) and (2.10) ,

$$\beta\Phi_G(K, f_1, f_3; V) \;=\; \frac{V}{a^3}\left[\frac{1}{2}\left(\frac{1}{z} \;-\; K\right)\psi_0^2 \;-\; f_1\sqrt{K}\psi_0 \;-\; \frac{f_3 K^{\frac{3}{2}}}{6}\psi_0^3 \;+\; \frac{K^2}{8}\psi_0^4\right], \tag{2.12}$$

with a the lattice spacing. Inserting Eq. (2.12) in Eq. (2.3) and using Eq. (2.11) we find the average monomer density

$$c_b = N_b/V = \frac{\psi_0^2}{2za^3}. \tag{2.13}$$

At this point it is convenient to define a new field $\phi(x)$ such that in the mean field approximation the average value $\phi_0^2 = \psi_0^2/2za^3$ is the monomer density. The grand potential written in terms of the new field is

$$\frac{\beta\Phi_G(K, F_1, F_3; V)}{V} = (1 - zK)\phi_0^2 - f_1\sqrt{\frac{2zK}{a^3}}\phi_0 - \frac{1}{6}f_3(2zaK)^{\frac{3}{2}}\phi_0^3 + \frac{1}{2}K^2z^2a^3\phi_0^4. \quad (2.14)$$

By comparing Eq. (2.14) with the expression for the grand potential for a linear polymer in good solvent [60], we can identify $r = (1 - zK)$ as the chemical potential of monomers (such that $\partial\beta\Phi_G/\partial r = c_b$) and $v = K^2z^2a^3$ as the excluded volume. It is also convenient to absorb $\sqrt{2zK}$ and $(2zK)^{\frac{3}{2}}$ constants in the end and branch point fugacities, respectively such that the grand potential can be written in a much simpler form

$$\frac{\beta\Phi_G(K, f_1, f_3; V)}{V} = r\phi_0^2 - \frac{f_1}{\sqrt{a^3}}\phi_0 - \frac{f_3}{6}\sqrt{a^3}\phi_0^3 + \frac{v}{2}\phi_0^4. \quad (2.15)$$

The parameters f_1 and f_3 are to physical quantities. Using Eqs. (2.4) and (2.5), we find the end-point and branch-point densities

$$c_1 = \frac{f_1}{\sqrt{a^3}}\phi_0, \quad (2.16)$$

$$c_3 = \frac{1}{6}f_3\sqrt{a^3}\phi_0^3. \quad (2.17)$$

2.3 Results and Discussion: Branched polymers adsorption onto different surfaces

We now apply the field theory presented in the previous section to a semi-dilute system of annealed branched polymers and investigate their adsorption to different surfaces. More specifically, we consider a solution of branched polymers with a monomer density c_b, where the polymers all have a fixed length L, and a tunable average branching number N_b.

The adsorption mean field energy of the branched polymer, $F - F_0$ can then be written as

$$F - F_0 = -\gamma a^3 \int dS\phi^2 + \int dV \bigg(\frac{a^2}{6\beta}(\nabla\phi)^2 + \frac{1}{2}\frac{\nu}{\beta}(\phi^4 - c_b^2) \\ - \frac{1}{\beta\sqrt{a^3}}(f_1(\phi - c_b^{1/2}) + f_3\frac{a^3}{6}(\phi^3 - c_b^{3/2})) \bigg) \quad (2.18)$$

The first term in Eq. (2.18) is a surface integral that gives the contact energy between the surface and the polymer. The first term in the volume integral is associated with the entropic cost of a non-uniform polymer distribution[69]. The rest of the terms in Eq. (2.18) are related to the free energy of a branched polymer in mean field approximation, see Eq. (2.15).

Considering the constraint that the total number of monomers is fixed

$$\int dV\ \phi^2 = constant = N, \quad (2.19)$$

Eq. (2.18) can be rewritten as

$$F - F_0 = -\gamma a^3 \int dS\phi^2 + \int dV \bigg(\frac{a^2}{6\beta}(\nabla\phi)^2 + \frac{1}{2}\frac{\nu}{\beta}(\phi^4 - c_b^2) \\ - \frac{1}{\beta\sqrt{a^3}}(f_1(\phi - c_b^{1/2}) + f_3\frac{a^3}{6}(\phi^3 - c_b^{3/2})) - \lambda(\phi^2 - c_b) \bigg) \quad (2.20)$$

where λ is the Lagrange multiplier.

Minimizing Eq. (2.20) with respect to the field $\phi(x)$ gives the following Euler-Lagrange equation

$$\frac{a^2}{6}\nabla^2\phi = -\lambda\phi + \nu\phi^3 - \frac{1}{2\sqrt{a^3}}(f_1 + \frac{a^3}{2}f_3\phi^2) \quad (2.21)$$

subject to the boundary condition

$$[\hat{n}.\nabla\phi + (6\beta a\gamma)\phi]_s = 0 \quad (2.22)$$

For simplicity, we rescale the Lagrange multiplier $E = \frac{6\lambda}{a^2}$ and introduce a length that characterizes the strength of the attraction between the surface and monomers as $\kappa^{-1} = \frac{1}{6\beta a\gamma}$. The other boundary condition is natural, far from the surface the field should be uniform, $\nabla^2\phi \to 0$, and and take on the bulk value $\phi(x) \to \sqrt{c_b}$. Using Eq. 2.21, the Lagrange multiplier can be written as

$$E = \frac{6}{a^2}(\nu c_b - \frac{1}{2}\frac{c_1}{c_b} - \frac{3}{2}\frac{c_3}{c_b}) \tag{2.23}$$

where $c_1 = f_1 c_b^{1/2}/\sqrt{a^3}$ and $c_3 = f_3\sqrt{a^3}c_b^{3/2}/6$ are respectively the end-point and branch-point concentrations far from the surface. For long polymers with no loops in which $N_1 = N_3 + 2$, the second and third terms in Eq. (2.23) correspond to the ratio of the end-points and branch-points to monomers. To make all the quantities dimensionless, we rescale the field with respect to the bulk value $\theta(x) = \phi(x)/\sqrt{c_b}$ and the spatial coordinate with respect to the Edwards correlation length[70] $x = \tilde{x}\xi_E$, where $\xi_E = \frac{a}{\sqrt{3c_b v}}$, the equation of motion simply becomes

$$\tilde{\nabla}^2\theta = 2(\theta^3 - \theta) + A_1(\theta - 1) + 3A_3(\theta - \theta^2) \tag{2.24}$$

with $A_1 = (c_1/c_b)/(\nu c_b)$, $A_3 = (c_3/c_b)/(\nu c_b)$.

The A_1 and A_3 quantities measure the relative importance of the branching structure of the polymer to the steric effect or excluded volume interaction between monomers in solution . The numerators c_1/c_b and c_3/c_b are the ratios of the concentrations of end-points and branch-points to the total number of monomers, respectively. For large polymers, these ratios can approach 1/2 for a maximally branched polymer. The denominator vc_b is a filling fraction of the polymer; dilute solutions will have small values of vc_b, and a dense polymer system with no solvent will have a value of one. With the new scaled coordinates, the

boundary conditions become

$$\left[\frac{\partial\theta}{\partial\tilde{n}} + \tilde{\kappa}\theta\right]_s = 0 \tag{2.25}$$

$$\lim_{x\to\infty} \theta = 1 \tag{2.26}$$

with $\tilde{\kappa} = \kappa\xi_E$. In terms of the the new mean-field order parameter, θ, we can rewrite Eq. 2.20, the adsorption energy, as

$$F - F_0 = \frac{a^2}{6\beta}\xi_E c_b \Bigg(- \tilde{\kappa}\int d\tilde{S}\theta^2 + \int d\tilde{V}\big((\tilde{\nabla}\theta)^2 + (\theta^2-1)^2 \\ - A_1(\theta-1) - A_3(2\theta^3 - 3\theta + 1)\big)\Bigg) \tag{2.27}$$

In the following sections we employ Eq. (2.27) to calculate the free energy of a polymer next to a flat, curved, spherical and sinusoidal surfaces. We note that while the Lagrange multiplier λ acts like a chemical potential in open surfaces for fixing the density of bulk polymers, in the closed systems, like inside a sphere, it's used to fix the number of monomers inside the shell.

Analytical Calculations

Flat wall

Next to a flat wall, the Euler-Lagrange equation, Eq. (2.24), subject to the boundary conditions given in Eqs. (2.25) and (2.26) can be solved perturbatively. We assume the attractive interaction between monomers and wall is smaller than the monomer-monomer repulsion ($\tilde{\kappa} \ll 1$). The solution to Eq. (2.24) to the second order in $\tilde{\kappa}$ can then be written as

$$\theta \approx 1 + \frac{\tilde{\kappa}}{A}(1 + \frac{\tilde{\kappa}}{A})e^{-A\tilde{z}} \tag{2.28}$$

where $A = \sqrt{4 + A_1 - 3A_3}$ with A_1 and A_3 proportional to the number of end (N_1) and branch (N_3) points as given below Eq. (2.24). For a single long polymer with no loops Eq. (2.1) yields $N_1 = N_3 + 2$. So we can simply write $A_1 - 3A_3 = 2V(1 - N_3)/(\nu N^2)$ implying that $A_1 - 3A_3 < 0$ and $|A_1 - 3A_3| < 4$ should be satisfied for real solutions. For $A_1 = A_3 = 0$, Eq. (2.28) converges to the profile of a linear polymer next to the flat wall [60]. As clearly shown in Eq. (2.28), the density of branched polymers are larger than the linear ones next to a flat wall.

Inserting Eq. (2.28) into Eq. (2.27), we find the change in surface tension, energy per unit area, to the second order in $\tilde{\kappa}$

$$\gamma - \gamma_0 \approx \frac{a^2 c_b}{6\beta\xi_E}\big(- \tilde{\kappa} - \frac{\tilde{\kappa^2}}{2} - \Gamma_b(\tilde{\kappa}, A_1, A_3)\big) \tag{2.29}$$

where Γ_b is the difference in tension from a linear chain due to branching and is given by

$$\Gamma_b(\tilde{\kappa}, A_1, A_3) = (\frac{A_1 + 3A_3}{A^2})\tilde{\kappa} + (\frac{3A^2 - A^3 - 4 + 2A_1 + 12A_3}{2A^3})\tilde{\kappa}^2 \tag{2.30}$$

Since the quantity Γ_b for all acceptable values of A_1 and A_3 is always positive, according to Eq. (2.29) the surface tension due to adsorption of a linear polymer to a flat wall is always higher than that of a branched one. In the next section, we calculate the impact of curvature on the polymer density profile and the free energy of the system.

Curved wall

To investigate the effect of curvature analytically, we assume that the flat wall is slightly bent to form a large sphere. The curvature could be either toward or away from the polymer. The radius of curvature b is considered to be large compared to the correlation length ($\tilde{b} = b/\xi_E \gg 1$). We can obtain the perturbative solutions in spherical coordinates through Eq. 2.24 assuming that $\theta = 1 + \delta$ ($\delta \ll 1$) at the weak adsorption limit ($\tilde{\kappa} \ll 1$). According to the direction of wall curvature, polymer solution is considered either to be in the interior (in) or the exterior (out) of a sphere of radius b. The perturbative solutions are then

$$\theta_{out}(\tilde{r}) \approx 1 + \frac{\tilde{\kappa}\tilde{b}}{1 + A\tilde{b} - \tilde{\kappa}\tilde{b}} \big(\tfrac{\tilde{b}}{\tilde{r}}\big) e^{-A(\tilde{r}-\tilde{b})} \tag{2.31}$$

for ($\tilde{b} < \tilde{r} < \infty$) and

$$\theta_{in}(\tilde{r}) \approx 1 + \tilde{\kappa}\tilde{b}\big(\tfrac{\tilde{b}}{\tilde{r}}\big) \frac{\sinh A\tilde{r}}{A\tilde{b}\cosh(A\tilde{b}) - \sinh(A\tilde{b})(1 + \tilde{\kappa}\tilde{b})} \tag{2.32}$$

for ($0 < \tilde{r} < \tilde{b}$). Asumming $\tilde{b} \gg 1$, we can write the monomer concentration on the surface to the second order in $\tilde{\kappa}$

$$\theta_{out} \approx 1 + \frac{\tilde{\kappa}}{A}\big(1 - \frac{1}{A\tilde{b}} + \frac{\tilde{\kappa}}{A}\big) \tag{2.33}$$

$$\theta_{in} \approx 1 + \frac{\tilde{\kappa}}{A}\big(1 + \frac{1}{A\tilde{b}} + \frac{\tilde{\kappa}}{A}\big) \tag{2.34}$$

Comparison of Eqs. (2.33) and (2.34) with Eq. 2.28 reveals that branched polymer concentration in the vicinity of a flat wall increases if the wall bends toward the polymer and decreases if the wall bends away from the polymer, consistent with the results obtained for linear polymers [60]. In order to obtain the change in surface tension due to the wall

curvature, we insert the concentration profiles given in Eqs. (2.31) and (2.32) into Eq. (2.27) and keep terms up to the first order in $1/\tilde{b}$. Then we have

$$(\gamma - \gamma_0)_{out} \approx \gamma - \gamma_0 + \frac{a^2 c_b}{6\beta\xi_E}\Big(\frac{\tilde{\kappa}^2}{4\tilde{b}} + \Gamma_g(\tilde{\kappa}, \tilde{b}, A_1, A_3)\Big) \tag{2.35}$$

and

$$(\gamma - \gamma_0)_{in} \approx \gamma - \gamma_0 - \frac{a^2 c_b}{6\beta\xi_E}\Big(\frac{\tilde{\kappa}^2}{4\tilde{b}} + \Gamma_g(\tilde{\kappa}, \tilde{b}, A_1, A_3)\Big) \tag{2.36}$$

where $\gamma - \gamma_0$ is the surface tension for the polymer next to a flat wall based on Eq. (2.29) and $\frac{\tilde{\kappa}^2}{4\tilde{b}}$ is the difference in surface tension due solely to the geometry of the wall. The quantity

$$\Gamma_g(\tilde{\kappa}, \tilde{b}, A_1, A_3) = \frac{8A^2 - A^4 - 16 + 4A_1 + 36A_3}{4A^4\tilde{b}}\tilde{\kappa}^2 \tag{2.37}$$

is due to the the coupling between the geometry and branched structure of the polymers. If we set $A_1 = 0$ and $A_3 = 0$ then $A = 2$ and Eqs. (2.35) and (2.36) reveal the impact of wall curvature on the surface tension for the case of linear polymers. As expected, the surface tension decreases if the wall bends toward the polymer and it increases if the wall bends away from the polymer.

Quite interestingly, the Γ_g expression shows the importance of coupling between wall curvature and polymer branching on the surface tension. A glance through Table 1. shows that the sum or difference $\Gamma_b \pm \Gamma_g$ corresponds to the change in the surface tension due to the coupling between branching and wall curvature. Since this term is positive for both the convex and concave surfaces, we see that branching always decreases the surface tension. However, the coupling between branching and curvature further lowers the surface

	Concentration	Surface Tension
Flat surface	$\theta_{\mathbf{flat}} = 1 + \frac{\widetilde{\kappa}}{A}\big(1 + \frac{\widetilde{\kappa}}{A}\big)$	$\Delta\gamma_{\mathbf{flat}} \approx \frac{a^2 c_b}{6\beta\xi_E}\big(- \tilde{\kappa} - \frac{\tilde{\kappa^2}}{2} - \Gamma_b\big)$
Convex surface	$\theta_{\mathbf{out}} = \theta_{\mathbf{flat}} - \frac{\widetilde{\kappa}}{A^2\widetilde{b}}$	$\Delta\gamma_{\mathbf{out}} \approx \Delta\gamma_{\mathbf{flat}} + \frac{a^2 c_b}{6\beta\xi_E}\big(\frac{\widetilde{\kappa}^2}{4\widetilde{b}} + \Gamma_g\big)$
Concave surface	$\theta_{\mathbf{in}} = \theta_{\mathbf{flat}} + \frac{\widetilde{\kappa}}{A^2\widetilde{b}}$	$\Delta\gamma_{\mathbf{in}} \approx \Delta\gamma_{\mathbf{flat}} - \frac{a^2 c_b}{6\beta\xi_E}\big(\frac{\widetilde{\kappa}^2}{4\widetilde{b}} + \Gamma_g\big)$

Table 2.1: Summary of the analytic results. The table shows the impact of the wall geometry and the topology of the polymer (embedded in Γ_g and Γ_b, see the text) for two physical quantities of θ and $\Delta\gamma$. The quantity θ is related to the concentration of the polymer at the surface by $c_b\theta^2$ with c_b monomer density. $\Delta\gamma$ is the change in the surface tension due to the presence of the wall.

tension when the wall curves toward the polymers compared to when the wall curves away from the polymer.

Numerical Calculations

The above analytical calculations were related to the surfaces with large radius of curvature. To study polymer adsorption on the surfaces with higher curvature, we need to numerically solve Eq. (2.24) for both the interior and exterior of smaller spheres. Since this is mainly a phenomenological model, we follow similar parameters to previous numerical works and values that are typical for virus capsids and RNA sizes [31, 9, 60, 61, 2, 71, 72].

Outside the sphere

In this section, we consider smaller spheres and obtain the polymer concentration profile outside the sphere vs the scaled distance from the surface of the sphere, $\tilde{r} - \tilde{b}$. We obtain numerical results by solving the nonlinear differential equation Eq. (2.24) subject to the boundary conditions given in Eqs. (2.25) and (2.26). The results are presented

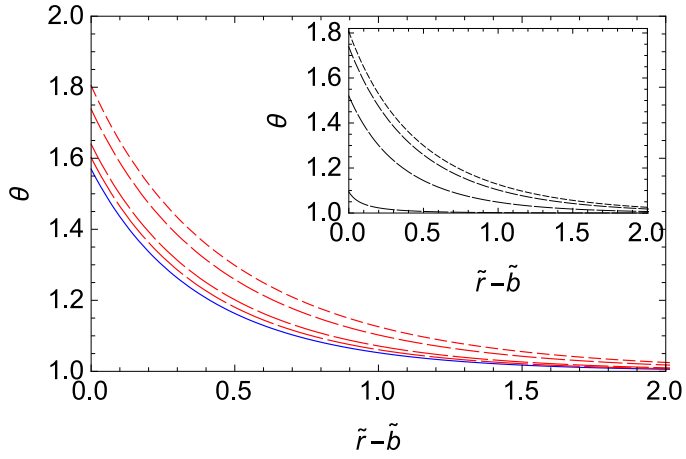


Figure 2.1: Profiles of the scaled polymer density amplitude θ vs scaled distance from the surface of the sphere $\tilde{r} - \tilde{b}$ for several relative branching density A_3. The solid line gives the profile for a linear polymer with no branching, and dashed lines give the profiles for branched polymers with branching density $A_3 = 0.2, 0.4, 0.8, 1.0$. Other parameters used are $\tilde{\kappa} = 1$, $\tilde{b} = 5.0$ and $\nu = 0.5$. The inset of Fig. 2.1 is the concentration profile for several values of the radius of the sphere ($\tilde{b} = 0.1, 1, 5, 100$) for the branching density $A_3 = 0.8$ with $\tilde{\kappa} = 1$ and $\nu = 0.5$. As the radius of the sphere increases, monomer concentration on the surface increases.

in Fig. 2.1, which illustrates that both the monomer concentration at the wall and the thickness of the adsorption layer increase as the branching density increases. The surface excess adsorbed onto the sphere can also be calculated using the concentration profile θ,

$$n_{ex} = \xi_E c_b \int_{\tilde{b}}^{\infty} d\tilde{r}(\theta^2 - 1)(\frac{\tilde{r}}{\tilde{b}})^2 \tag{2.38}$$

Figure 2.2 shows the surface excess as a function of the branching density. As the branching density increases, the surface excess also increases. For $A_3 = 0.2, 0.4, 0.6, 0.8, 1.0$, the ratio of surface excess of a branched to a linear polymer is $n_{ex}^{branched}/n_{ex}^{linear}$ =1.118, 1.264, 1.451, 1.696, 2.030, respectively. Inset of Fig. 2.2 is a plot of surface excess vs the

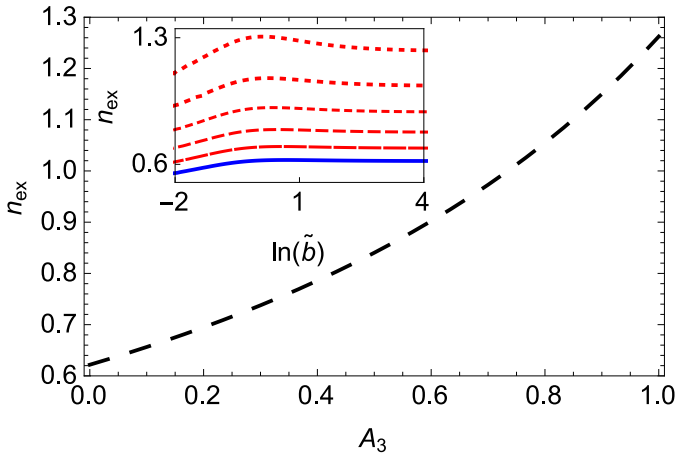


Figure 2.2: Surface excess (in units of $\xi_E c_b$) vs branching density for $\tilde{\kappa} = 1.0$, $\tilde{b} = 5.0$ and $\nu = 0.5$. Inset shows the excess on different size of spheres for $\tilde{\kappa} = 1.0$ and $\nu = 0.5$. The solid line represents the linear chain with $A_3 = 0$, and the dashed lines give the surface excess for branched polymers with branching density $A_3 = 0.2, 0.4, 0.6, 0.8, 1.0$ as dashing gets smaller.

radius of the sphere for different branching densities A_3 =0.2, 0.4, 0.6, 0.8, 1.0. According to the inset of Fig. 2.2, for a given branching density, there is a critical radius r^* for which the surface excess has a maximum. We find that the position of the critical radius decreases linearly as branching density goes up, *i.e.,* $r^* \propto -A_3$.

Using the numerical solution for θ and Eq. (2.27), the surface tension can be written as

$$\gamma_{out} - \gamma_0 = \frac{a^2 c_b}{6\beta\xi_E}\Bigg(- \tilde{\kappa}\theta^2(\tilde{b}) + \int_{\tilde{b}}^{\infty} dr\big((\theta^2-1)^2 - A_1(\theta-1) - A_3(2\theta^3 - 3\theta + 1) + (\tilde{\nabla}\theta)^2\big)(\frac{\tilde{r}}{\tilde{b}})^2\Bigg). \quad (2.39)$$

Figure 2.3 is a plot of the surface tension as a function of the branching density, which shows that as the branching density increases, the surface tension decreases. The inset shows the tension vs the sphere radius for different branching densities A_3 =0.2, 0.4, 0.6, 0.8, 1.0.

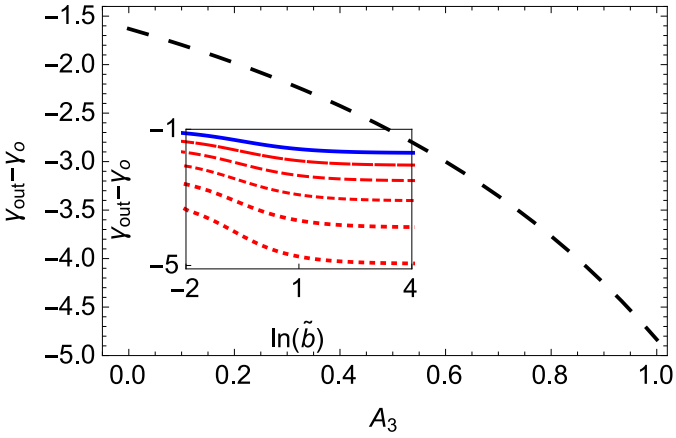


Figure 2.3: Surface tension (in units of $\frac{a^2 c_b}{6\beta\xi_E}$) vs branching density for $\tilde{\kappa} = 1.0$, $\tilde{b} = 5.0$ and $\nu = 0.5$. Inset is a plot of surface tension as a function of radius of the sphere for $\tilde{\kappa} = 1.0$ and $\nu = 0.5$. The solid line represents the linear chain with $A_3 = 0$, and the dashed lines give the surface tension for branched polymers with branching density $A_3 = 0.2, 0.4, 0.6, 0.8, 1.0$ as dashing gets smaller.

As the radius of sphere increases, the tension decreases consistent with the perturbative results.

Inside a sphere

In this section, we obtain the polymer concentration profile inside a sphere ($r < b$) by solving the nonlinear differential equation Eq. (2.24) subject to the boundary conditions given in Eq. (2.25). In addition, because the polymer is now confined inside an impermeable sphere, the total number of monomers, N is fixed. In terms of normalized length scale and the order parameter θ, we have

$$N = c_b \xi_E^3 \int_0^{\tilde{b}} 4\pi \tilde{r}^2 \theta^2 d\tilde{r}. \tag{2.40}$$

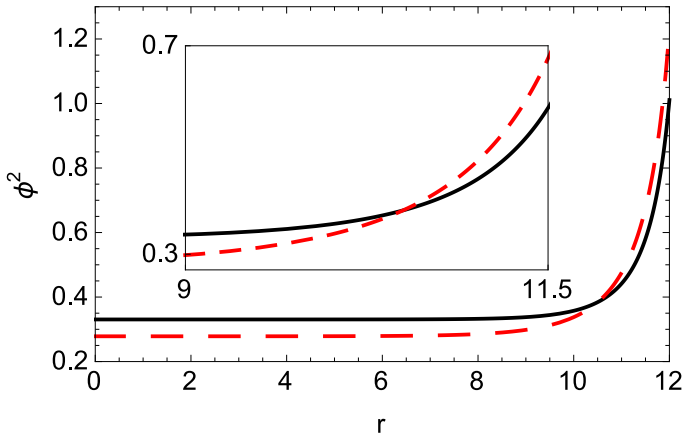


Figure 2.4: Concentration profile ($\phi(x) = \theta(x)\sqrt{c_b}$) inside the sphere for branching density $A_3 = 0.05$ (solid line) and $A_3 = 1.04$ (dashed lines) for $b = 12$, $\tilde{\kappa} = 1.0$, $\nu = 0.5$ and $N = 3000$ where the lengths are in units of a, the Kuhn length. Inset shows the details of the plot next to the surface.

The surface tension in this case can be written as

$$\gamma_{in} - \gamma_0 = \frac{a^2 c_b}{6\beta\xi_E}\Big(- \tilde{\kappa}\theta^2(\tilde{b}) + \int_0^{\tilde{b}} dr\big((\theta^2-1)^2 - A_1(\theta-1) - A_3(2\theta^3 - 3\theta + 1) + (\tilde{\nabla}\theta)^2\big)(\frac{\tilde{r}}{\tilde{b}})^2\Big). \quad (2.41)$$

The concentration profile as a function of r, the distance from the center of the sphere is shown in Fig. 2.4 for a branched polymer with branching density $A_3 = 0.05$ (solid lines) and $A_3 = 1.04$ (dashed lines). As illustrated in the figure, due to the attraction between the wall and the polymer, more monomers are attracted to the surface of sphere as the branching density increases.

Figure 2.5 illustrates the surface tension as a function of the radius of the sphere. Quite interestingly, we find that there is an optimal size for the radius of sphere for a given fixed chain length. For the parameters $\tilde{\kappa} = 1.0$, $\nu = 0.5$ and $N = 3000$, the optimal radius is $b = 11.25$ with $A_3 = 0.05$ and is $b = 10.90$ for $A_3 = 1.04$. When the polymer is more

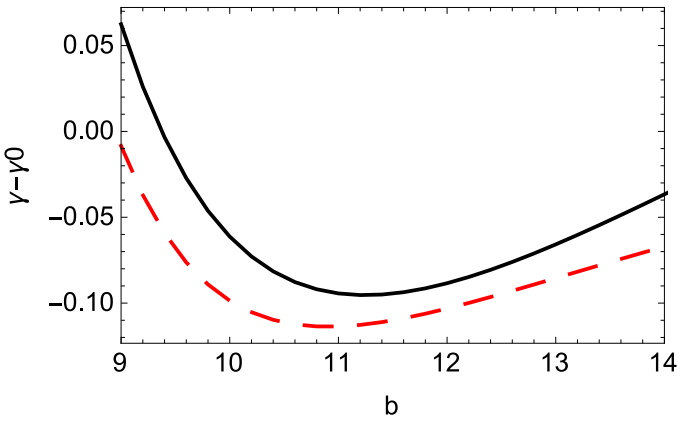


Figure 2.5: Surface tension for polymer adsorbed in the interior of the sphere at $A_3 = 0.05$ (solid line) and $A_3 = 1.04$ (dashed lines) for $\tilde{\kappa} = 1.0$, $\nu = 0.5$ and $N = 3000$ where the lengths are in units of a, the Kuhn length and the energy is in units of k_BT.

branched, the optimal radius becomes smaller. This is mainly due to lower cost for the excluded volume interaction.

Sinusoidal grating

In this section, we consider a polymer solution next to a sinusoidal surface, $z = z_0 cos((2\pi/\lambda)x)$. Here z_0 is the amplitude and λ is the wavelength of the surface. As mentioned in the introduction, this should give some insight into the behavior of branched polymer next to a rough random surface because the qualitative features of the results are the same. To obtain the concentration profile, θ we solve Eq. (2.24) subject to the boundary condition given in Eq. (2.25) using a finite element method in 2D. The numerical results are shown in Figs. 2.6, 2.7 and 2.8 as contour plots of the polymer density next to the sinusoidal adsorbing surface. In all cases we keep the strength of the attractive interaction between the surface and monomers the same.

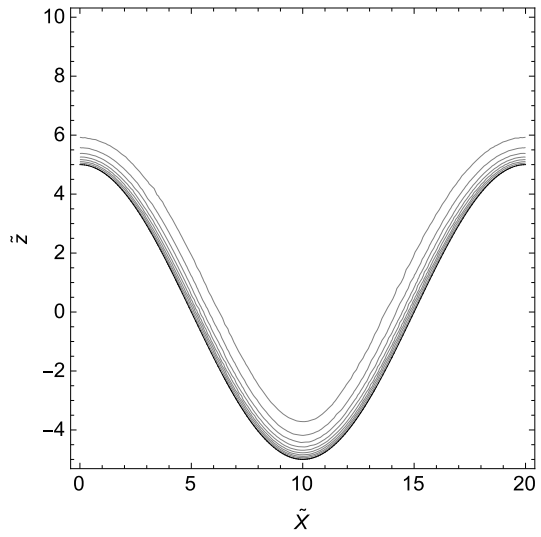


Figure 2.6: Contour plot of the concentration profile for the polymer with branching density $A_3 = 0.8$ for $\tilde{\kappa} = 1$, $\tilde{\lambda} = 20$ and $\tilde{z_0} = 5$. The solid thick black line is the position of the surface. The upper contour corresponds to a local concentration of $\theta^2 = 1.25$ that increases to $\theta^2 = 3.6$ towards the bottom of the valley. The $\tilde{x}$ and $\tilde{z}$ coordinates are distances along and perpendicular to the corrugations respectively.

Figure 2.6 shows that the profile of a branched polymer next to a sinusoidal surface is similar to that of a flat wall if the wavelength of surface fluctuations is large, for example $\tilde{\lambda} = \lambda/\xi_E = 20$. The figure also illustrates that the concentration of genome is higher in the valley compared to the peaks. This is consistent with the perturbative results presented in previous section, in that if the wall curves away from the genome, the monomer concentration decreases, otherwise, it increases. The non-uniformity in the concentration profile at the wall becomes more apparent as we decrease $\tilde{\lambda}$; *i.e.,* the genome concentration becomes much higher at the valley compared to the peak, see Fig. 2.7. In the figure, the amplitude of surface fluctuations, $\tilde{z_0} = z/\xi_E = 0.5$, is chosen to be relatively small to emphasize on the difference between the genome profile next to the flat and sinusoidal walls. Note that the amplitude $\tilde{z_0} = 0.5$ in Fig. 2.7 is 10 times smaller than that in Fig. 2.6. Nevertheless,

since $\tilde{\lambda} = 2$ is 10 times smaller in Fig. 2.7, the impact of surface fluctuations are more pronounced.

In addition, we find that not only the concentration profile at the wall is not uniform, the distribution of branch-points is not homogeneous either. Figure 2.8 illustrates the ratio of branch density to the monomer density at a sinusoidal surface. The figure shows that the branching concentration is higher at the valley with respect to peaks consistent with our perturbative results in section 2, where we found that the branching density increases if the surface is curved toward the polymer and decreases if the surface is curved away.

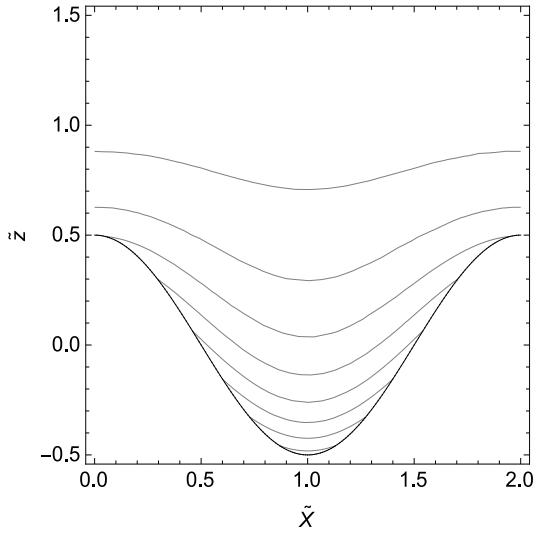


Figure 2.7: Contour plot of the concentration profile for the polymer with branching density $A_3 = 0.8$ for $\tilde{\kappa} = 1$, $\tilde{\lambda} = 2$ and $\tilde{z_0} = 0.5$. The solid thick black line is the position of the surface. The upper curve corresponds to a local concentration of $\theta^2 = 1.5$ that increases to $\theta^2 = 5$ towards the bottom of the valley. The $\tilde{x}$ and $\tilde{z}$ coordinates are distances along and perpendicular to the corrugations respectively.

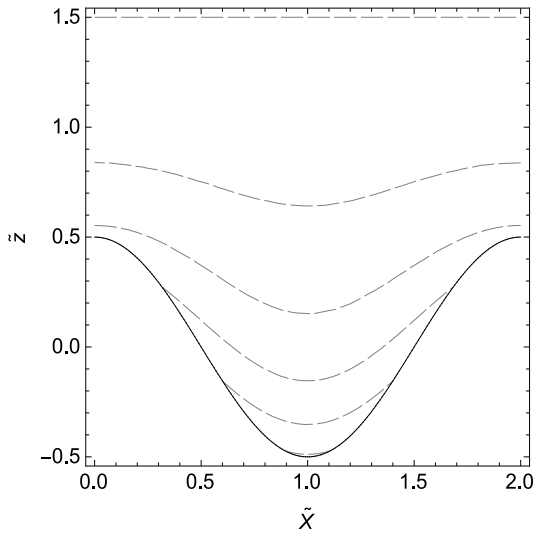


Figure 2.8: Contour plot of the ratio of branch to monomer concentrations with branching density $A_3 = 0.8$ for $\tilde{\kappa} = 1$, $\tilde{\lambda} = 2$ and $\tilde{z_0} = 0.5$. The solid thick black line is the position of the surface. The upper curve corresponds relative concentration of branch points of $c_3/c = 0.4$ and it increases to $c_3/c = 0.9$ towards the bottom of the valley. The $\tilde{x}$ and $\tilde{z}$ coordinates are distances along and perpendicular to the corrugations respectively.

2.4 Conclusion

In this paper, we use the field theory methods based on the $n \to 0$ limit of an $\mathcal{O}(n)$ model to describe the statistics of an annealed branched polymer. We, in particular, carefully examine the behavior of branched polymers next to various adsorbing walls both analytically and numerically.

We show that for the annealed branched polymers, increasing the branching density will increase the concentration of polymer but decrease the surface tension next to flat, inward curving, and outward curving walls. In comparison to a flat adsorbing wall, we find that when the wall curves toward the polymer solution, the tension decreases but when it curves away from it, the tension increases. While these results are consistent with those found for linear polymers next to different type of walls, we interestingly found a correla-

tion between the branching and curvature that causes a further lowering of surface tension when the wall curves towards the polymer, but decreases the amount of lowering of surface tension when the wall curves away from the polymer.

Our numerical solutions for the adsorption of polymer to the exterior of small spheres show that increasing the branching lead to an increase in the surface excess. For the polymers adsorbed in the interior of small spheres, we find a minimum in the surface tension as a function of radius of sphere. This clearly demonstrates the interplay between monomer-wall attraction and monomer-monomer excluded volume interaction. Our findings also indicate that branching decreases the optimal radius of sphere, as more polymers can sit in the vicinity of the wall without a huge cost for monomer-monomer repulsion. This result has a considerable consequence for the encapsulation of RNA by virus shell proteins, and could suggest a non-specific mechanism for the preferential packaging of viral RNA to cellular RNA *in vivo* [73].

Furthermore, we find similar effect for the adsorption of polymers onto sinusoidal surfaces. The concentration of polymers increases in the valleys compared to peaks and also branching density goes up in the valley section compared to the peaks. This effect becomes more pronounced as wave-length decreases.

Understanding the mechanisms involved with the adsorption of annealed branched polymers onto different surfaces will play a critical role in biomedical technologies. In particular, the paper was inspired by the idea of using functionalized gold nano-particles to bind RNA for gene delivery[46], which has industrial applications for biosensors and microfluidic

devices, and even possible medical application for gold nano-particles encapsulated by virus coats as potential tools for gene therapy. [45, 46, 47, 48, 49, 50].

Acknowledgment

This work was supported by the National Science Foundation through Grant No. DMR-13-10687.

2.5 Appendix A: $O(n)$ model of a magnet

In this appendix, we derive the equivalence between the grand canonical partition function, Eq. 2.2, for the polymer system and the partition function for the $n \to 0$ limit of an $\mathcal{O}(n)$ model. Note that the $\mathcal{O}(n)$ model corresponds to a magnet whose magnetic dipole of its atoms has n components . The Ising model commonly studied in most statical mechanics courses is the O(n) model for n=1. The n=0 limit is an interesting case as it reproduces the statistics of a self-avoiding linear polymer.

The Hamiltonian of the $O(n)$ model of a magnet is

$$H(\{\boldsymbol{S}\}, K, f_1, f_3; V) = -K \sum_{\langle x,x' \rangle}^{V} \boldsymbol{S}_x \cdot \boldsymbol{S}_{x'} - f_1 \sum_{x}^{V} J_1[\boldsymbol{S}_x] - f_3 \sum_{x}^{V} J_3[\boldsymbol{S}_x], \tag{2.42}$$

where S_x is an n dimensional vector of fixed length at each lattice point x and K, f_1, and f_3 are the coupling constants. The first sum in Eq. 2.42 is over all pairs of nearest neighbors and $J_1[\boldsymbol{S}_x]$ and $J_3[\boldsymbol{S}_x]$ are the source terms for end-points and branch-points respectively. We have used some prescience in giving the coupling constants K, f_1, and f_3 the same

symbol as the fugacities in Eq. (2.2). The partition function for the $\mathcal{O}(n)$ model is then

$$\mathcal{Z}_n(K, f_1, f_3; V) = \prod_x^V \mathrm{tr}_{\boldsymbol{S}_x}\, e^{-H(\{\boldsymbol{S}\},K,f_1,f_3;V)}, \tag{2.43}$$

where $\mathrm{tr}_{\boldsymbol{S}}$ defined as

$$\mathrm{tr}_{\boldsymbol{S}}\ e^{\boldsymbol{k}.\boldsymbol{S}} = \frac{\prod_i (\int_{-\infty}^{\infty} dS_i)\delta(\sum_i S_i^2 - n))e^{\boldsymbol{k}.\boldsymbol{S}}}{\prod_i (\int_{-\infty}^{\infty} dS_i)\delta(\sum_i S_i^2 - n))} \tag{2.44}$$

and the size of the spin is subject to the condition $|\boldsymbol{S}| = \sqrt{n}$ or $\sum_i S_i^2 = n$. Using the power series definition of the exponential, Eq. (2.43) can be written as

$$\mathcal{Z}_n(K, f_1, f_3; V) = \sum_{N_b=0}^{\infty} \frac{K^{N_b}}{N_b!} \sum_{N_3=0}^{\infty} \frac{f_3^{N_3}}{N_3!} \sum_{N_1=0}^{\infty} \frac{f_1^{N_1}}{N_1!} I_n(N_b, N_3, N_1; V), \tag{2.45}$$

with I_n defined as

$$I_n(N_b, N_3, N_1; V) = \prod_x^V \mathrm{tr}_{\boldsymbol{S}_x} \left[\Big(\sum_{\langle x,x' \rangle}^V \boldsymbol{S}_x \cdot \boldsymbol{S}_{x'} \Big)^{N_b} \Big(\sum_x^V J_3[\boldsymbol{S}_x] \Big)^{N_3} \Big(\sum_x^V J_1[\boldsymbol{S}_x] \Big)^{N_1} \right] \tag{2.46}$$

$$= \prod_x^V \mathrm{tr}_{\boldsymbol{S}_x}\, N_b! N_3! N_1! \sum_\alpha \mathcal{C}_\alpha[\{\boldsymbol{S}\}], \tag{2.47}$$

The comparison of the grand canonical partition functions defined in Eq. (2.2) with the partition function for the $\mathcal{O}(n)$ model in Eq. (2.45) reveals the similarity between the two models. It is now obvious why the coupling constants in the $\mathcal{O}(n)$ model were chosen to be labeled the same as the fugacities in the polymer system on a lattice with the excluded volume interaction. To show full equivalence between the grand canonical ensemble of self-avoiding branched polymers Ξ in Eq. (2.2) and the $n \to 0$ limit of partition function for the $O(n)$ model of a magnet (Eq. (2.43), we only need to show that in the $n \to 0$ limit the I_n expression gives the multiplicity Ω, or counts the number of ways of arranging a self-avoiding branched polymer on the lattice.

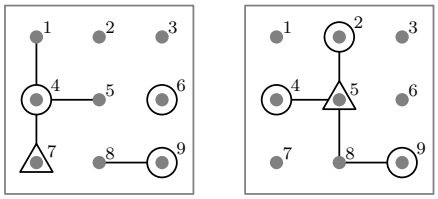


Figure 2.9: Two possible configurations of 4 bonds, 3 end-points, and 1 branch-point on a 3×3 two-dimensional square lattice. The left diagram is not a valid configuration while the right diagram is a valid configuration and is counted.

Each $\mathcal{C}_\alpha[\{\boldsymbol{S}\}]$ in Eq. (2.46) can be represented graphically. For the lattice, a product of neighboring $\boldsymbol{S}$-vectors that lie on points x and x' can be represented by a line drawn between the points. The $J_1[\boldsymbol{S}_x]$ and $J_3[\boldsymbol{S}'_x]$ source terms at point x and x' can be represented by circles and triangles placed on their respective points.

As an example we present two possible configurations, one that represents a valid branched polymer configuration and one that does not, with $N_b = 4$, $N_1 = 3$, and $N_3 = 1$ on a 3×3 square lattice in Fig. 2.9. The corresponding $\mathcal{C}_\alpha[\{\boldsymbol{S}\}]$ terms for both graphs in Fig. 2.9 contain 4 dot products of neighboring $\boldsymbol{S}$-vectors (one for each bond), 3 $J_1[\boldsymbol{S}]$ terms (3 circles), and 1 $J_3[\boldsymbol{S}]$ term (1 triangle). The $\mathcal{C}_\alpha[\{\boldsymbol{S}\}]$ term for the right diagram in Fig. 2.9 is

$$\mathcal{C}_b[\boldsymbol{S}] = \big(\boldsymbol{S}_2 \cdot \boldsymbol{S}_5\big)\big(\boldsymbol{S}_4 \cdot \boldsymbol{S}_5\big)\big(\boldsymbol{S}_5 \cdot \boldsymbol{S}_8\big)\big(\boldsymbol{S}_8 \cdot \boldsymbol{S}_9\big) J_1[\boldsymbol{S}_2] J_1[\boldsymbol{S}_4] J_1[\boldsymbol{S}_9] J_3[\boldsymbol{S}_5]. \tag{2.48}$$

where the indices indicate which lattice site is associated with each term.

We will show below that in the $n \to 0$ limit the trace of a single configuration gives one if the graph corresponds to a physically valid branched polymer configuration and zero

otherwise, i.e.,

$$\lim_{n\to 0}\prod_{x}^{V}\mathrm{tr}_{\boldsymbol{S}_x}\,\mathcal{C}_\alpha[\{\boldsymbol{S}\}]=\begin{cases}1 & \text{for } \alpha \text{ valid,}\\ \\ 0 & \text{for } \alpha \text{ invalid.}\end{cases} \tag{2.49}$$

This indicates, in the $n\to 0$ limit, the I_n term given in Eq. (2.45) counts the number of valid physical configurations, which is how multiplicity Ω is defined in Eq. (2.2). The condition in Eq. (2.49) establishes the equivalence between the partition function for the $\mathcal{O}(n)$ model in Eq. (2.45) with the grand canonical partition function for a flexible branched polymer in Eq. (2.2), *i.e.*

$$\lim_{n\to 0}\mathcal{Z}_n(K,f_1,f_3;V)=\Xi(K,f_1,f_3;V). \tag{2.50}$$

Evaluation of the trace

The trace of the configurations $\mathcal{C}_\alpha[\{\boldsymbol{S}\}]$ in Eq. (2.49) takes the form of products of $\boldsymbol{S}$-vectors on each lattice site x. These products can be evaluated by using the following generating function

$$\mathrm{tr}_{\boldsymbol{S}}\,S_{i_1}\cdots S_{i_p}=\frac{\partial}{\partial k_{i_1}}\cdots\frac{\partial}{\partial k_{i_p}}\,\mathrm{tr}_{\boldsymbol{S}}\,e^{\boldsymbol{k}\cdot\boldsymbol{S}}\Big|_{\boldsymbol{k}\to 0}. \tag{2.51}$$

The trace of the generating function $\mathrm{tr}_{\boldsymbol{S}}\,e^{\boldsymbol{k}\cdot\boldsymbol{S}}$ can be evaluated in closed form

$$\mathrm{tr}_{\boldsymbol{S}}\,e^{\boldsymbol{k}\cdot\boldsymbol{S}}=\sum_{p=0}^{\infty}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n}{2}+p)}\frac{1}{p!}\left(\frac{n|\boldsymbol{k}|^2}{4}\right)^p. \tag{2.52}$$

In the limit of $n\to 0$ it simplifies to the form

$$\lim_{n\to 0}\mathrm{tr}_{\boldsymbol{S}}\,e^{\boldsymbol{k}\cdot\boldsymbol{S}}=1+\frac{1}{2}|\boldsymbol{k}|^2. \tag{2.53}$$

A detailed step by step derivation of Eq. (2.52) is presented in appendix A of Ref. [74]. Inserting Eq. (2.53) into Eq. (2.51), we find the product of the components of the $\boldsymbol{S}$-vectors

in the absence of source terms is

$$\lim_{n\to 0} \mathrm{tr}_{\boldsymbol{S}}\, S_{i_1}\cdots S_{i_p} = \begin{cases} 1 & \text{if } p=0, \\ \delta_{i_1 i_2} & \text{if } p=2, \\ 0 & \text{otherwise.} \end{cases} \tag{2.54}$$

The expression in Eq. (2.54) evaluates to a non-zero value only in the lattice sites with exactly 0 or 2 bonds terminating on them.

To describe the generating function for end- and branch-points, we construct J_1 and J_3 functions, respectively, such that they satisfy the following equations

$$\lim_{n\to 0} \mathrm{tr}_{\boldsymbol{S}}\, S_i J_1[\boldsymbol{S}] = \lim_{n\to 0} \frac{1}{\sqrt{n}} \sum_j \delta_{ij}, \tag{2.55a}$$

$$\lim_{n\to 0} \mathrm{tr}_{\boldsymbol{S}}\, S_i S_j S_k J_3[\boldsymbol{S}] = \lim_{n\to 0} \sqrt{n} \sum_l \delta_{il}\delta_{jl}\delta_{kl}, \tag{2.55b}$$

while all other traces involving the sources such as $\mathrm{tr}_{\boldsymbol{S}}\, S_i S_j J_1[\boldsymbol{S}]$, $\mathrm{tr}_{\boldsymbol{S}}\, S_i J_3[\boldsymbol{S}]$ and $\mathrm{tr}_{\boldsymbol{S}}\, S_i S_j J_3[\boldsymbol{S}]$ are equal to zero in the $n \to 0$ limit. Using Eq. (2.52) it is straightforward to derive the following expressions for J_1 and J_3,

$$J_1[\boldsymbol{S}] = \frac{1}{\sqrt{n}} \sum_i^n S_i, \tag{2.56a}$$

$$J_3[\boldsymbol{S}] = \frac{4}{3n^{\frac{3}{2}}} \sum_i^n \Big(S_i^3 - \frac{3}{n} |\boldsymbol{S}|^2 S_i + 3S_i \Big). \tag{2.56b}$$

The sum in Eq. (2.55a) evaluates to a non-zero value only if for every lattice site containing the source term J_1 there exists exactly one bond terminating on that site. Similarly the sum in Eq. (2.55b) evaluates to a non-zero value only if for the sites with branch-point J_3, there are exactly three bonds terminating on that site. The factors of $1/\sqrt{n}$ in Eq. (2.55a) and $\sqrt{n}$ in Eq. (2.55b) enforce the no-loops condition necessary to obtain the statistics of a self-avoiding polymer.

In general, for every valid configuration with N_p connected graphs, N_1 end-points, and N_3 branch-points the product of all the lattice site integrals gives

$$\lim_{n\to 0}\prod_{x}^{V}\mathrm{tr}_{\boldsymbol{S}}\,\mathcal{C}_{\alpha}[\{\boldsymbol{S}\}]=\lim_{n\to 0}n^{N_p+\frac{1}{2}(N_3-N_1)}. \tag{2.57}$$

Using Eq. (2.1), the exponent in Eq. (2.57) is equal to the number of loops, so in the $n\to 0$ limit only those valid configurations with no loops evaluate to 1, while all others evaluate to 0.

It is important to note that the conditions given in Eqs. (2.55a) and (2.55b) do not forbid the multiple source terms sharing the same lattice site. This oversight can be remedied by redefining the partition function (see Eqs. (2.42) and (2.43)) such that the exponential of the source terms is replaced by the constant and linear terms in the power series expansion

$$\mathcal{Z}_n(K,f_1,f_3;V)=\prod_{x}^{V}\mathrm{tr}_{\boldsymbol{S}_x}\,e^{K\sum\limits_{\langle x,x'\rangle}\boldsymbol{S}_x\cdot\boldsymbol{S}_{x'}}\times\prod_{x}^{V}\big(1+f_1J_1[\boldsymbol{S}_x]+f_3J_3[\boldsymbol{S}_x]\big). \tag{2.58}$$

The structure in Eq. 2.58 ensures that there will be at most one single source term per lattice site and otherwise does not have any impact on the derivation presented above. From here on, we will only consider the partition function presented in Eq. 2.58 for the branched polymers system.

Mean Field Hamiltonian

We can now convert the lattice model over a discrete set of $\boldsymbol{S}$-vectors into a continuous field theory $\psi(x)$ using a Hubbard-Stratonovich transformation and connect the lattice fugacities K, f_1, and f_3 to physical quantities of chemical potential and concentration in the mean field approximation.

It is necessary to carefully treat the sum over nearest neighbors on the lattice given in Eq. (2.58) in order to change the lattice model into a continuous field theory. The sum can be written as a double sum over lattice sites multiplied by a nearest neighbor delta function

$$\sum_{\langle x,x'\rangle} = \frac{1}{2}\sum_{x}^{V}\sum_{x'}^{V}\delta_{\langle x,x'\rangle}. \tag{2.59}$$

The function $\delta_{\langle x,x'\rangle}$ is similar to a Kronecker delta function, and is explicitly defined as an operator that evaluates to 1 when x and x' are neighbors and 0 otherwise. The additional factor of half prevents double counting. Using Eq. (2.59), we now perform a Hubbard-Stratonovich transformation to introduce the auxiliary field $\psi(x)$

$$e^{\frac{K}{2}\sum\limits_{x,x'}\delta_{\langle x,x'\rangle}\boldsymbol{S}_x\cdot\boldsymbol{S}_{x'}} = \int_{-\infty}^{\infty}\mathcal{D}\psi\; e^{-\frac{1}{2}\sum\limits_{x,x'}\delta^{-1}_{\langle x,x'\rangle}\boldsymbol{\psi}(\boldsymbol{x})\cdot\boldsymbol{\psi}(\boldsymbol{x}')+\sum\limits_{x}\sqrt{K}\boldsymbol{\psi}(\boldsymbol{x})\cdot\boldsymbol{S}_x}, \tag{2.60}$$

where $\delta^{-1}_{\langle x,x'\rangle}$ is the inverse of the $\delta_{\langle x,x'\rangle}$ operator. Using Eq. (2.60), the partition function Eq. (2.58) can be written as

$$\begin{aligned}\mathcal{Z}_n(K,f_1,f_3;V) = \int_{-\infty}^{\infty}\mathcal{D}\psi\; e^{-\frac{1}{2}\sum\limits_{x,x'}\delta^{-1}_{\langle x,x'\rangle}\boldsymbol{\psi}(\boldsymbol{x})\cdot\boldsymbol{\psi}(\boldsymbol{x}')} \\ \times\prod_{x}^{V}\mathrm{tr}_{\boldsymbol{S}_x}\, e^{\sqrt{K}\boldsymbol{\psi}(\boldsymbol{x})\cdot\boldsymbol{S}_x}\big(1+f_1J_1[\boldsymbol{S}_x]+f_3J_3[\boldsymbol{S}_x]\big).\end{aligned} \tag{2.61}$$

The first term in the second line of Eq. (2.61) is simply the generating function performed in Eq. (2.51). We now use Eqs. (2.55a) and (2.55b) to evaluate the integrals associated with the sources J_1 and J_3 in Eq. (2.61) in the $n\to 0$ limit. Without loss of generality, the source terms can pick a special direction. To simplify the integrations, we choose the $(1,0,\ldots,0)$ direction and thus Eqs. (2.55a) and (2.55b) can be written

$$\lim_{n\to 0}\mathrm{tr}_{\boldsymbol{S}}\, S_iJ_1[\boldsymbol{S}] = \delta_{i1}, \tag{2.62}$$

$$\lim_{n\to 0}\mathrm{tr}_{\boldsymbol{S}}\, S_iS_jS_kJ_3[\boldsymbol{S}] = \delta_{i1}\delta_{j1}\delta_{k1}. \tag{2.63}$$

Inserting Eqs. (2.62) and (2.63) into Eq. (2.61) and performing the integral over $\boldsymbol{S}$-vector, the partition function in the $n \to 0$ limit becomes

$$\mathcal{Z}_n(K, f_1, f_3; V) = \int \mathcal{D}\psi \; e^{-\frac{1}{2}\sum\limits_{x,x'} \delta^{-1}_{\langle x,x' \rangle} \boldsymbol{\psi}(\boldsymbol{x})\cdot\boldsymbol{\psi}(\boldsymbol{x}')} \times \prod_{x}^{V} \left(1 + \frac{K}{2}|\boldsymbol{\psi}(\boldsymbol{x})|^2 + f_1\sqrt{K}\psi_1(\boldsymbol{x}) + \frac{f_3 K^{\frac{3}{2}}}{6}\psi_1^3(\boldsymbol{x})\right). \quad (2.64)$$

The $f_1\psi_1(x)$ and $f_3\psi_1^3(x)$ terms are proportional to the end and branch-point densities, while ψ^2 is proportional to the monomer density. Since for most physically relevant systems the ratio of branch or end-points to monomers is low, the $f_1\psi_1(x)$ and $f_3\psi_1^3(x)$ terms will be much smaller than the $\psi^2(x)$ term. Raising the second line of Eq. (2.64) into the exponent ($1 + X = e^{ln(1+X)}$) and expanding the logarithm, we define a new effective Hamiltonian as given in Eq. (2.8).

Chapter 3

RNA topology remolds electrostatic stabilization of viruses

3.1 Introduction

Simple viruses encapsulate their genetic material into a protein shell, measuring no more than about 15 nm across for the smallest viruses and about 28 nm for a typical (plant) virus [75, 2]. Under many circumstances, *in vitro* virus assembly is spontaneous and driven primarily by electrostatic interactions between negative charges on the backbone of the polynucleotide, usually single-stranded (ss) RNA, and positive charges on the virus coat proteins [76, 77, 78, 62, 79, 80, 81, 82]. However, recent experiments indicate that RNA plays a role that goes beyond its polyelectrolyte (PE) nature, as some RNAs are encapsulated more efficiently than others [9]. For example, when viral RNA1 of BMV (brome mosaic virus) and CCMV (cowpea chlorotic mottle virus) are mixed in solution with the capsid

proteins from CCMV, the BMV RNA is packaged three times more efficiently [9]. As the two RNAs differ in the amount of branching and their tertiary structure, both a straightforward consequence of their primary sequence, there must be a tight connection between capsid packing preferences and the structure of RNA [83, 84].

In many viruses the number of negative charges on the ssRNA is larger than the number of positive charges on the virus coat proteins [85, 71, 86, 87]. This overcharging phenomenon is intriguing and has been the subject of many papers. Belyi and Muthukumar examined a sample of actual viruses and found the ratio of the RNA charge to structural charge on the inner capsid surface to be around 1.6 [86]. While it seems to be feasible to encapsulate linear polymers with a number of charges as much as nine times that on the capsid proteins [55], which would result in strong "overcharging" of the virion, recent experiments show that the *optimal* number of charges residing on the linear PE is *less* than the total number of charges on the inner surface of viral shells [56], implying "undercharging" of the complex of capsid proteins with *linear* polyelectrolytes. This naturally leads to the question of which RNA features are implicated in the overcharging of the virion.

In what follows we show that RNA secondary structural features, such as branching, have a pronounced effect on the genome encapsulation capacity and thus could explain the phenomena of overcharging observed in viral particles.

In virtually all theoretical studies published to date investigating the overcharging in viruses, the genome is modeled as a simple linear polyelectrolyte chain [71, 63, 88, 89]. Thus the phenomenon of overcharging is associated with many factors other than the structure of RNA [86, 85, 71, 87, 63]. Our numerical solutions of the polyelectrolyte Poisson-

Boltzmann theory, without any additional assumptions regarding the effective monomer charge, do not support a universal overcharging of the virion. In fact, we find that mean-field PE theories for linear polymer chains lead to an undercharging phenomenon with fewer negative charges on the chain than positive ones on the capsid. This is consistent with several other recent numerical studies [63, 87, 78]. The overcharging in the viral particles has been explained through the genome-capsid N-terminal interactions and/or Manning condensation [86, 85, 71, 63]. However, it is important to note that the phenomenon of overcharging is also observed in viruses in which the number of charges on N-terminals is not significant, *e.g.*, Dengue and yellow fever viruses [85].

While the theoretical studies of linear polymers shed some light on the overcharging phenomenon, recent experiments reveal the importance of RNA structure that goes beyond its polyelectrolyte nature as a linear charged chain [9, 83, 90, 91]. Intra-chain base paring, promoted by hydrogen bonding between mutually complementary nucleotides along the backbone, leads to a highly branched structure of the RNA molecule that furthermore promotes its compaction in free solution.

In this paper, we present a model that takes into account the combined effect of RNA branching and the genome-capsid protein electrostatic interactions. We find that the larger the inherent propensity to form branch points quantified by the fugacity, f_b, the larger is the optimal chain length that can be accommodated in the capsid, consistent with *in vitro* experiments. These results are evident from Fig. 3.1, showing the displacement in the position of the encapsulation free energy minimum towards longer chains as the

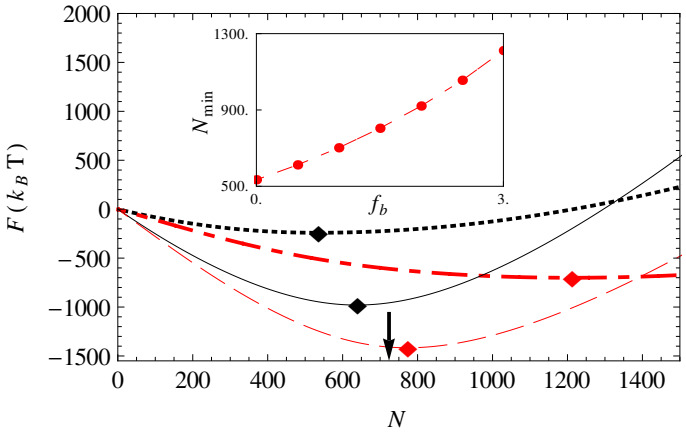


Figure 3.1: Encapsulation free energy as a function of monomer numbers for a linear polymer with f_b=0 (solid and dotted lines) and a branched polymer with f_b=3 (dashed and dot-dashed lines) at two different values of μ, corresponding to salt concentrations 10 mM (solid and dashed lines) and 100 mM (dotted and dotted-dashed lines). The arrow indicates the monomer number at which the full virus particle (capsid + polyelectrolyte) becomes neutral. Inset shows the position of the minimum $N_{\min}$ vs. the branching fugacity f_b for 100 mM salt concentration. Other parameters used are $\upsilon = 0.5$, $\tau = -1$, $\sigma = 0.4$, $b = 12$, and $T = 300K$, typical for RNA and virus capsids [2, 61, 72, 71].

branching fugacity increases. The inset to Fig. 3.1, showing the position of the minimum $N_{\min}$ as a function of the propensity for branching f_b, directly demonstrates this effect.

The figure also illustrates that the encapsulation free energy becomes more negative with increasing propensity of RNA to form branch points for a given number of monomers. This stabilization behavior suggests that branching is not only conducive to more efficient packing of the genome material into the virus shell, but also allows viral RNAs that have more branch points than other types of cellular RNAs [83] to out-compete the latter during replication in infected, susceptible host cells.

3.2 Theory

To obtain the optimal length and the free energy associated with the encapsulated RNA inside a capsid, we model RNA as a generic flexible branched polyelectrolyte. Because of the physical character of the base pairing, the degree of branching of RNAs is statistical and may in the process of encapsulation be affected by interaction with the charges located on or near the inner surface of the protein coat. To this end, we consider only the case of annealed branched polymers in this letter [92].

Further, we consider that the RNA interacts with positive charges residing on the inner surface of a sphere, where for simplicity we additionally presume that the charges are not localized but smeared out uniformly. For a large proportion of viruses the positive charges are indeed located on the inner surface of the capsid that in reality is not a perfect sphere but has a structure on the nanometer scale [72]. For some viruses positively charged disordered domains on the coat proteins point into the capsid cavity in a brush-like fashion [85], a feature that we do not include in our coarse-grained model at this stage.

In the mean-field, ground-state approximation, the free energy of a negatively charged polymer chain confined to a positively charged, infinitely thin spherical shell can be written as

$$F = \int \mathrm{d}^3 r \Big[\tfrac{1}{6} |\nabla \Psi(\mathbf{r})|^2 + W\big[\Psi(\mathbf{r})\big] - \tfrac{1}{8\pi\lambda_B} |\nabla \Phi(\mathbf{r})|^2 - 2\mu \cosh\big[\Phi(\mathbf{r})\big] + \tau \Phi(\mathbf{r}) \Psi^2(\mathbf{r}) \Big] + \int \mathrm{d}^2 r \Big[\sigma\, \Phi(\mathbf{r})\Big]. \quad (3.1)$$

All quantities in Eq. 4.1 are dimensionless, so energies are in units of thermal energy k_BT and lengths in units of the statistical step length (Kuhn length) of the polymer, a. Here, τ denotes the linear charge density of the polymer, σ the surface charge density of the shell, $\Psi^2(\mathbf{r})$ the monomer density at position $\mathbf{r}$, and $\Phi(\mathbf{r})$ the mean electrostatic potential. The parameter μ is the fugacity (density) of the monovalent salt ions, and corresponds to the concentration of salt ions in the bulk. The (dimensionless) Bjerrum length, λ_B, is a measure of the dielectric constant of the solvent, corresponding to about 0.7 nm for water at room temperature. The square gradient term in the first line of Eq. (4.1) describes the entropic cost for a non-uniform polymer density, and the last two lines in Eq. (4.1) describe the electrostatic interactions between the polyelectrolyte, the salt ions, and the charged capsid at the level of Poisson-Boltzmann theory [63]. The full derivation of the standard form of the free energy can be found in Refs. [59, 93, 94, 95, 96]. In addition to the standard form, we add the $W[\Psi]$ term that describes the statistics of an annealed branched polymer [97, 25, 98, 68], given explicitly by

$$W[\Psi] = \frac{1}{2}\upsilon\Psi^4 - f_e\Psi - \frac{f_b}{6}\Psi^3, \tag{3.2}$$

where υ is the (dimensionless) excluded volume and f_e and f_b are the fugacities of the end- and branch-points, respectively.

In our description, the stem-loop or hair-pin configurations in RNA structures are counted as end points. The number of end and branch points N_e and N_b of the polymer depend on the fugacities f_e and f_b through

$$N_e = -f_e\frac{\partial F}{\partial f_e} \qquad \text{and} \qquad N_b = -f_b\frac{\partial F}{\partial f_b}. \tag{3.3}$$

Since we consider only the case of a single encapsulated polymer with no closed loops, there is a constraint on the number of end and branch points,

$$N_e = N_b + 2, \tag{3.4}$$

with the degree of branching controlled by the fugacity of branch points f_b. The chain is linear if $f_b = 0$, and becomes more branched as f_b increases. The fugacity of endpoints f_e is not a free parameter in the system, it is set through the above constraint, Eq. (4.5). In addition, the total number of polyelectrolyte monomers inside the capsid is fixed [99, 60], *i.e.*,

$$N = \int \mathrm{d}^3 r \; \Psi^2(\mathbf{r}), \tag{3.5}$$

which we enforce by introducing a Lagrange multiplier, E, when minimizing the free energy.

We obtain the polyelectrolyte profile and electrostatic potential by varying the free energy functional with respect to fields $\Psi(r)$ and $\Phi(r)$ [50]. The resulting coupled set of non-linear equations describes the monomer density field, Ψ, and the electrostatic potential Φ_{in}, in the interior of the capsid, and the usual Poisson-Boltzmann equation for the electrostatic potential, Φ_{out}, in the exterior of the capsid, *viz.*,

$$\frac{1}{6}\nabla^2\Psi(\mathbf{r}) = -E\Psi(\mathbf{r}) + \tau\Phi_{in}(\mathbf{r})\Psi(\mathbf{r}) + \frac{1}{2}\frac{\partial W}{\partial \Psi} \tag{3.6a}$$

$$\nabla^2\Phi_{in}(\mathbf{r}) = \tfrac{1}{\lambda_D^2}\sinh\left[\Phi_{in}(\mathbf{r})\right] - \tfrac{\tau}{2\lambda_D^2\mu}\Psi^2(\mathbf{r}) \tag{3.6b}$$

$$\nabla^2\Phi_{out}(\mathbf{r}) = \tfrac{1}{\lambda_D^2}\sinh\left[\Phi_{out}(\mathbf{r})\right] \tag{3.6c}$$

where $\lambda_D = 1/\sqrt{8\pi\lambda_B\mu}$ is the (dimensionless) Debye screening length. The polymer segment concentration outside the capsid is assumed to be zero, $\Psi = 0$. Equations (3.6) along with the constraints in Eqs. (4.5) and (4.4) represent a set of coupled nonlinear differential

equations that, subject to appropriate boundary conditions, can only be solved numerically for the unknown parameters f_e and E and fields Ψ and Φ.

Assuming that the positive surface charge density, σ, is fixed, the electrostatic boundary condition (BC) is obtained by minimizing the free energy with respect to Φ on the surface, $\hat{n}.\nabla\Phi_{in} - \hat{n}.\nabla\Phi_{out} = 4\pi\lambda_B\sigma$. The choice of boundary conditions for Ψ depends on how the polymer interacts with the capsid surface through non-electrostatic forces. The strong short-ranged repulsion (as would be the case if we had included an excluded volume term between the polyelectrolyte monomers and the capsid proteins) leads to Dirichlet BCs. However, it turns out that for the large surface charge densities relevant to viruses, our conclusions are robust and do not depend on the choice of BC; we come back to this below. In this paper, we focus on Neumann BCs, directly obtained from the minimization of the free energy in Eq. 4.1 with respect to the polymer density field, ψ, on the surface.

3.3 Results

The overall dimensionless monomer density profiles $\mathcal{C}(r) = \Psi(r)^2$ as a function $r = |\mathbf{r}|$ the distance from the center of the cavity are shown in Fig. 3.2 for a linear polymer with $f_b = 0$ and a branched polymer with $f_b = 3.0$ of an equal number of segments, $N = 1000$, enclosed in a spherical shell. The radius of the capsid is taken to be $b = 12$ in units of the polymer Kuhn length that ftor our purpose is of the order of 1 nm [61]. Both types of polymers can adsorb onto the surface and, interestingly, the branched polymer is adsorbed more strongly onto the surface than the linear chain.

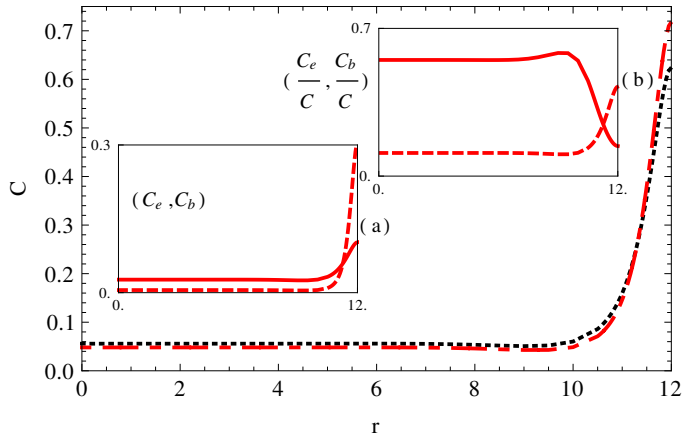


Figure 3.2: Concentration profile for *N=1000* with μ corresponding to salt concentration 100 mM and two different branching fugacities, $f_b = 0$ (linear polymer) for the dotted line, and $f_b = 3.0$ (branched polymer) for the dotted-dashed line. Inset (a) shows the concentration profile of endpoints (solid line) and branch points (dashed lines). Inset (b) shows the fraction of endpoints (solid line) and branch points (dashed lines). Other parameters used are $\upsilon = 0.5$, $\tau = -1$, $\sigma = 0.4$, $b = 12$, and $T = 300K$.

We also investigated the spatial inhomogeneity in our annealed branched polymer model of RNA. In Fig. 3.2 (a), we plot the dimensionless density of endpoints $C_e(r) = f_e\Psi(r)$ (solid line) and branches $C_b(r) = \frac{f_b}{6}\Psi^3(r)$ (dashed lines), obtained from Eq. (4.3). Figure 3.2 (b) illustrates the fractions of end points C_e/C (solid line) and fraction of branches C_b/C (dashed lines) as a function of the distance from the center of the capsid. We observe that most branching takes place within a two Debye length layer, thus very near the capsid wall where the concentration of segments is high and the local gradient in density is the largest. This is straightforward to understand as branching increases the local density allowing more segments to interact with the wall. Figure 3.2 also shows that end points are dominantly distributed over the capsid interior. Thus branching can affect not only the segment distribution but also the structure of the adsorbed layer making both quite non-uniform.

If we insert the parameter f_e and fields Ψ and Φ numerically obtained from Eqs. (3.6), Eq. (4.5), and Eq. (4.4) into Eq. (4.1), we can calculate the free energy of the chain-capsid complex, F. To obtain the encapsulation free energy using Eq. 4.1, we need to find the free energy difference between the chain-capsid complex and a free chain in solution and a positively charged capsid. The free energy of the free self-interacting chains (both linear and branched) is negligible under the stated conditions and is ignored. The capsid self-energy resulting from electrostatic interaction is calculated solving the system in the limit as $N \to 0$. Obviously, the capsid self-energy does not depend on the genome topology, but is not negligible and is explicitly subtracted. Here, we emphasize that since capsid proteins spontaneously self-assemble in the absence of genome in different kind of viruses, we only focused on the free energy of complexation of chain-capsid interaction.

A plot of the encapsulation free energy F vs the monomer number N, as shown in Fig. 3.1, confirms that the free energy minimum moves towards longer chains, *i.e.*, allowing more monomers to be encapsulated into the viral shell. In the inset to Fig. 3.1, we plot the optimal polymer length $N_{\min}$ (defined as the position of the free energy minimum) versus the branch point fugacity f_b for μ corresponding to 100 mM. As illustrated in Fig. 3.1, this effect is more pronounced at high salt concentration.

For low salt concentrations, electrostatics overwhelms all the other interactions and branching becomes less pronounced, but it still has an effect so that a branched polymer is packaged more efficiently than a linear polymer. For instance, at 10 mM salt the free energy has a minimum around N=638 for the linear polymer and N=773 for the branched polymer, see Fig. 3.1. At 100 mM salt, the optimal number of monomers for the linear

polymer is N = 534, but for the branched polymer increases by more than two times to N=1211. The arrow on the N axis of Fig. 3.1 corresponds to $N = 4\pi b^2\sigma$, representing a neutral system where the number of positive charges on the capsid wall is equal to the number of negative charges on the polymer chain. The aforementioned results correspond to the Neumann BC. As noted above, our conclusions do not depend on the type of BCs (Dirichlet vs. Neumann) that we employed. For the Dirichlet BC, at 100 mM, salt the optimal number of monomers is 202 for a linear polymer, and 930 for a branched polymer, consistent with the results for the Neumann BC.

This observation reveals that overcharging in viral particles could be a direct result of the secondary, i.e., branched, structure of viral RNA. We emphasize here that we repeated the above calculations for different surface charge densities, relevant to different viral capsids ($0.3 \leq \sigma \leq 0.9$), and found that for all cases, the number of charges on linear polymers is less than the number of positive charges on the capsid. Quite interestingly, we also found that for any given linear charge density of the chain, the optimal length of encapsulated branched polymers is always larger than that of linear polymers.

Figure 3.1 also reveals the second important point, *viz.* that the free energies associated with branched polymers have deeper minima than those for linear polymers for a set of salt concentrations. This effect explains why some RNAs are encapsulated more efficiently than other RNAs, or other linear polyelectrolytes for that matter.

If coupling between RNA branching and electrostatics represents a robust mechanism, the details of its description should not be qualitatively important. To this end, it is interesting to compare our results for encapsulated charged branched polymer based

on a field theoretic *Ansatz* for the statistics of branched polymers [97] with a very simple model with short-ranged attractive interaction between different chain segments mimicking the self-pairing of RNA bases [59]. We consider a linear polymer and now define the $W[\Psi]$ term as

$$W[\Psi] = \frac{1}{2}(v - sw)\Psi^4 + u\Psi^6, \tag{3.7}$$

with s the average fraction of base pairs and w the binding energy. We also include the next term in the virial expansion in order to stabilize the free energy since the total coefficient in front of the ψ^4 term can become negative. Calculating F vs. N curves for increasing values of s, the average fraction of self-paired bases, we find the same qualitative behavior as for increasing branching fugacity: the position of the minimum moves towards longer polymers (larger N) and the depth of the minimum increases for increasing s. For example, at 10 mM the free energy minimum is located at $N = 632$ for s=0 and at $N = 740$ for $s = 0.04$. At 100 mM salt the location of minimum moves from N=524 for s=0 to N=903 for s=0.04. Furthermore, as was the case for the branching model, the free energy minimum becomes deeper as the number of base-pairs increases.

3.4 Discussion and Conclusions

Both models described above show that the total charge of the genome inside the capsid is larger than the one residing on the capsid interior and that the virion is thus overcharged. Our analysis clearly reveals that the genomic function of RNA, as encoded in its sequence that engenders its branched secondary structure, plays an important role in the self-assembly of ssRNA viruses. The branched secondary structure of RNA, treated with

either branching or self-pairing models, promotes overcharging of the virion and stabilize its equilibrium configuration. We emphasize that while our results differ from previous studies [87], a very recent coarse-grained molecular dynamics simulation of the assembly of viral particles completely confirms the importance of the structure of RNA in the assembly process [29]. Note that within our field theory formalism, we do not observe overcharging for linear chains. The condition of the solution and protein charge distribution in the simulations of Ref. [29] are such that overcharging could be observed for linear chains; nevertheless, polymer branching enhances overcharging, consistent with our studies.

In order to explain the experiments noted in the introduction on the competition between RNA of CCMV and BMV through the theory presented above, we calculated the number of branch points for both RNAs. In particular, we used RNASubopt, a program in the Vienna RNA package [100], to generate an ensemble of secondary structures for genome sequences of RNA1 of BMV and CCMV. We then calculated the thermally averaged number of branch points from the secondary structures of each RNA. We found that RNA1 of BMV has higher average number of branch points (65) than CCMV (60.5) confirming that in the absence of specific interactions RNA1 of BMV would be preferentially packaged over RNA1 of CCMV, consistent with the experimental results of Comas-Garcia et al. [9, 101]. While one has to be cautious about results for longer sequences at high salt concentrations, the Vienna RNA Package [100] has been used to calculate thermally averaged properties of viral genomes with lengths of 2500-7000 nt, and important results have been obtained [83].

A comprehensive investigation of the physico-chemical parameters that impact capsid formation could have great potential in the development of anti-viral therapies and a systematic understanding of the processes involved in viral infection.

Acknowledgments

The authors would like to thank Mehran Kardar and Aaron Yoffe for useful discussions. This work was supported by the National Science Foundation through Grant No. DMR-13-10687 (R.Z.).

Chapter 4

Effects of RNA branching on the electrostatic stabilization of viruses

4.1 Introduction

Many single-stranded (ss) RNA viruses package their genome concurrently with the self-assembly of the whole capsid in such a way, that small protein subunits spontaneously assemble around the nucleic acid to built a complete protein shell (capsid) [102]. In the prevailing paradigm this assembly is predominantly driven by generic, nucleotide sequence independent, electrostatic interactions [103] between the negative charges on the RNA phosphate backbone and the positive charges on the virus capsid proteins (CP) [78, 62, 79, 81, 82, 104]. Recent experiments have indeed abundantly verified the importance of the "charge-matching hypothesis", based on the preponderance of electrostatic interactions between the capsid proteins and the RNA for proper genome packaging [6].

However, besides the importance of electrostatics, packaging experiments suggest that there must exist a correlation between the specific details of the nucleic acid structure and the efficient virus assembly [9, 83, 84, 105]. In a beautifully designed experiment Comas-Garcia et al. [9] have set the viral RNA1 of Brome Mosaic Virus (BMV) and the RNA of Cowpea Chlorotic Mottle Virus (CCMV) to compete against each other for capsid proteins belonging to CCMV exclusively. Although both RNAs are of similar length, BMV RNA was shown to outcompete the CCMV RNA, therefore suggesting that electrostatics alone is not enough for efficient genome encapsidation and that further structural details of RNA, apart from its generic charge, could play a role in the genome encapsidation [9, 10].

Even further away from the presumed non-specificity of the genome - CP interactions are indications, from both *in vitro* and *in vivo* studies, that the capsid self-assembly is achieved via a directed capsid assembly mediated by the highly specific, non-electrostatic interactions between sections of RNA and capsid proteins; these sections of RNA are thought to contain *packaging signals* and are repeated along the genome according to the symmetry of the capsid [106]. Contrary to the generic electrostatic charge matching, the essence of the packaging signal hypothesis is thus that the viral genomes have local secondary or tertiary structures with high CP affinity, serving as heterogeneous nucleation sites for the formation of capsids [107, 108]. Quite interestingly, in a recent experiment on Satellite Tobacco Mosaic Virus (STMV), Sivanandam *et al.* find that reducing the number of charges on the N-terminal section of capsid proteins through mutations results into the encapsidation of shorter RNAs than the wild type ones. However, unexpectedly a single mutation in one specific location along the N-terminal completely stops the self-assembly [105]. Investigat-

ing the nature of how and which structural details of RNA could be important for virus assembly is thus urgently required to ascertain on which point along the axis of "charge-matching" to "packaging signals" hypotheses the viruses actually drive and regulate their assembly.

Viral RNAs are found to be compact and highly branched [31] due to the base-pairing between the nucleotides, engendering compactification and folding of the molecule. Indeed, it appears that the compactness of the ssRNA wild-type viral genomes is one of the principal characteristics of their nucleotide sequence, setting them distinctly apart from randomized sequences [83, 109], and that the physical compactness of the viral genome can be regarded as a primary factor among evolutionary constraints [110].

While theoretical arguments suggest that the details of the RNA structure are important for its efficient packaging in the small volume of the virus capsid [61, 27, 29, 73, 111, 105], it remains overall poorly understood how the RNA sequence chemical composition together with its length affect the compactification and the packaging efficiency. Based on simple scaling arguments, it has been shown that genome secondary structures, or more specifically branching, lower the free energy of RNA encapsidation [61, 27]. As far as the length of RNA is concerned, there is a clear correlation with the number of positive charges on the virus coat proteins, structurally due to their extended N-tails, for many ssRNA viruses [85, 86, 29, 63, 27]. This correlation ratio is ~ 1.6 for many wild type viruses [86], implying that the number of negative charges on the RNA is in fact larger than the number of positive charges on the protein motifs, making these viruses *overcharged*.

Furthermore, when virus coat proteins encapsidate a linear polymer, *e.g.*, poly(styrene sulfonate) (PSS), two different results are obtained: both highly overcharged (correlation ratio ~ 9 [55]) and undercharged (correlation ratio between 0.45 and 0.6 [56]) virus-like particles (VLP). The overcharging phenomenon has been discussed in many theoretical papers with different conclusions dependening mostly on the details of the model under consideration [86, 85, 71, 87, 63, 55, 56, 112]. What one would hope for is that the important characteristics of the RNA genome packaging would robustly depend on some well defined characteristics of the genome, a hypothesis recently proposed in our work [73], where we showed that the secondary structure of RNA, as quantified by its branchiness, coupled to electrostatic interactions enhances the genome encapsidation capacity and could robustly explain the overcharging actually observed in virions.

While understanding the detailed role of electrostatics and structure of RNA on the self-assembly is the focus of what follows, we also aim additionally to understand what controls the virions or VLP stability or what the main factors are that enhance this stability before the disassembly of the capsid. Viruses seem to release their genome during the disassembly [113], which would imply that the genome not just leaves, but is in fact actively pushed from the capsid - a scenario that has been shown as specifically valid for bacteriophages, where the repulsive DNA-DNA interactions act like a coiled osmotic spring ejecting the genome. The corresponding osmotic pressure is in fact quite large and positive, surpassing even 50 atm, and stemming mostly from the combination of electrostatic and hydration interactions that are dominant in the range of DNA densities relevant for bacteriophage packing [103].

Contrary to DNA in bacteriophages, the osmotic pressure in ssRNA viruses is not easy to measure directly and in the absence of experiments one thus has to rely on theoretical estimates. There have been several theoretical studies that investigate the osmotic pressure of ssRNA viruses [2, 71, 63, 114, 115]. Siber and Podgornik showed that the filled ssRNA virions exhibit a small residual negative osmotic pressure, which depends strongly on the amount of capsid charges and can be turned positive with relatively higher capsid charge [63]. In addition, Javidpour et al. studied the effects of multivalent ions, which can fundamentally change the nature of electrostatic interactions [116], on the osmotic pressure and the stability of the virus like empty shells, showing that the multivalent ions can turn a positive electrostatic osmotic pressure into a negative one [114]. Furthermore, recent all atom molecular dynamics simulations showed that the osmotic pressure inside an empty Poliovirus capsid is negative, suggesting that the mechanism might be connected with excess charges on the capsid that prevent the solution ion to exchange with the capsid [115], a scenario at odds with what we know about the permeability of capsids. While there have thus been several lines of investigation regarding the nature and specifically the sign of the capsid osmotic pressure, there exist no studies taking into account the role of the secondary structure of RNA in the osmotic pressure of ssRNA viruses or virus like particles, another aspect that we elucidate further below.

In this paper, we extend our previous analysis and investigate how the secondary structure of the RNA affects the osmotic pressure of ssRNA viruses and what are the repercussions for stability of the virions. We show that the secondary structure of RNA may indeed result in negative osmotic pressures at conditions where a linear polymer would ex-

hibit positive osmotic pressures. This may suggest that having a branched structure makes not only RNA more effectively packaged but also makes a virion more stable. The paper is organized as follows. In the next section, we introduce the model and the fundamentals of the theory together with the basic quantities that we will calculate. In Sec. 4.3, we present the results for osmotic pressure as well as the effect of RNA branching on the free energy minimum, defining the optimum length of RNA, the optimum number of branched points and the optimum charge ratios of the system, together with the corresponding ion concentration and RNA density profiles. Section 6.5 discusses effects of different models, boundary conditions and different parameterizations that might correspond to different types of viruses. Finally, we summarize our findings. In the appendix, we derive in detail the model free energy of the encapsidation.

4.2 Model

To elucidate the role of genome in the assembly of spherical RNA viruses, we model RNA as a generic, negatively charged, flexible branched polyelectrolyte that interacts with positive charges residing on the inner surface of the capsid. More specifically, we consider only the case of *annealed branched polymers* because the strength of RNA base-pairing is relatively weak and may easily be affected by the interaction with the positive inner surface charges of the shell during encapsidation. For simplicity, we model the capsid as a thin sphere and assume that the charges are not localized but smeared out uniformly on the inner surface of the sphere. We note that while a thin shell is a good approximation for the capsid of some viruses like Dengue and yellow fever [72], the capsid proteins of some

other viruses contain N-terminal tails which are highly positively charged and point into the capsid cavity in a brush-like fashion [85].

The mean-field free energy functional of a polyelectrolyte chain confined within a charged shell in a univalent salt solution, under the ground state approximation, can be written as

$$\beta F = \int \mathrm{d}^3 r \Big[\tfrac{a^2}{6} |\nabla \Psi(\mathbf{r})|^2 + W\big[\Psi(\mathbf{r})\big] - \tfrac{\beta^2 e^2}{8\pi\lambda_B} |\nabla \Phi(\mathbf{r})|^2 - 2\mu \cosh\big[\beta e \Phi(\mathbf{r})\big] + \beta\tau \Phi(\mathbf{r}) \Psi^2(\mathbf{r}) \Big] + \int \mathrm{d}^2 r \Big[\beta\sigma\, \Phi(\mathbf{r}) \Big]. \quad (4.1)$$

Here β denotes the inverse of the thermal energy $k_B T$, a the statistical step (Kuhn) length of the polymer, τ the linear charge density of the polymer, σ the surface charge density of the shell, $\Psi(\mathbf{r})$ the monomer density field at position $\mathbf{r}$, and $\Phi(\mathbf{r})$ the mean electrostatic potential. The parameter μ is the fugacity of the monovalent salt ions corresponding to the concentration of salt ions in the bulk. $\lambda_B = e^2\beta/4\pi\epsilon\epsilon_0$, is the Bjerrum length, a measure of the dielectric constant (ϵ) of the solvent and is about 0.7 nm for water at room temperature.

The first term of Eq. (4.1) is the entropic cost of non-uniform polymer density and the last two lines of Eq. (4.1) correspond to the electrostatic interactions between the polymer, the shell and the salt ions on the level of the Poisson-Boltzmann theory [63]. The standard form of this free energy can be found in references [59, 63]. For completeness we also provide a step by step derivation of Eq. (4.1) for a linear polymer in the appendix.

The self-interaction term $W[\Psi]$ in Eq. (4.1) is associated with the self repulsion of the polyelectrolyte and the energy of an annealed branched polymer [97, 25, 98, 68],

$$W[\Psi] = \frac{1}{2}\upsilon\Psi^4 - \frac{1}{\sqrt{a^3}}(f_e\Psi + \frac{a^3}{6}f_b\Psi^3), \tag{4.2}$$

where υ is the excluded volume term and f_e and f_b are the fugacities of the end- and branch-points of the annealed polymer, respectively. A detailed derivation of Eq. 4.2 is given in [117]. In this model, the stem-loop or hair-pin configurations of RNA are counted as the end points. The number of end- and branch-points N_e and N_b of the polymer are related to the fugacities f_e and f_b in a standard way by

$$N_e = -\beta f_e \frac{\partial F}{\partial f_e} \qquad \text{and} \qquad N_b = -\beta f_b \frac{\partial F}{\partial f_b}. \tag{4.3}$$

We have two additional constraints in the problem. First, the total number of monomers inside the capsid is fixed [60],

$$N = \int \mathrm{d}^3 r \; \Psi^2(\mathbf{r}), \tag{4.4}$$

a constraint that we enforce by introducing a Lagrange multiplier, E, when minimizing the free energy. Second, the number of the end points depends on the number of branched points so that

$$N_e = N_b + 2, \tag{4.5}$$

since we consider only a single polymer with no closed loops. Thus, f_e is not a free parameter. For our calculations, we change f_b and find f_e through Eqs. (4.3) and (4.5). The polymer is linear if $f_b = 0$, and the number of branched points increases with f_b.

By varying the free energy functional with respect to fields $\Psi(\mathbf{r})$ and $\Phi(\mathbf{r})$, we obtain a coupled set of non-linear differential equations coupling the monomer density with

the electrostatic potential in the interior of the capsid, and the usual Poisson-Boltzmann equation for the exterior of the capsid. The monomer density field in fact satisfies the modified Edwards equation

$$\frac{a^2}{6}\nabla^2\Psi(\mathbf{r}) = -E\Psi(\mathbf{r}) + \beta\tau\Phi_{in}(\mathbf{r})\Psi(\mathbf{r}) + \frac{1}{2}\frac{\partial W}{\partial \Psi}, \tag{4.6}$$

while the electrostatic potential satisfies the modified Poisson-Boltzmann equation in the interior of the capsids

$$\nabla^2\Phi_{in}(\mathbf{r}) = \frac{1}{\lambda_D^2\beta e}\sinh\left[\beta e\Phi_{in}(\mathbf{r})\right] - \frac{\tau}{2\lambda_D^2\mu\beta e^2}\Psi^2(\mathbf{r}), \tag{4.7}$$

and the standard Poisson-Boltzmann equation in the exterior

$$\nabla^2\Phi_{out}(\mathbf{r}) = \frac{1}{\lambda_D^2\beta e}\sinh\left[\beta e\Phi_{out}(\mathbf{r})\right], \tag{4.8}$$

where $\lambda_D = 1/\sqrt{8\pi\lambda_B\mu}$ is the Debye screening length. The boundary condition (BC) for the electrostatic potential is obtained by minimizing the free energy, $\hat{n}\cdot\nabla\Phi_{in} - \hat{n}\cdot\nabla\Phi_{out} = 4\pi\lambda_B\sigma/\beta e^2$ assuming the surface charge density σ is fixed. The concentration of the polymer outside of the capsid is assumed to be zero. The BC for the inside monomer density field Ψ is of Neumann type ($\hat{n}\cdot\nabla\Psi|_s = 0$), that can be obtained from the energy minimization [60]. However, due to the short-ranged self-repulsions of the polymer, Dirichlet type BC ($\Psi|_s = 0$) might be preferable so that the polymer density goes to zero on the surface of the capsid. In our calculations we use both types of BCs and find that our conclusions do not depend on their detailed nature so that our conclusions are robust. We start with the Neumann BC but discuss the impact of the Dirichlet BC later in Sec. 6.5.

Using Eq. (4.1), we can also obtain the osmotic pressure due to the genome encapsidation, *i.e.*, the force exerted on the virus capsid by the genome per unit surface area,

defined as

$$P(N) = -\left(\left.\frac{\partial F}{\partial V}\right|_{Q_c,N} - \left.\frac{\partial F}{\partial V}\right|_{Q_c,N=0}\right), \tag{4.9}$$

where V is the volume of the capsid and we subtracted the part of the osmotic pressure for the empty capsid. In the calculation of the pressure, we keep the total number of monomers N and the total number of charges on the capsid $Q_c = 4\pi b^2\sigma$ constant with b the radius of the capsid.

4.3 Results

We numerically solve the nonlinear coupled differential equations, Eqs. (4.6), (4.7), (4.8), subject to the constraints given in Eqs. (4.4) and (4.5) to obtain the fields Ψ and Φ and the parameter f_e. Electrostatic potential and polymer concentration profiles as a function of r, the distance from the center of the shell, are shown in Fig. 4.1(a) and (b), respectively for $10mM$ (solid and dashed lines) and for $100mM$ (dotted and dotted-dashed lines) salt concentrations for a linear polymer with $f_b = 0$ (solid and dotted lines) and a branched polymer with $f_b = 3.0$ (dashed and dotted-dashed lines). The total number of monomers enclosed in the shell is $N = 1000$ for both profiles shown in the figure. Independent of the amount of salt and degree of branching, the polymer concentration is always larger right next to the surface due to the electrostatic attraction between the polymer and capsid, but it is higher for the branched polymers than the linear one (Fig. 4.1 (b)). Note that in all cases the genome profiles remain nearly constant inside the shell but increase noticeably in the vicinity of the capsid wall.

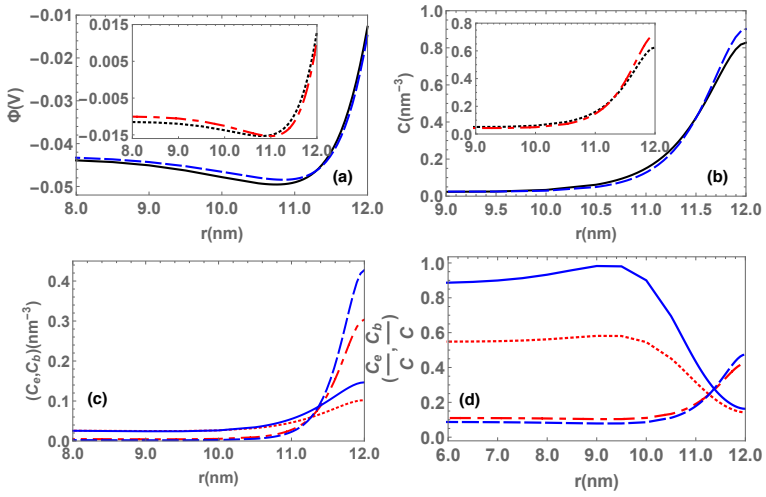


Figure 4.1: For $N = 1000$ and two different salt concentrations μ corresponding to 10 mM (solid and dashed lines) and 100 mM (dotted and dotted-dashed lines), (a) Electrostatic potential profile for a linear polymer with $f_b = 0$ (solid and dotted lines) and branched polymer with $f_b = 3.0$ (dashed and dotted-dashed lines) (b) Concentration profile corresponding to two different degree of branching for a linear polymer with $f_b = 0$ (solid and dotted lines) and for a branched polymer with $f_b = 3.0$ (dashed and dotted-dashed lines). (c) Concentration profile of endpoints (solid and dotted lines) and branch points (dashed and dotted-dashed lines) for a branched polymer with $f_b = 3.0$. (d) Fraction of endpoints (solid and dotted lines) and branch points (dashed and dotted-dashed lines) for a branched polymer with $f_b = 3.0$. Other parameters are $\upsilon = 0.5\ nm^3$, $\tau = -1\ e$, $\sigma = 0.4\ e/nm^2$, $b = 12\ nm$, $a = 1\ nm$ and $T = 300\ K$.

In addition, we investigated the distribution of branch and end points inside the capsid for $10mM$ and for $100mM$ salt concentrations. Figure 4.1(c) illustrates the concentration of endpoints $C_e(r) = \frac{1}{\sqrt{a^3}} f_e \Psi(r)$ (solid line for 10 mM and dotted line for 100 mM) and branch points $C_b(r) = \frac{\sqrt{a^3}}{6} f_b \Psi^3(r)$ (dashed lines for 10 mM and dotted-dashed lines for 100 mM), obtained from Eq. (4.3). As shown in Fig. 4.1(c), the number of branch points increases in the vicinity of the capsid wall at both salt concentrations; however, it increases even more at the lower salt concentration indicating more segments interact with the wall. The end points, on the other hand, mainly distributed over the interior of the shell. Figure

4.1(d) shows the fractions of end points C_e/C (solid lines for 10 mM and dotted line for 100 mM) and fraction of branch points C_b/C (dashed lines for 10 mM and dotted-dashed lines for 100 mM) as a function of r.

Once the fields Ψ and Φ are obtained, we insert them into Eq. (4.1) to calculate the free energy of chain-capsid complex, F. To obtain the encapsidation free energy, F, we need to calculate the free energy of a chain free in solution and that of a positively charged capsid and then subtract them both from the chain-capsid complex free energy, F given in Eq. (4.1).

The capsid self-energy ($F(N = 0)$) due to the electrostatic interactions is calculated through Eqs. (4.7) and (4.8) in the limit as $N \to 0$, and should be explicitly subtracted from the encapsidation free energy. The focus of this paper is on the solution conditions in which the capsid proteins can spontaneously self-assemble in the absence of genome as seen in different kind of experiments [118, 81]. Note that the free energy associated with a free chain (both linear and branched) is negligible under the experimental conditions [71, 63, 27]. To avoid the problem of proper free energy rescaling, we furthermore calculate the osmotic pressure of RNA trapped inside the capsid and investigate the impact of its secondary structure on the stability of capsid. Through the calculation of osmotic pressure, we have been able to confirm all our conclusions obtained through the free energy calculation.

In order to get the osmotic pressure, we first calculate the free energy of the system as a function of the monomer number N for both linear and branched chains and then insert it in Eq. (4.9). A plot of the osmotic pressure P vs. the monomer number N is given in Fig. 4.2(a) for both linear and branched polymers at two different salt concentrations. The

solid and dotted lines correspond to linear polymers with $f_b = 0$ and dashed and dotted-dashed lines to branched polymers with $f_b = 3.0$. The salt concentrations are 10 mM (solid and dashed lines) and 100 mM (dotted and dotted-dashed lines). As is clear from the figure, the osmotic pressure goes through a minimum and this minimum is displaced towards longer chains as we increase the degree of branching, *i.e.*, more monomers can be encapsidated with increasing f_b. For example, the minimum of pressure is at $N \approx 523$ for a linear polymer $f_b = 0$, and increases to $N \approx 851$ for a branched polymer with $f_b = 3$ at 100 mM salt. At 10 mM salt, the minimum of the free energy is at $N \approx 628$ for $f_b = 0$ and at $N \approx 719$ for $f_b = 3$.

Figure 4.2 (b) shows the osmotic pressure in terms of the degree of branching f_b for 10 mM (dotted lines) and 100 mM (dotted-dashed lines) salt concentrations with $N = 1200$. When $f_b = 0$ (linear polymer), the osmotic pressure is positive but changes the sign as f_b increases regardless of the salt concentration. The figure shows that the pressure becomes more negative as the degree of branching increases indicating that the secondary structure of the genome makes the virus more stable.

To further investigate the role of branching on the assembly of viral shells, we study the impact of branching on the minimum free energy, the optimal number of monomers, the optimal number of branched points, and the ratio of the chain charge to the capsid charge. A plot of the encapsidation optimum free energy F_{min} vs. the branching fugacity f_b is given in Fig. 4.3(a) at two different salt concentrations. For branched polymers, the free energy becomes deeper, indicating that compared to the linear polymers, the branchiness confers more stability to the capsid at both salt concentrations. This effect could explain why some

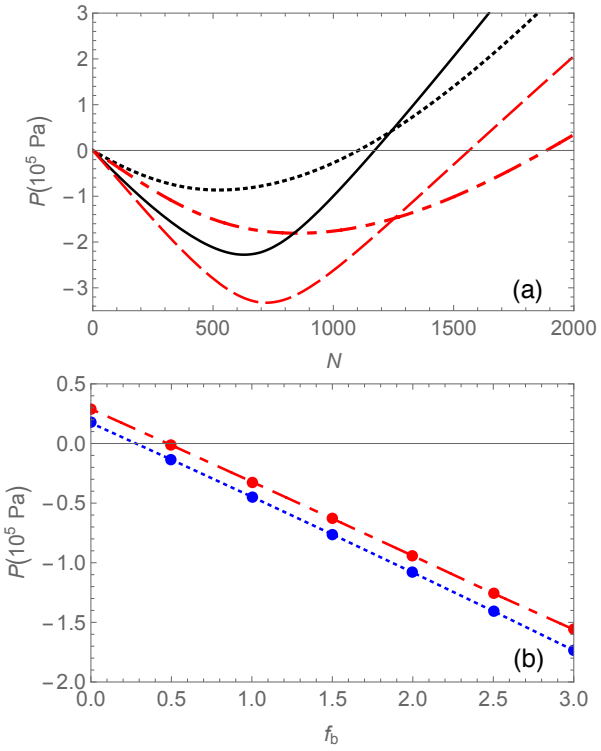


Figure 4.2: (a) Osmotic pressure as a function of monomer numbers for a linear polymer with $f_b = 0$ (solid and dotted lines) and a branched polymer with $f_b = 3$ (dashed and dotted-dashed lines). Solid and dashed lines correspond to the salt concentration, $\mu = 10$ mM and dotted and dotted-dashed lines represent the salt concentration, $\mu = 100$ mM. (b) Osmotic pressure for $N = 1200$ as a function of fugacity of branch points, f_b, at 10 mM (dotted lines) and 100 mM (dotted-dashed lines) salt concentrations. Other parameters are $\upsilon = 0.5$ nm^3, $\tau = -1$ e, $\sigma = 0.4$ e/nm^2, $b = 12$ nm, $a = 1$ nm and $T = 300$ K.

RNAs are encapsidated more efficiently than others, or indeed linear polyelectrolytes. Note that the effect of branching is more apparent at high salt concentrations. Expectedly, for low salt concentrations, electrostatics overwhelms all the other interactions and the impact of branching becomes less pronounced; nevertheless, the minimum moves towards the longer chains for branched polymers compared to linear ones.

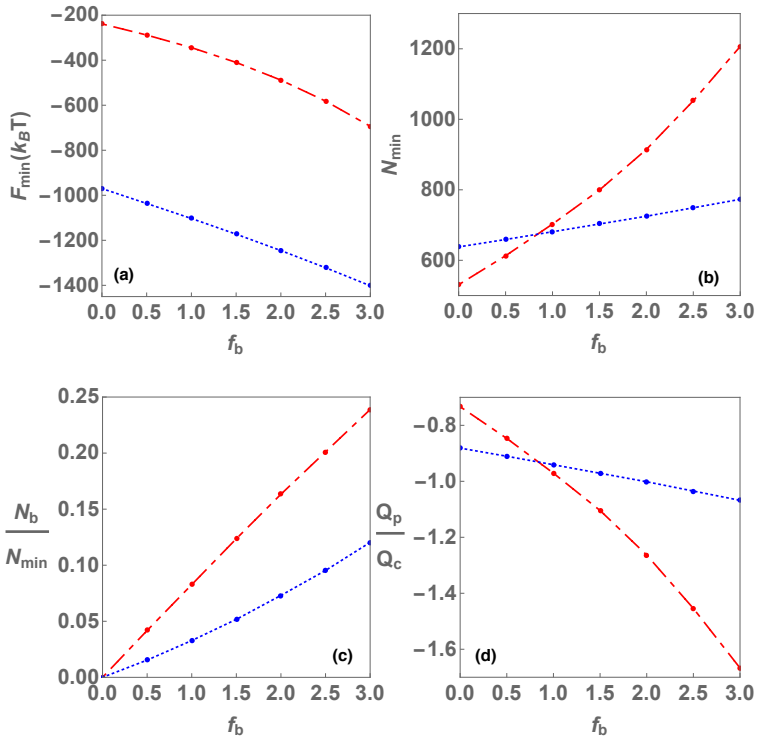


Figure 4.3: For 10 mM (dotted lines) and 100 mM (dotted-dashed lines) salt concentrations, (a) Optimum free energy (units of k_BT) (b) Optimum number of monomers (c) Ratio of number of branched points to the number of monomers at the minima (d) Ratio of number of polymer charges to the capsid charges at the minima as a function of fugacity of branch points, f_b. Other parameters are $\upsilon = 0.5\ nm^3$, $\tau = -1\ e$, $\sigma = 0.4\ e/nm^2$, $b = 12\ nm$, $a = 1$ nm and $T = 300\ K$.

Figure 4.3(b) shows the optimal number of encapsidated monomers associated with the minimum of free energy as a function of f_b. As illustrated in the figure, more monomers are packaged as the degree of branching increases. For example, at 100 mM for a linear polymer, $f_b = 0$, the optimum number of monomers is $N \approx 534$ and it increases to $N \approx 1211$ for a branched polymer with $f_b = 3.0$. At 10 mM salt, the optimum monomer number for linear polymer is $N \approx 638$ and for branched one is $N_{min} \approx 773$, with $f_b = 3.0$. Figure 4.3(c) is a plot of the ratio of number of branched points to the optimal number of monomers vs. the branching fugacity. As expected, the ratio increases for higher f_b values.

The fact that longer, branched chains can be more easily encapsidated by capsid proteins could straightforwardly explain one of the reasons why viruses are overcharged. The total charge of the virion is $Q = Q_p + Q_c = \tau N + 4\pi b^2 \sigma$ where the first term corresponds to the genome charge and the second one to that of the capsid. Figure 4.3(d) shows the charge ratio of the genome to the capsid vs. the fugacity of branched points for two different salt concentrations at the minima of the free energy for $\upsilon = 0.5\ nm^3$, $\tau = -1\ e$, $\sigma = 0.4\ e/nm^2$, $b = 12\ nm$, $a = 1\ nm$ and $T = 300\ K$. The virion becomes overcharged for the values of $f_b > 2$ at 10 mM and $f_b > 1$ at 100 mM.

4.4 Discussion and Summary

We have investigated the role of RNA sequence specificity, as it transpires through the RNA branchiness in the electrostatic encapsidation of RNA viruses. Specifically, we addressed in detail the dependence of the free energy and the osmotic pressure of a confined self-interacting RNA constrained within a spherical, charged capsid. The sequence specificity was modeled through an annealed distribution of RNA end- and branch-points, and the electrostatics was addressed within a mean-field Poisson-Boltzmann framework, allowing us to study explicitly the impact of branching and genome-capsid electrostatic interaction on the optimal length of the encapsidated genome. While the details of our model can be subject to criticism and RNA sequence specificity could enter on other more detailed levels of description, we do believe that the coupling between RNA self-interaction and capsid electrostatics represents a robust mechanism of encapsidation and virion stabilization.

To confirm that the results derived within our model of RNA branching, corresponding to a simple description of the RNA secondary structure, are indeed robust we also propose an alternative self-interacting linear chain model of RNA based on the assumption that RNA can be described as a linear polymer, *i.e.*, possesses no branch-points and only two end-points, but self-interacts with short-ranged attractive interactions describing the self-pairing of RNA segments [59]. As for the rest, we assume again that the capsid wall can be modeled as a thin, charged spherical shell with uniform surface charge density. The free energy corresponding to this model is again given by Eq. (4.1), except that the polymer chain is now linear, implying that

$$f_e, f_b \longrightarrow 0, \tag{4.10}$$

and the self interaction term $W[\Psi]$ thus changes to

$$W[\Psi] = \frac{1}{2}(v - a^3\beta s w)\Psi^4 + \frac{1}{6}u\Psi^6, \tag{4.11}$$

with s the average fraction of self-interacting chain segments, *i.e.*, base-pairs, and w is the corresponding short-range binding energy. Note that we included the next, Ψ^6 term in the virial expansion in Eq. 4.11, with $u > 0$ in order to stabilize the free energy since $(v - a^3\beta s w)$ can in general become negative. Variation of the free energy yields the same Euler-Lagrange equations as given in Eqs. (4.6), (4.7), (4.8) subject to the constraint, Eq. (4.4). The results of this calculation are presented in Fig. 4.4 that illustrates the encapsidation free energy as a function of the number of monomers, N. As illustrated in the figure, the positions of the free energy minima move towards longer polymers (larger N) and the depth of the minima increase with increasing s, the average fraction of bound segments. At 10 mM salt, Fig. 4.4 shows that the minimum of the encapsidation free energy is located at $N = 632$ for $s = 0$

and at $N = 740$ for $s = 0.04$. The effect is again more pronounced at 100 mM salt in which the location of the minimum moves from $N = 524$ for $s = 0$ to $N = 903$ for $s = 0.04$. w is chosen 1 k_BT and $u = 0.5\ nm^6$ in our calculations.

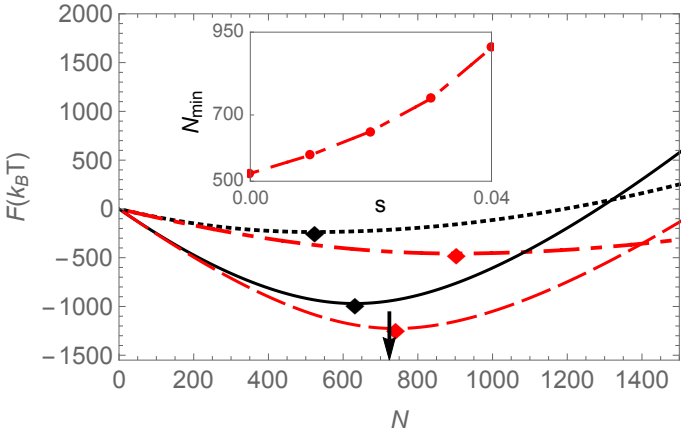


Figure 4.4: Encapsidation free energy (units of k_BT) as a function of monomer number for a self-interacting linear chain model with s=0 (solid and dotted) and s=0.04 (dashed and dotted-dashed lines) at two different values of μ, corresponding to salt concentrations 10 mM (solid and dashed lines) and 100 mM (dotted and dotted-dashed lines). The arrow indicates the monomer number at which the full virus particle is neutral ($Q_p = Q_c$). Inset shows the position of the minimum N_{min} vs. the average fraction of self-paired bases, s, for 100 mM salt concentration. Other parameters are $\upsilon = 0.5\ nm^3$, $w = 1\ k_BT$, $u = 0.5\ nm^6$, $\tau = -1\ e$, $\sigma = 0.4\ e/nm^2$, $b = 12\ nm$, $a = 1\ nm$ and $T = 300\ K$.

It thus seems that this rather different model, though presenting the same salient features of the system, yields the same qualitative behavior as discussed above for branched polymers. This substantiates our claim that the coupling between RNA self-interaction and capsid electrostatics represents a robust mechanism of encapsidation and virion stabilization.

In addition to investigating the different ways of modeling the secondary structures of RNA, we also studied the impact of different boundary conditions on the encapsidation

free energy and osmotic pressure. While all the results presented above correspond to the Neumann BC, $\hat{n}.\nabla\Psi|_s = 0$, we found that our conclusions do not depend on the type of BCs in that we obtained qualitatively the same results for the Dirichlet BC, $\Psi|_s = 0$. Although the Dirichlet BC changes the polymer density profile (see the inset of Fig. 4.5), the behavior of the free energy and the osmotic pressure remains qualitatively remarkably unaffected in that the minimum of the free energy does get deeper and moves towards longer chains as branching increases. As is clear from Fig. 4.5, at 100 mM salt the minimum of the free energy at $N \approx 401$ for a linear polymer with $f_b = 0$, is displaced to $N \approx 1103$ for a branched polymer with $f_b = 8.5$ when the Neumann BC is replaced by the Dirichlet BC for the polymer density field. Furthermore, for the Dirichlet BC at 10 mM salt, the free energy minimum is displaced from $N = 599$ for $f_b = 0$ to $N = 735$ for $f_b = 8.5$. Note that the value of f_b used for Dirichlet is chosen such that the ratio of number of branch points to the number of total monomers is almost the same as those for Neumann case.

We also calculated the osmotic pressure for Dirichlet BC using both branched and self-interacting linear chains. Consistent with the free energy results, we found that as the degree of branching or the average fraction of self-interacting chain segments increases, the osmotic pressure as a function of N becomes more negative and its minimum moves towards longer chains.

Further, we examined the impact on the free energy of the capsid surface charge density ($0.3 \leq \sigma \leq 0.9$), polymer charge density ($-2.0 \leq \tau \leq -0.5$) and Kuhn length ($0.5 \leq a \leq 2.0$). For both Dirichlet and Neumann BCs, we found that the optimal number of encapsidated monomers for linear chains is always such that number of charges on

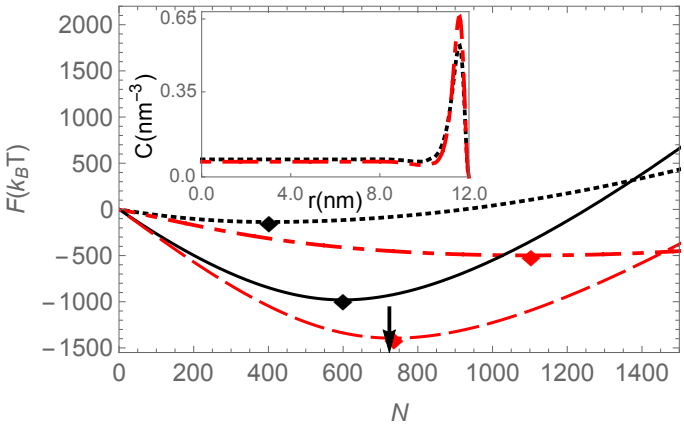


Figure 4.5: Encapsidation free energy (units of k_BT) vs monomer numbers for a linear chain with $f_b = 0$ (solid and dotted lines) and a branched chain with $f_b = 8.5$ (dashed and dotted-dashed lines) at two different sal concentrations μ, 10 mM (solid and dashed lines) and 100 mM (dotted and dotted-dashed lines) with the Dirichlet BC. The arrow indicates the monomer number at which the full virus particle is neutral ($Q_p = Q_c$). Other parameters take the values $\upsilon = 0.05\ nm^3$, $\tau = -1\ e$, $\sigma = 0.4\ e/nm^2$, $b = 12\ nm$, $a = 0.5\ nm$ and $T = 300\ K$. Inset shows the concentration profile for N=1000 with two different branching fugacities, $f_b = 0$ (linear chain) for the dotted line, and $f_b = 8.5$ (branched chain) for the dotted-dashed lines.

the polymer is less than those on the capsid, *i.e.*, the virus-like particles (VLP) are undercharged. In contrast, we found that the optimal length of the encapsidated branched polymers is larger than that of the linear polymers for all cases examined, resulting in overcharging of VLPs in many cases. We emphasize that while our findings are consistent with previous mean-field PE theories in that the VLPs with a linear polymer is undercharged [63], our results for linear polymers differ from recent numerical simulations [29] and the scaling theories [27] on the assembly of viral particles. While the overcharging for linear polymers, observed in Ref. [27] is due to the charges on the N-terminals and in Ref. [29] could be due to the solution conditions or the protein charge distribution, it is found that the branched structure of the polymer enhances overcharging, consistent with our studies.

It is difficult to determine the topology of large single-stranded viral RNAs in solution, but recent experiments indicate that the secondary structure does play an important role in the efficient packaging of RNA [9, 10]. The secondary structures can be predicted using a number of softwares, such as RNAsubopt (a program in the Vienna RNA package [100]), RNAfold (another program in the Vienna RNA package [100]) and mfold [119]. All these software tools, that are progressively unreliable for longer chains, estimate the free energy changes according to the base-pairing and the loop closure of ssRNA and the secondary structure of RNA results from base-pairing of G, U, C and A nucleotides. RNAfold and mfold calculate the possible sets of base-pairing corresponding to the minimum free energy, while RNAsubopt has an option to generate Boltzmann weighted secondary structures which can be used to calculate a meaningful ensemble average of any quantity. This software was successfully used [83, 110] to calculate the maximum ladder distance (MLD) and we applied RNAsubopt to calculate the thermally averaged number of branch points for RNA1 of BMV and CCMV to shed light on the experiments noted in the introduction on the competition between RNA1 of CCMV and BMV. We generated the ensemble of secondary structures using the RNA1 sequences of both BMV and CCMV obtained form the National Center for Biotechnology Information Genome Database [120], and then calculated the thermally averaged number of branched points of RNA1 of BMV and CCMV. We found that RNA1 of BMV has 65 branched points vs. 60.5 branched points of RNA1 of CCMV [101]. These numbers confirm the experimental results of Comas-Garcia et al. [9] that RNA1 of BMV would be preferentially packaged over RNA1 of CCMV. We note that although these programs were designed for the short RNAs, many important results

have been extracted through finding the ensemble average of the desired quantities for viral genomes of length $2500 - 10000$ nucleotides [83, 110].

The theoretical models presented in this paper clearly indicate the important role of the secondary structure of RNA on the assembly of ssRNA viruses. The secondary structure can be indeed invoked to explain the overcharging observed in RNA viruses, while it promotes the efficiency of RNA packaging by increasing the compactness of RNA in order to better fit into a small capsid. As shown above, the secondary structure of RNA clearly effects the osmotic pressure of the capsid; regardless of the details of the model as well as calculational details such as the form of the BCs, we obtain consistently negative osmotic pressures resulting from the presence of the negatively charged chain. The osmotic pressure becomes more negative for a branched polymer compared to the linear one.

Non-specific electrostatic interactions have emerged as the driving force for virus assembly through both the experimental as well as the theoretical studies [86, 10, 6, 63, 73]. In our two simple models we generalized the implementation of electrostatic interactions by coupling it to RNA topology. While this is an important step in realism of the modeling, the present level of description still cannot include the specific interactions (or packaging signals) into a complete picture of virus assembly. Further investigations on both specific and non-specific interactions could help understanding the structure of viruses and take steps on the development of antiviral drugs.

Acknowledgments

This work was supported by the National Science Foundation through Grant No. DMR-13-10687 (R.Z.). R. P. acknowledges the financial support of the Agency for research and development of Slovenia (ARRS) under Grants No. P1-0055 and J1-7435. The authors would like to thank the Aspen Center for Physics where some of the work was discussed during the Physics and Mathematics of Viral Assembly workshop.

4.5 Appendix

Derivation of the free energy

We consider RNA as a single polyelectrolyte in a good solvent in the presence of salt ions. There are N monomers of the polyelectrolyte chain, N^+ positive and N^- negative salt ions in the solvent. The microscopic degrees of freedom are the position of the monomers ($\mathbf{r}(s)$) and positive ($\mathbf{r}_\mathbf{i}^+$) and negative ($\mathbf{r}_\mathbf{i}^-$) ions . The partition function can be written as path integral over all configurations

$$\mathcal{Z} = \int \mathcal{D}\mathbf{r}(s)\mathcal{D}\mathbf{r}_\mathbf{i}^+\mathcal{D}\mathbf{r}_\mathbf{i}^- e^{-\beta\mathcal{H}} \tag{4.12}$$

where

$$\beta\mathcal{H} = \frac{3}{2a^2}\int_0^N \mathrm{d}s\,\dot{\mathbf{r}}^2(s) + \frac{\upsilon}{2}\int \mathrm{d}\mathbf{r}\,\hat{\rho}_m^2(\mathbf{r}) + \int_0^N \mathrm{d}s V(\mathbf{r}(s)) + \frac{\beta}{2}\int\int \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}'\;\hat{\rho}_c(\mathbf{r})\upsilon_c(\mathbf{r}-\mathbf{r}')\hat{\rho}_c(\mathbf{r}'). \tag{4.13}$$

The first term in Eq. (4.13) describes the ideal entropy of the chain, the second corresponds to the short range steric repulsions between monomers and the third term is an external

potential acting on the chain. The last term corresponds to the electrostatic interactions between the charges of monomers and ions. In Eq. (4.13) υ_c is the Coulomb interaction

$$\upsilon_c = \frac{1}{4\pi\epsilon\epsilon_0}\frac{1}{|\mathbf{r}-\mathbf{r}'|}, \tag{4.14}$$

and $\hat{\rho}_c$ is the charge density operator given by

$$\hat{\rho}_c(\mathbf{r}) = \tau\int_0^N \mathrm{d}s\ \ \delta(\mathbf{r}-\mathbf{r}(s)) + e\sum_i^{N^+}\delta(\mathbf{r}-\mathbf{r}_i^+) - e\sum_i^{N^-}\delta(\mathbf{r}-\mathbf{r}_i^-) + \rho_0(r). \tag{4.15}$$

Here, τ is the uniform monomer charge density along the polyelectrolyte and $\rho_0(r)$ is the charge density of the inner wall capsid in this system. To calculate the following integral in the partition function

$$\mathcal{Z}_{\text{salt}} = \int \mathcal{D}[\mathbf{r}_i^+]\mathcal{D}[\mathbf{r}_i^-]e^{-\frac{\beta}{2}\int\int \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}'\hat{\rho}_c(\mathbf{r})\upsilon_c(\mathbf{r}-\mathbf{r}')\hat{\rho}_c(\mathbf{r}')}, \tag{4.16}$$

we introduce a local charge density $\rho_c(\mathbf{r})$ and its auxiliary field $\phi(\mathbf{r})$ using the following identity

$$\begin{aligned} 1 &= \int \mathcal{D}[\rho_c(\mathbf{r})]\delta(\rho_c(\mathbf{r}) - \hat{\rho}_c(\mathbf{r})) \\ &= \int \mathcal{D}[\rho_c(\mathbf{r})]\mathcal{D}[\phi(\mathbf{r})]e^{i\beta\int \mathrm{d}\mathbf{r}(\rho_c(\mathbf{r})-\hat{\rho}_c(\mathbf{r}))\phi(\mathbf{r})} \end{aligned} \tag{4.17}$$

where the second line is the Fourier transform of the delta function. The auxiliary field $\phi(\mathbf{r})$ will turn out to be the electrostatic potential. We then replace the density operator $\hat{\rho}_c$ by the corresponding fluctuating density field ρ_c [121]. Multiplying Eq. 4.16 by Eq. 4.17 and using Eqs. 4.14 and 4.15 and the Hubbard-Stratonovich transformation, we find

$$\begin{aligned} \mathcal{Z}_{\text{salt}} = \int \mathcal{D}[\phi(\mathbf{r})]\,(\int \mathrm{d}\mathbf{r}e^{-i\beta e\phi(\mathbf{r})})^{N^+}(\int \mathrm{d}\mathbf{r}e^{i\beta e\phi(\mathbf{r})})^{N^-} \\ e^{-\frac{\beta\epsilon\epsilon_0}{2}\int \mathrm{d}\mathbf{r}(\nabla\phi(\mathbf{r}))^2}e^{-i\beta\tau\int_0^N \mathrm{d}s\phi(\mathbf{r}(s))}e^{-i\beta\int \mathrm{d}\mathbf{r}\rho_0(r)\phi(\mathbf{r})}. \end{aligned} \tag{4.18}$$

We use the same procedure as above to obtain the contribution of excluded volume interaction to the partition function,

$$e^{-\frac{1}{2}\upsilon\int \mathrm{d}\mathbf{r}\hat{\rho}_m^2(\mathbf{r})} = \int \mathcal{D}[\psi(\mathbf{r})]e^{-\frac{1}{2}\upsilon\int \mathrm{d}\mathbf{r}\psi^2(\mathbf{r})}e^{-i\upsilon\int_0^N ds\ \psi(\mathbf{r}(\mathbf{s}))}, \tag{4.19}$$

with ψ the auxiliary field representing the monomer density field. Plugging Eqs. 4.18 and 4.19 into Eq. 4.12, we find the partition function

$$\begin{aligned}\mathcal{Z}[N^+,N^-] = \int \mathcal{D}[\mathbf{r}(s)]\mathcal{D}[\phi(\mathbf{r})]\mathcal{D}[\psi(\mathbf{r})](\int \mathrm{d}\mathbf{r}e^{-i\beta e\phi(\mathbf{r})})^{N^+}(\int \mathrm{d}\mathbf{r}e^{i\beta e\phi(\mathbf{r})})^{N^-} \\ e^{-\frac{3}{2a^2}\int_0^N \mathrm{d}s\,\dot{\mathbf{r}}^2(s)-\int_0^N \mathrm{d}sV(\mathbf{r}(s))} \\ e^{-\frac{\beta\epsilon\epsilon_0}{2}\int \mathrm{d}\mathbf{r}(\nabla\phi(\mathbf{r}))^2-i\beta\tau\int_0^N \mathrm{d}s\phi(\mathbf{r}(s))-i\beta\int \mathrm{d}\mathbf{r}\rho_0(r)\phi(\mathbf{r})} \\ e^{-\frac{1}{2}\upsilon\int \mathrm{d}\mathbf{r}\psi^2(\mathbf{r})-i\upsilon\int_0^N ds\ \psi(\mathbf{r}(\mathbf{s}))}.\end{aligned} \tag{4.20}$$

We now switch to the grand-canonical ensemble modifying only the terms associated with the salt ions

$$\Xi[\mu] = \sum_{N^\pm}^{\infty}\frac{\mu^{N^++N^-}}{N^+!N^-!}\mathcal{Z}[N^+,N^-], \tag{4.21}$$

with μ the fugacity (density) of the monovalent salt ions related to the concentration of salt ions in the bulk. Inserting Eq. 4.20 into Eq. 4.21, the grand canonical partition function can be written as

$$\Xi = \int \mathcal{D}[\phi(\mathbf{r})]\mathcal{D}[\psi(\mathbf{r})]e^{-\beta\mathcal{H}_1[\phi(\mathbf{r}),\psi(\mathbf{r})]}\int \mathcal{D}[\mathbf{r}(\mathbf{s})]e^{-\beta\mathcal{H}_2[\mathbf{r}(s)]} \tag{4.22}$$

with the effective free energies

$$\beta\mathcal{H}_1[\mathbf{r}(s)] = \int_0^N \mathrm{d}s\Big(\frac{3}{2a^2}\dot{\mathbf{r}}^2(s) + V(\mathbf{r}(s)) + i\beta\tau\phi(\mathbf{r}(s)) + i\upsilon\ \psi(\mathbf{r}(\mathbf{s}))\Big) \tag{4.23}$$

and

$$\beta\mathcal{H}_2[\phi,\psi] = \int \mathrm{d}\mathbf{r}\Big(\frac{\beta\epsilon\epsilon_0}{2}(\nabla\phi(\mathbf{r}))^2 + i\beta\rho_0(r)\phi(\mathbf{r}) - 2\mu\cos(\beta e\phi(\mathbf{r})) + \frac{1}{2}\upsilon\psi(\mathbf{r})^2\Big) \tag{4.24}$$

The polymer part of the partition function is similar to the Feymann integral of the Hamiltonian $\mathcal{H} = -\frac{a^2}{6}\nabla^2 + U(\mathbf{r})$ with the potential $U(\mathbf{r}) = V(\mathbf{r}) + i\beta\tau\phi(\mathbf{r}) + i\upsilon\ \psi(\mathbf{r})$ and imaginary time $t \to is$ [59]. We assume that the chain is very long (total number of monomers $N \to \infty$) with a well defined energy gap such that the ground state approximation is valid. Thus, we have

$$\begin{aligned}\int \mathcal{D}[\mathbf{r}(\mathbf{s})] e^{-\beta\mathcal{H}_1[\mathbf{r}(s)]} \approx e^{-NE_0} &= e^{-N min\{\frac{<\Psi_0|\mathcal{H}|\Psi_0>}{<\Psi_0|\Psi_0>}\}} \\ &= \exp\Bigg(- \int \mathrm{d}\mathbf{r} \Bigg\{ \frac{a^2}{6}|\nabla\Psi_0(\mathbf{r})|^2 + V(\mathbf{r})|\Psi_0(\mathbf{r})|^2 \\ &\quad + i\beta\tau\phi(\mathbf{r})|\Psi_0(\mathbf{r})|^2 + i\upsilon\ \psi(\mathbf{r})|\Psi_0(\mathbf{r})|^2 - \lambda(\Psi_0(\mathbf{r})^2 - \frac{N}{V}) \Bigg\} \Bigg) \end{aligned} \quad (4.25)$$

with Ψ_0 the eigenfunction and E_0 the eigenenergy of the ground state. The Lagrange multiplier λ is introduced to normalize the wave function. Plugging Eq. 4.25 into Eq. 4.22 and integrating out the ψ field, we find the grand canonical partition function as

$$\Xi = \int \mathcal{D}[\Phi(\mathbf{r})] e^{-\beta\mathcal{F}} \quad (4.26)$$

with

$$\begin{aligned}\beta\mathcal{F} = \int \mathrm{d}\mathbf{r} \Bigg\{ &\frac{a^2}{6}|\nabla\Psi_0(\mathbf{r})|^2 + V(\mathbf{r})|\Psi_0(\mathbf{r})|^2 + \beta\tau\Phi(\mathbf{r})|\Psi_0(\mathbf{r})|^2 \\ &+ \frac{1}{2}\upsilon|\Psi_0(\mathbf{r})|^4 - \lambda(\Psi_0(\mathbf{r})^2 - \frac{N}{V}) - \frac{\beta\epsilon\epsilon_0}{2}|\nabla\Phi(\mathbf{r})|^2 \\ &+ \beta\rho_0(r)\Phi(\mathbf{r}) - 2\mu\cosh(\beta e\Phi(\mathbf{r})) \Bigg\} \end{aligned} \quad (4.27)$$

where we introduce the transformation $\Phi \to i\phi$ with Φ being the mean electrostatic potential. Due to the absence of an external potential, $V(\mathbf{r}) = 0$ and the capsid charge density is $\rho_0(\mathbf{r}) = \sigma\delta(z)$ with σ the surface charge density. This leads then to Eq. 4.1 considering

the constraint given in Eq. 4.4. Note that Eq. (4.27) is for a linear chain with $f_1 = 0$ and $f_3 = 0$. For branched polymers in the absence of electrostatic interactions, see Ref. [117].

Chapter 5

Functional analysis of the N-terminal basic motif of a eukaryotic satellite RNA virus capsid protein in replication and packaging

5.1 Introduction

Satellite tobacco mosaic virus (STMV) was originally found in association with helper virus (HV) Tobacco mosaic virus (TMV strain U5) in natural infections of tree tobacco (*Nicotiana glauca*) [122]. The replication of STMV depends on the RNA-dependent

RNA polymerase (RdRp) encoded by TMV U5, which has since been renamed Tobacco mild green mosaic virus. Other strains of TMV such as U1 and U2 also have been shown to support the replication of STMV [122]. The genome of STMV is composed of a single-stranded, positive-sense RNA of 1059 nucleotides (nt)[123, 124]. STMV RNA does not show appreciable sequence homology with helper TMV RNA, except for the 3'-terminal150 nt that exhibit 65% homology with the corresponding tRNA-like structure (TLS) located at the 3' end of TMV RNA [123]. Biologically active icosahedral virions of STMV are 17-18 nm in diameter with a $T = 1$ symmetry [125] and are assembled from 60 identical capsid protein (CP) subunits of 17.5 kDa [123]. The 3D structure of STMV has been determined at 1.8 Å [126].

A hallmark feature of STMV is that it is the only satellite virus, whose HV is a rod-shaped tobamovirus. Additionally, STMV stands out among other spherical viruses because of the high degree of order adopted by the RNA within the capsid [31]. Despite its simple genome organization, information concerning STMV replication is limited [127, 128]. Previous studies [127] revealed that a mutation engineered to completely block CP synthesis severely reduced the replication of STMV while a variant designed to express only the N-terminal 13 amino acid (aa) motif retained wild type replication via an unknown mechanism.

Using an *Agrobacterium*-based transient expression system (agroinfiltration) we recently reported several interesting properties of STMV when expressed in the presence and absence of its HV-RdRp [129]. In the current study, we extended the agroinfiltration system to further shed light on the contribution of the CP, more specifically the N-terminal 13aa motif to STMV replication. We have engineered a series of mutations affecting the

positively charged amino acids located within the N-terminal 13aa motif and evaluated their effect on replication and packaging. Application of a wide range of molecular techniques revealed that replication and packaging are regulated by protein-protein (i.e. CP and HV-RdRp) and protein-RNA-protein (i.e. HV-RdRp-STMV RNA-CP) interactions. First, we found that, the N-terminal 13aa motif has a regulatory role in STMV replication. Second, the genomic context of the N-13aa motif has a profound influence in modulating (+) and (−)-strand accumulation. Third, a single positively charged arginine residue located at the third position in the N-terminal 13aa motif of the CP is obligatory for the interaction between CP and HV-RdRp. Fourth, a Far-Northwestern analysis revealed that CP is obligatory to promote interaction between HV-RdRp and STMV RNA (i.e. protein-protein-RNA interaction). Finally, we observed that STMV CP bearing alanine substitutions for positively charged residues located at positions 5, 7, 10 and 11 specifically are defective in packaging full-length STMV genomic RNA. This study shows that packaging and replication of STMV RNA are regulated by the N-terminal 13aa motif of the CP.

5.2 Results

Evidence showing that the N-terminal 13aa motif of STMV CP is obligatory for replication

The genome organization of STMV RNA and the sequence of the first N-terminal 25 aa of the CP gene are shown in Fig. 7.1A. Among these 25 aa, the first 13aa are rich in basic residues and are predicted to interact with RNA inside the assembled capsid [130]. Application of DisProt [131] predicted that the N-terminal 25 aa region is disordered while

the remaining 134 aa are highly ordered (Fig. 7.1B). Preliminary data on the replication of STMV in plant protoplasts suggested a role for STMV CP and more specifically for the N-terminal13aa in replication [127]. We opted to extend the *Agrobacterium*-mediated transient expression system (agroinfiltration) developed recently for STMV in our lab [129] in dissecting the mechanism involving the CP-regulated STMV replication. To this end, we constructed three independent variant agrotransformants of wt STMV (pSTMV; Fig. 7.1C). These are as follows: (i) Agrotransformant CPKO was constructed by substituting the translation initiating methionine codon located at nucleotide 162-164 with a stop codon (i.e. ^{162}AUG$^{164}\rightarrow{}^{162}$UAG^{164}) (Fig. 7.1C); (ii) agrotransformant CPΔ13aa was designed to express the CP devoid of the first 13aa by engineering a stop codon at the translation initiation site and a start codon at aa14 (Fig. 7.1 C) and (iii) agrotransformant CP13aa was designed to express only the first 13aa by engineering a stop codon at aa position 14 (Fig. 7.1C).

Each of the above-mentioned three agrotransformants was mixed with agrocultures of pRP to provide the HV-RdRp (Fig. 7.1C) and co-infiltrated into N. *benthamiana* leaves. Plants co-infiltrated with pSTMV and pRP served as positive controls. At five days post infiltration (dpi), total RNA and protein preparations were isolated and subjected respectively to duplicate Northern blot hybridization to detect progeny (+) and (−)-RNA and Western blot analysis. Duplicate Northern blots probed for assessing the affect of CP on the accumulation of progeny (+) and (−)-strands are shown in Fig. 7.2A. Quantitative analysis of progeny RNA with respect to the wt control is tabulated and shown in Fig. 7.2B.

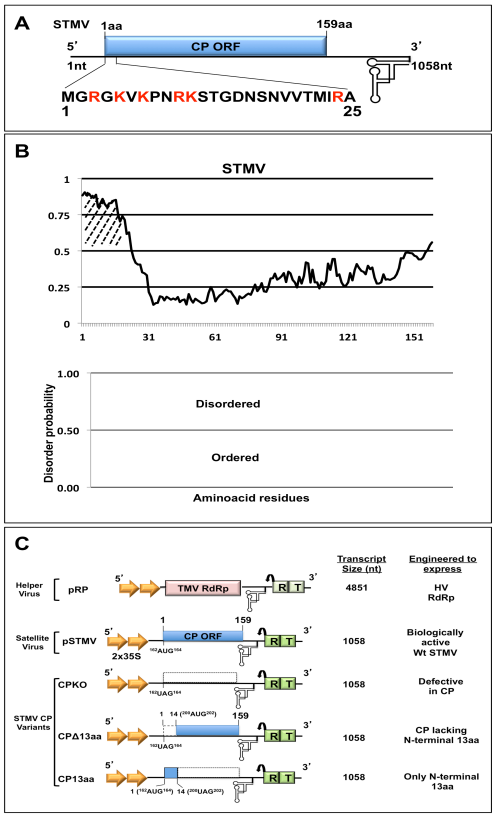
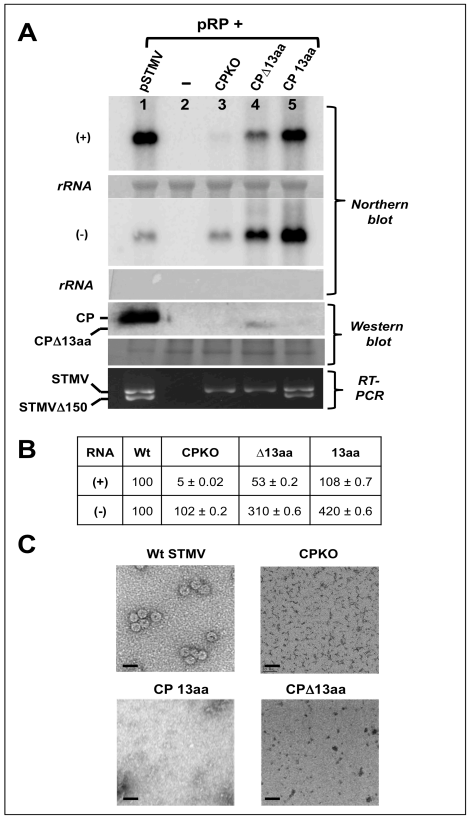
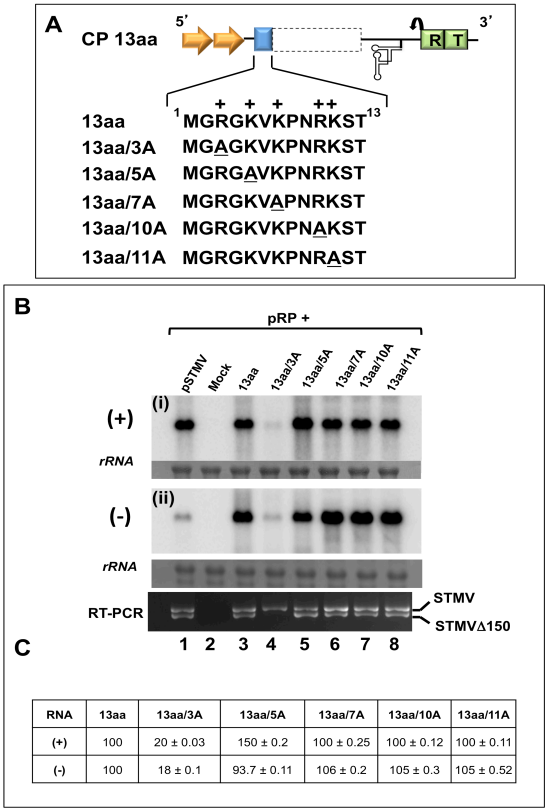


Figure 5.1: Genome organization of STMV. (A) Genome organization of Satellite tobacco mosaic virus (STMV) highlighting the 25 N-terminal amino acids (aa) of its coat protein ORF. The positively charged arginine (R) and lysine (K) residues are highlighted in red. (B) Graphic representation of the disordered probability of the 159 aa residues of STMV CP was by DisProt analysis. The hatched region indicates a previously identified RNA-binding domain (see text for details). In the bottom panel, aa having a score equal or above 0.5 and below 0.5 are considered to be disordered and ordered, respectively. (C) T-DNA based agro-constructs of STMV and its CP derivatives. Characteristic features of agroconstructs pRP and pSTMV were described previously [129]. CPKO, an STMV CP variant is characterized by having a ^{162}UAG^{164} stop codon in the place of an ^{162}AUG^{164} start codon. CP variant CPΔ13aa was engineered to express CP lacking the N-terminal 13aa motif due to the presence of a stop codon between nt 162-164 (^{162}UAG^{164}) and a start codon between nt 200-202 (^{200}AUG^{202}). CP variant CP13aa is engineered to express only the N-terminal 13aa motif by having a stop codon at amino acid position 14 located between nt 200-202 (^{200}UAG^{202}). Each agroconstruct contains in sequential order (L to R), double *Cauliflower mosaic virus* (CaMV) 35S promoters (indicated by arrowheads), a ribozyme (R, indicated by a bent arrow) derived from *Tobacco ring spot virus* (TRSV) and a NOS terminator (T). The size in nucleotides (nt) of ectopically expressed RNA transcripts and the expected nature of proteins in each case are shown to the right. The numbers shown in parentheses are the numbers of non-viral RNA nucleotides left after self-cleavage by the ribozyme (R).



RNA	Wt	CPKO	Δ13aa	13aa
(+)	100	5 ± 0.02	53 ± 0.2	108 ± 0.7
(-)	100	102 ± 0.2	310 ± 0.6	420 ± 0.6

Figure 5.2: STMV progeny analysis. (A) STMV plus (+) and minus (−) strand progeny RNA and CP were respectively probed by Northern and Western blot analysis. RT-PCR assays of wt STMV and the indicated CP variants were performed using total RNA isolated from leaves harvested at 4 dpi. (B) Quantitative analysis of (+) and (−) progeny RNA of wt STMV and each CP variant was performed by scanning the autoradiographs using a Typhoon Phosphoimage scanner (Model 9410, GE Health Care) followed by ImageQ so ware analysis. Data was quantitated and compared internally with each levels of (+) and (−)-strands to that of wild type. Conditions for Northern and Western blots and RT-PCR analysis are as described previously [129]. The positions of full length and truncated CP, full length and truncated STMV RNA are indicated on the left. (C) Virion analysis [129]. Electron microscopic images of negatively stained (with uranyl acetate) purified virions of wt STMV, CPKO, CP 13aa and CPΔ13aa are shown. Bar = 50 nm.



RNA	13aa	13aa/3A	13aa/5A	13aa/7A	13aa/10A	13aa/11A
(+)	100	20 ± 0.03	150 ± 0.2	100 ± 0.25	100 ± 0.12	100 ± 0.11
(-)	100	18 ± 0.1	93.7 ± 0.11	106 ± 0.2	105 ± 0.3	105 ± 0.52

Figure 5.3: Mutational analysis of the N-terminal 13aa motif of STMV CP. (A) Characteristic features of the agroconstruct of CP 13aa is shown in Fig. 7.1 C. The N-terminal 13aa sequence of the CP is indicated below the line diagram. In the clone 13aa, the location of the positively charged amino acids are indicated by a + sign. In each variant clone the location of the engineered alanine mutation is indicated by an underline. The nomenclature used to designate each variant is exemplified by 13aa/XA, where X is characterized by an alanine (A) substitution for arginine (R, at positions 3 and 10) or lysine (K at positions 5, 7 and 11). (B) Progeny analysis of N-terminal 13aa motif variants by Northern blot and RT-PCR. Duplicate blots containing total RNA of N. *benthamiana* leaves in ltrated with the indicated samples (lanes 1 through 8) were hybridized with a ^{32}P-labelled STMV probe to detect either plus (panel i) or minus (panel ii) sense RNA. Ribosomal RNA (rRNA) represents the loading control. Conditions used for Northern blot hybridization and RT-PCR is as described in the Fig. 7.2 A legend. Positions of STMV and STMVΔ150 are indicated to the right. (C) Progeny RNA quantification. Northern blots shown in panel B were scanned using a Phosphoimager. Since each mutation was engineered into the genetic background of 13aa, the absolute values of accumulated (+) and (−)-strand progeny RNA for each mutant were compared internally to the 13aa.

This experiment was repeated three times and we consistently observed that the presence of CP, more specifically the N-terminal13aa, had profound influence on progeny accumulation. Quantitative data in Fig. 7.2B represents the relative (+) and (-)-strand accumulation for each CP mutant when compared internally to that of wt. It was observed that, complete absence of CP, exemplified by the behavior of CPKO, decreased the plus-strand accumulation by 90% while minus-strand accumulation was not affected. By contrast, CP lacking the N-terminal 13aa motif (i.e. CPΔ13aa) reduced the plus-strand accumulation by 47% while a 3-fold increase in minus-strand accumulation was observed. Interestingly, expression of only the N-terminal 13aa motif had no detectable effect on plus-strand accumulation while a 4-fold increase in minus-strand accumulation was observed (Fig. 7.2 B). Taken together, these results suggested that the N-terminal13aa region of the STMV CP has two independent roles for regulating (−) and/or (+)-strand synthesis (see Discussion). With respect to CP accumulation, as expected, in addition to the wt control, a detectable level of a faster migrating CP was accumulated for variant Δ13aa but not for other variants (Fig. 7.2A). To detect the CP of 13aa by Western blot, the 13aa was FLAG tagged and its expression was confirmed by using anti-FLAG monoclonal antibodies (Fig 7.6A).

Recently, we demonstrated that production of a truncated form of STMV RNA is a hallmark feature associated with STMV replication in N. *benthamiana* [129]. It was observed that with either mechanical inoculation of STMV with its HV or co-expression of STMV with HV-RdRp via agroinfiltration, a truncated non-replicative form of STMV lacking the 3' terminal 150 nt (i.e. STMVΔ150) was also accumulated along with the wt genomic RNA. Additional experiments demonstrated that production of STMVΔ150 is linked to

HV-dependent replication and is not a product of ribonuclease activity [129].To verify the production of STMVΔ150 in leaves expressing each of the three variant agrotransformants (CPKO, Δ13aa and 13aa), total RNA preparations were subjected to RT-PCR. Production of STMVΔ150 was associated with either wt STMV (Fig. 7.2 A, lane 1 in RT-PCR panel) or a variant competent to express the 13aa (i.e. CP 13aa in Fig. 7.2A, lane 5 in RT-PCR panel) but not with those variants defective in expressing 13aa i.e. CPKO or CPΔ13aa (Fig. 7.2A, lanes 3 and 4 in RT-PCR panel). Since the production of STMVΔ150 is linked to HV-dependent replication [129] and the fact that expression of the 13aa plays a crucial role in STMV replication (Figs. 7.2A and 7.3B) [128]., these observations suggest that expression of the 13aa is linked to the production of STMVΔ150 by an unknown mechanism.

To test the effect of the N-terminal13aa deletion on virion assembly, agrotransformants of CPKO, CPΔ13aa and 13aa were co-expressed with agrotransformant pRP in N. *benthamiana* leaves. Plants co-infiltrated with pSTMV and pRP served as positive controls while those infiltrated only with empty vector served as negative controls. At 4 dpi, virions were purified, negatively stained and examined by Transmission Electron Microscopy (TEM). Icosahedral virions of 18 nm characteristic of STMV were recovered from leaves infiltrated with pSTMV and pRP (Fig. 7.2 C). By contrast, no virions were detected in leaves infiltrated with any of the three CP variants (Fig. 7.2 C). In agreement with previous findings [126], our observations further confirm that the N-terminal13aa region is required for virion assembly.

A positively charged amino acid at position 3 is obligatory for efficient replication of STMV

Results shown in Fig. 7.2A accentuate the importance of the N-terminal13aa region in STMV replication. To precisely identify the role of positively charged aa encompassing the N-terminal13aa of STMV CP on replication, an alanine residue was substituted for a desired arginine or lysine residues normally located within the 13aa N-terminal region resulting in the construction of a set of five variants of CP 13aa (i.e. 13aa/3A, 13aa/5A, 13aa/7A, 13aa/10A and 13aa/11A in Fig. 7.3A). Each of these CP agrotranformants was co-infiltrated with pRP into N. *benthamiana* leaves and progeny were evaluated using Northern blot and RT-PCR analyses (Fig.7.3). Northern blots were probed for assessing the affect of CP and its variants on the progeny accumulation (Fig. 7.3B) and quantitated as described above (Figs. 7.3C).

First, we evaluated the affect of the N-terminal 13aa motif and five variants on plus-strand accumulation (Fig. 7.3B, panel i). Except for variant 13aa/3A, plus-strand accumulation for the remaining four variants was indistinguishable from that of the internal control (i.e. wt) (Fig. 7.3C). Secondly, we evaluated the effect of these mutants on minus-strand accumulation (Fig. 7.3B, panel ii). Since each of the five alanine substitutions were engineered into the genetic background of the wt 13aa construct, data was normalized against wt 13aa. Except for variant 13aa/3A, where the (−)-strand accumulation was equivalent to that of the control (i.e. 13aa), a 3-fold increase in minus-strand accumulation was observed for the remaining four variants (Fig. 7.3C). Taken together, these observations

accentuate the significance of that having a positively charged residue at their respective positions, specifically at 3^{rd} position is obligatory for sustained replication.

Results of RT-PCR analysis evaluating the production of STMVΔ150 are shown in Fig. 7.3B. Production of STMVΔ150 was found associated with all variants except for 13aa/3A (Fig. 7.3B). Collectively, data encompassing replication profiles (Fig. 7.3 B), quantitative analysis of strand asymmetry (Fig. 7.3 C) and RT-PCR (Fig. 7.3B) analysis suggested that the arginine and the lysine residues at the N-terminus plays a significant role for maintaining wild type replication of the STMV RNA. More specifically a positively charged residue at the 3^{rd} position is intimately involved with the replication of STMV (also see below).

Replication of full-length STMV CP variants harboring N-terminal mutations

Next we wanted to evaluate the replication and packaging phenotypes of full-length STMV CP bearing each of the six N-terminal alanine substitution variants (Fig. 7.3A). Consequently, each of the five-alanine substitution variants (Fig.7.3A) was engineered into the CP ORF of the full-length STMV agrotransformant (Fig. 7.4A) and its effect on replication, translation, virion assembly and genome packaging was evaluated. Results are shown in Figs. 7.4 and 7.5.

Northern blot hybridization results on the effect of the five-alanine mutations on STMV replication are shown in Fig. 7.4 B and the quantitative data on the relative accumulation levels of (+) and (−)-strand progeny is shown in Fig. 7.4C. Consistent with the data shown in Fig. 7.3B, despite near wt levels of (−)-strand accumulation, a ¿ 90% reduction in

the accumulation of (+)-strand was observed for variant 3A (Fig. 7.4B, lane 1; Fig. 7.4C). For the remaining four variants (ie. 5A, 7A, 10A and 11A), a 30 to 50-fold stimulation (in relation to wt) in the accumulation of (+)-strand was observed (Fig. 7.4B, lanes 2-5) which was contrary to the accumulation level seen with the same variants engineered in the CP 13aa construct (Fig. 7.3). Results of RT-PCR assays used to evaluate the production of STMV Δ150 are similar to those shown in Fig. 7.3B, i.e. STMV Δ150 was present in total RNA preparations of all variants except 3A (Fig. 7.4B, bottom panel).

Results shown in Figs. 7.3 and 7.4 clearly indicate that at position 3, a positively charged amino acid residue is preferred. To substantiate this, the arginine residue at position 3 was substituted with another basic amino acid namely, a lysine residue (i.e. 3R to 3K, Fig. 7.4A) and its effect was examined on replication, CP synthesis and production of STMV Δ150. Results summarized in Fig. 7.4(D and E) confirm the importance of a positively charged residue at position 3.

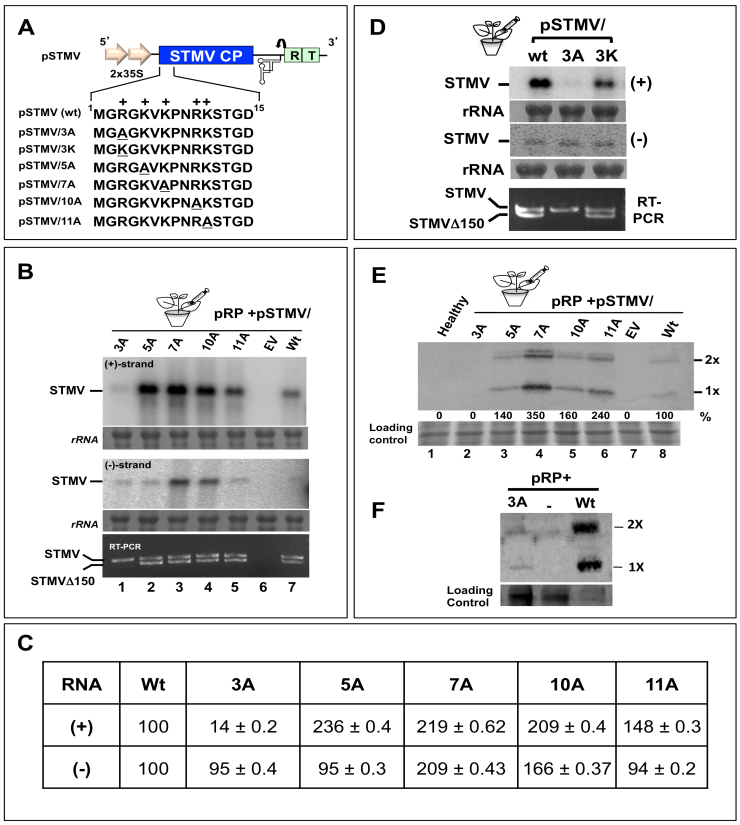
Effect of N-terminal mutations on STMV assembly and packaging

Western blot analysis revealed that all four other variants (5A, 7A, 10A and 11A) were competent to translate higher levels of CP than wt. Among four variants, CP production was more prominent with variant 7A than others (Fig. 7.4E). By contrast, CP for variant 3A was only detectable when a 10-fold excess of total protein was loaded (Fig. 7.4F). To verify whether these variants were competent to assemble into virions or virion like particles (VLPs), virions were purified from infiltrated N. *benthamiana* leaves. Fig. 7.5 summarizes the results. EM and Northern blot hybridization evaluated the physical morphology of purified virions and their packaging phenotypes, respectively. The biological

nature of the purified virions was tested by an infectivity assay. Results are summarized in Fig. 7.5. Icosahedral virions of 18 nm were detected in wt as well as in variants 5A, 7A, 10A and 11A but not in variant 3A (Fig. 7.5 A). Northern blot analysis of viron RNA profiles revealed interesting results. For example, only virions of wt contained full-length STMV RNA as well as truncated STMV (i.e. STMV Δ150) (Fig. 7.5 B, lane 6), a characteristic feature of STMV in N. *benthamiana*[129]. By contrast, virions of all four variants contained only the truncated STMV Δ150 but not full length (Fig. 7.5B, lanes 2-5). Presence of full length and STMV Δ150 in virion preparations of wt and the absence of full length and presence of STMV Δ150 in all four CP variants were further confirmed by an RT-PCR assay (Fig. 7.5C). In contrast to the wt control (Fig. 7.5D, lane 1), the inability to induce infection in N. *benthamiana* plants mechanically inoculated with virions of HV and each of the four CP variants (Fig. 7.5D, lanes 2-6) further confirm that CP variants are incompetent to package full length STMV RNA. Reasons for this defective packaging phenotype are considered under the Discussion.

Evidence for a physical interaction between HV-RdRp and STMV CP

Although replication of STMV depends on HV-RdRp, results shown in Figs. 7.2-7.5 demonstrate that STMV CP, more specifically its N-terminal 13aa motif, has a regulatory role in the replication. Therefore to test whether STMV CP physically interacts with HV-RdRp, we performed the following experiment. One set of N. *benthamiana* plants was infiltrated to co-express HV-RdRp (i.e. pRP, Fig. 1C) and either wt STMV or one of three selected N-terminal CP variants (3A, 5A and 7A). Another set of N. *benthamiana*



RNA	Wt	3A	5A	7A	10A	11A
(+)	100	14 ± 0.2	236 ± 0.4	219 ± 0.62	209 ± 0.4	148 ± 0.3
(-)	100	95 ± 0.4	95 ± 0.3	209 ± 0.43	166 ± 0.37	94 ± 0.2

Figure 5.4: Replication of full-length STMV RNA bearing mutations in the N-terminal 13aa motif. (A) Mutations engineered into the N-terminal 13aa motif of CP. Positively charged amino acid residues are indicated by a + sign. Alanine or lysine substitutions at designated positions are underlined. Nomenclature used to designate each mutation is the same as described under the Fig. 7.3 legend. (B-E) Progeny analysis. Total RNA and protein samples were isolated from leaves harvested at 4 dpi. Panels B and D summarize Northern blot analysis of (+) and (−)-strand accumulation for the indicated variants and RT-PCR analysis. Panel C summarizes the absolute values of accumulated (+) and (−)-strand progeny RNA. (E) Western blot analysis of CP expression for indicated variants. Following co-infiltration of pRP and each variant agrotransformant progeny were analyzed by Northern blot hybridization, RT-PCR, Western blot analysis and the absolute values of accumulated (+) and (−)-strand progeny RNA.as described in the Figs 7.2 and 7.3 legends. (F) Detection of CP 3A by Western blot analysis using a 10-fold excess of the indicated protein samples. Positions of STMV and STMVΔ150 are indicated to the left in panels B and D. Ribosomal RNA (rRNA) represents loading controls for Northern blots. Position of monomeric (1x) and dimeric (2x) forms of STMV CP is indicated to the right in panel E. Accumulation of (+) and (−)-strand progeny was quantitated as described in the Fig. 7.2 legend.

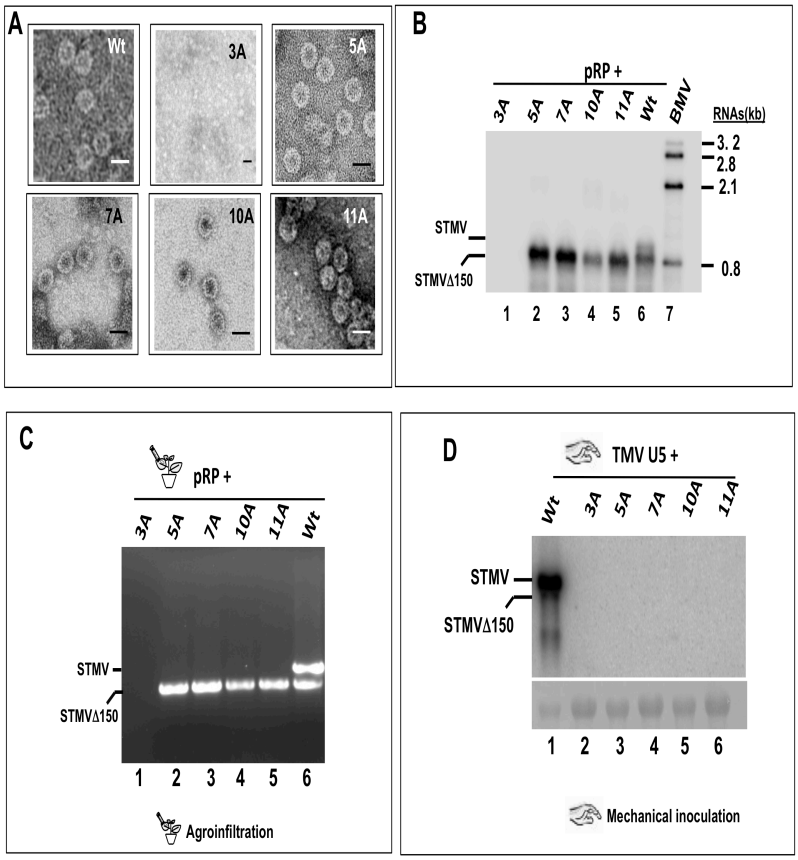


Figure 5.5: Electron microscopy and packaging phenotypes of N-terminal variants. (A) Electron microscopic images of virions purified from indicated samples. Bar = 20 nm. (B) Virion RNA analysis of indicated samples (lanes 1 through 6) by Northern blot hybridization. BMV RNA (lane 7) was used as a size marker. (C) RT-PCR analysis of virion RNA recovered from N. *benthamiana* plants co-in ltrated with pRP and either the indicated STMV variant inocula (lanes 1 through 5) or wt STMV (lane 6). (D) Biological assays in N. *benthamiana* plants mechanically inoculated with helper virus TMV-U5 and virion RNA of wt and the indicated STMV variant samples.

plants was infiltrated to co-express pRP and only 13aa tagged with FLAG ($13aa^{c-FLAG}$). Plants co-infiltrated with P19 and either pSTMV or $13aa^{c-FLAG}$ served as controls. At 4 dpi, total proteins were extracted from infiltrated leaves and incubated with anti-126 kDa antibodies, followed by the precipitation of the complex with protein G-agarose beads and Western blotting with anti-126 kDa, CP or Anti-FLAG antibodies. Results are shown in Fig. 7.6A. It was observed that wt CP, variants 5A and 7A, but not 3A, were co-precipitated with HV-RdRp (i.e. Anti-126 kDa) (Fig. 7.6A, panel ii). These results validate that a direct or indirect interaction exists between HV-RdRp and STMV CP and the arginine residue located at position 3 is obligatory in promoting this interaction. In addition, RNase A treatment did not disrupt the co-precipitation of anti-126 kDa and CP (Fig. 7.6B), suggesting that RNA is not involved in promoting the interaction between HV-RdRp and CP.

Evidence showing interaction of STMV RNA with HV-RdRp requires CP

Northern blot results (Figs. 7.2-7.4) demonstrate the importance of CP in STMV replication while results of Co-IP assays (Fig. 7.6A) demonstrate that CP physically interacts with HV-RdRp. Does interaction of HV-RdRp with STMV RNA require CP? To answer this question, we performed a Far-Northwestern assay. Results are shown in Fig. 7.7(A, B). Since variant 3A translates inefficiently in vivo (Fig. 7.4F), a Northwestern analysis was performed using in vitro translation products of wt and four CP variants (CPKO, 3A, 5A and 7A) and ^{32}P-labeled STMV RNA as a binding probe. Interaction between STMV RNA and CPs of wt, variants 5A and 7A, but not with either CPKO or variant 3A, was detectable [Fig. 7A, panel (ii)]. These Northwestern analyses confirm interaction between

CP and RNA while Co-IP assays (Fig. 7.6A) established that CP physically interacts with HV-RdRp. To further evaluate whether CP is required to promote interaction between STMV RNA and HV-RdRp, we performed another Far-Northwestern analysis. Results demonstrated that the presence of STMV CP enabled the interaction between HV-RdRp and STMV RNA [Fig. 7.7B, panels (I to iii)]. Taken together these results provide evidence that CP enhances selection of STMV RNA templates by HV-RdRp.

5.3 Discussion

Viral replication is a complex process involving numerous macromolecular interactions [132]. In positive strand RNA viruses, although replication is primarily catalyzed by virus encoded RNA-dependent-RNA polymerase (RdRp), several lines of evidence suggest that CP is intimately associated with this active process to stimulate replication [132]. For example, CP-regulated viral RNA synthesis occurs via CP-RNA interactions as in the case of Alfalfa mosaic virus (AMV) [133, 134] or CP-RdRp interactions as in the case of member viruses of the Coronaviridae [135] and rubella virus [136] or CP functioning as an RNA chaperone [131, 137, 138]. The results presented here demonstrate that a cooperative interaction between HV- RdRp and STMV CP (Fig. 7.2A), more specifically the N-terminal 13aa region (Figs 7.3 and 7.4), plays an important role in STMV replication and packaging. Mutational analysis further revealed that a positively charged residue at the third position in the N-terminal 13aa motif (Figs 7.3B, 7.4B, D and 7.6A) promotes the interaction between HV-RdRp and STMV CP. These features are commonly shared with AMV [134, 139] and rubella virus [136]. In AMV, minus-strand synthesis requires the viral RNA to be in

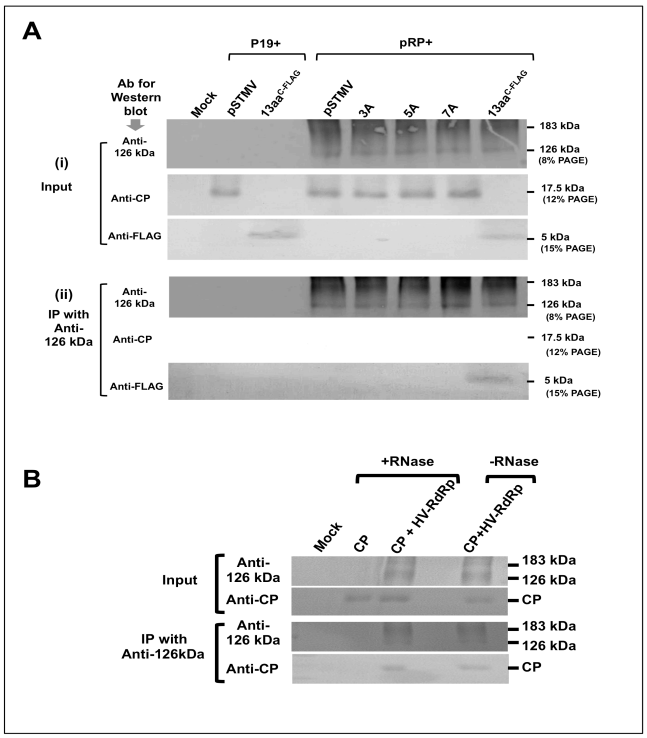


Figure 5.6: Co-immunoprecipiation assays.(A)HV-RdRp(pRP)was co-expressed with either pSTMV (i.e.wt) or the indicated STMV variants in N. *benthamiana* leaves by agroinfiltration. Mock inflatrated plants or plants inflatrated with P19 (a suppressor of RNA silencing) and either pSTMV or $13aa^{FLAG}$ served as controls. (i) Expression of HV-RdRp, STMV CP and its variants in total protein extracts was con rmed by Western blotting using anti-126kDa and anti-STMV CP antibodies, respectively. Total protein extracts from leaves in ltrated with the empty vector (mock) or CP only were used as controls. (ii) For co-immunoprecipitation, total protein extracts of each sample were incubated rst with anti-126-kDa antibody followed by complex precipitation with protein-G agarose. e resulting co-immunoprecipitated products were subjected to Western blotting using anti-126 kDa or anti-STMV CP or anti-FLAG antibodies. (B) STMV CP (CP) was expressed either independently or co-expressed with HV-RdRp in N. *benthamiana* leaves by agroinfiltration. Total protein extracts were divided into two batches: one batch was treated with RNase (+RNase) and the other remained untreated (-RNase). Co-immunopreciptation with anti-126 kDa followed by Western blot analysis was as described above.

a 3' pseudoknot conformation and to be free of CP; upon CP binding a conformational switch blocks minus-strand synthesis and favors plus-strand synthesis [140, 141]. However, since the 3' UTR of STMV inherently mimics a TLS, a conformational switch analogous to AMV is not required to initiate minus-strand synthesis. Similar levels of minus-strand accumulation between wt and the CP defective variant CPKO (Fig. 7.2A, compare lanes 1 and 3) suggest that in STMV minus-strand synthesis by HV-RdRp is independent of the CP.

Our results exemplify that STMV CP (Fig. 7.2), specifically the basic N-terminal 13aa motif (Figs 7.3A and 7.4A) actively participates in modulating (+) an/or (−)-strand synthesis. As discussed below, this active participation in STMV RNA replication depends on the genomic context of the N-terminal 13aa motif. First, the critical role played by the CP is exemplified when the replication profile of wt and mutant CPKO is compared (Fig. 7.2A). Since the levels of (−)-strand accumulation for wt and mutant CPKO are indistinguishable (Fig. 7.2A, compare lanes 1 and 3). It is reasonable to conclude that CP has no detectable effect on (−)-strand synthesis and a significant reduction in (+)-strand synthesis by mutant CPKO (Fig. 7.2A, lane 3) suggests CP plays a critical role in (+)-strand synthesis.

Co-expression of the wt 13aa (Fig. 7.2) along with HV-RdRp stimulated both (+) and (−) strands accumulation equally (Fig. 7.2A lane 5) whereas co-expression of Δ13aa significantly reduced (+)-strand accumulation compared to that of (−)-strand (Fig. 7.2 A compare lane 4 and 5). This suggests that the role of the N-terminal 13aa motif is attributed to (+) strand synthesis. However, on mutating four of the five positively charged

amino acids to alanine in the wt 13aa (i.e. 5A, 7A, 10A and 11A), the balanced accumulation of (+) and (−)-strand RNA was altered resulting in increased accumulation of (−)-strands (Fig. 7.3B). Interestingly, this scenario was changed when the above-mentioned four mutants were incorporated into the background of full length CP. This is exemplified by increased accumulation of (+)-strands over the (−)-strands (Fig. 7.4B). Collectively, these observations suggest that (+) and (−)-strand accumulation is regulated by the context of the N-terminal 13aa motif.

Another interesting outcome of our study is the intrinsic role played by a positively charged residue at position 3 of the N-terminal 13aa motif. For example, substitution of an alanine residue for an arginine at position 3 (i.e. 3R→3A; Figs 7.3A and 7.4A) had very little effect on (−)-strand synthesis (Fig. 7.3B, line 4; Fig. 7.4B, line 1) while plus-strand synthesis was down regulated by nearly 80% (Figs 7.3B, C and 7.4B, C). An analogous scenario was observed in AMV. Mutational analysis of the N-terminal basic residue of AMV CP revealed that substitution of alanine for lysine at position 18 completely abolished replication [139]. As discussed below, this down-regulation of progeny accumulation is due the absence of interaction of CP with HV-RdRp.

Results from co-immunoprecipitation analysis revealed that STMV CP interacts with HV-RdRp (Fig. 7.6A lane 4). Although this phenomena was previously observed in other RNA viruses such as Brome mosaic virus and Flock House Virus (15) but not for satellite viruses. Results of this study represents the first example demonstrating that STMV RNA replication requires a interaction between CP and its HV-RdRp. Furthermore, North-Western and Far North Western analyses (Fig. 7.7A,B) accentuated that the presence

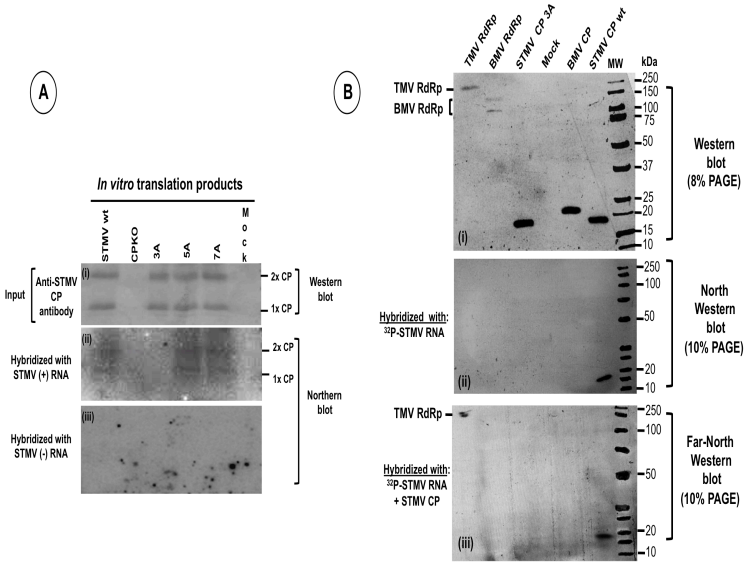


Figure 5.7: Northwestern and Far Northwestern blot assays. (A) RNA-CP interaction detected by Northwestern blot analysis: Triplicate blots containing in vitro translation products of the indicated samples were generated following 12% SD-PAGE analysis. The first blot (i) was subjected to Western blot analysis using anti-STMV CP antibody while the second (ii) and third (iii) blots were respectively subjected to hybridization with ^{32}P-labeled STMV (+) and (−)-strand riboprobes and visualized using autoradiography. The positions of monomeric (1x) and dimeric (2x) forms of STMV CP are indicated to the right. (B) Requirement of CP to promote interaction between HV-RdRp and STMV RNA was detected by Far Northwestern blot analysis: HV-RdRp (183 kDa) or, BMV RdRp (109 and 94 kDa) and BMV CP (19 kDa) (negative controls) and Wt and 3A variant of STMV CP (17.5 kDa) were fractionated by either 8% (panel i) or 10% SDS-PAGE (panels ii and iii), transferred to a nitrocellulose membrane and successively denatured and re-natured prior to hybridizing with32P-labeled STMV RNA probe only (ii) or radiolabelled probe and STMV CP (iii). Western blot shown in panel (i) represents input control analyzed using a mixture of antibodies raised against TMV-RdRp, BMV- RdRp, BMV CP and STMV CP antibodies.

of a positively charged residue at the 3^{rd} position is obligatory to promote CP-HV-RdRp interaction.

We recently demonstrated that replication of STMV in N. *benthamiana*, but not in N. tabacum, generates a truncated RNA (i.e. Δ150 characterized by lacking the 3' terminal 150nt) along with the full length RNA (9). Results of this study revealed that production of Δ150 is linked to HV-dependent replication since it was not detected in plants infiltrated with replication defective variants of STMV (i.e. CPKP and 3A; Figs 7.3 and 7.4).

In general, it is widely accepted that electrostatic interactions are the driving force for virion assembly in that the positively charged domains on viral CPs neutralize most of the negative charges of the RNA [88, 142, 55, 82, 78, 29, 50, 143]. Many in vitro viral self-assembly studies show that RNA packaging can be largely explained through electrostatic interactions [55]. Based on these experimental observations, we recently introduced a very simple generic model to study the spontaneous encapsulation of the genome by the CP [73]. In this model, we consider RNA as a negatively charged polymer interacting attractively with positively charged CP. More specifically, we use a phenomenological model to investigate the total free energy of the virion, $F_{virion} = F_{capsid} + F_{RNA} + F_{int}$, with F_{capsid} the free energy of the capsid, F_{int} the free energy associated with capsid-RNA interactions and F_{RNA} the free energy of RNA. Since STMV CP is positively charged, we model the capsid as a charged sphere so the term F_{capsid} involves the repulsive electrostatic interaction between positive charges on the CP. In our approach, RNA is modeled as a self-avoiding flexible branched chain. Thus the term F_{RNA} includes the contribution of RNA entropy, the secondary structure of RNA, and the excluded volume interaction related to the second

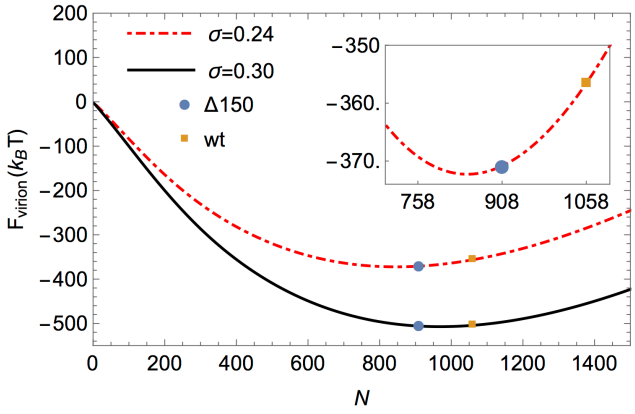


Figure 5.8: Free energy of assembly. Encapsidation free energy vs the genome length for wt CPs (solid black curve) and mutated CPs (dashed red curves). The blue dots correspond to the encapsidation free energy of STMVΔ150 RNA and the orange squares to wt STMV RNA. Energies are in units of thermal energy k_BT. See the text and reference [73] for further details of the model.

viral coefficient. We find the optimal length of RNA and its distribution within the capsid through minimization of the free energy F_{virion} explained above.

Data shown in Fig. 5.8 demonstrates the result of our calculations, F_{virion} vs genome length. The σ in the figure corresponds to the charge density of the inner wall of the capsid. It is 0.24 $electron/nm^2$ for the capsid constructed by mutated coat proteins (5A, 7A, 10A and 11A) and it is 0.3 $electron/nm^2$ for the wt capsid. The solid black curve shows the free energy of RNA encapsulated by wt proteins while the dashed red curves correspond to that by the mutated coat proteins. Blue dots and orange squares respectively represent the encapsidation free energy of STMVΔ150 (908nt) and wt RNA (1058nt). As illustrated in the figure, in the case of the mutated capsid, the encapsidation free energy of STMVΔ150RNA is significantly lower than that of wt RNA (14.28 k_BT) leading to the exclusive encapsulation of STMVΔ150 RNA. However, in the case of wt CP,

the encapsulation free energy of STMVΔ150 RNA is lower than that of wt RNA by only 1.47 k_BT, explaining why both STMVΔ150 and wt RNA can be encapsidated by wt CP. As a matter of fact, while the ratio of free STMVΔ150 to wt RNA is very low (24:76) [129], the ratio of packaged STMVΔ150 to wt RNA is much larger (55:45) indicating a clear preference for encapsidation of the shorter RNA. More specifically, these calculations reveal that the electrostatic interaction between mutated capsid proteins and full length RNA is not strong enough to compact the genome to fit in a small size capsid. In other words, due to the size of the STMV capsid, the energetic cost of encapsidation of negatively charged wt RNA is very high and thus the shorter genome is encapsidated more favorably even by wt capsid proteins. While the number of charges on the wt capsid proteins are sufficient to encapsidate the full length genome, it is important to note that wt RNA gets encapsidated much less efficiently than STMVΔ150 RNA. These theoretical results offer the basis to explain results presented in Fig. 7.5C.

As noted above, electrostatics clearly explains the results corresponding to variants 5A, 7A, 10A and 11A presented in Fig. 7.5C. However, a question that naturally arises would be why does the CP bearing the 3A mutation not encapsulate either STMVΔ150 or wt RNA? We offer the following explanation. Our observations (Fig. 7.5C) suggest that non-specific interactions are insufficient to drive the self-assembly of a ssRNA virus by CP. The presence of an arginine residue located at position 3 reveals the necessity of a specific interaction between RNA and CP for the successful assembly of a virion. The interaction could be due to a specific nucleation site on the CP for RNA or the specific CP-RNA interaction could alter the conformation of the CP resulting in a stronger protein-protein

interaction leading to virion assembly. The experiments presented in this paper show that an amino acid residue (i.e. arginine) at a specific site (i.e. position 3) is crucial for successful packaging. Identification and disruption of contacts between particular CP residues and RNA could contribute to our knowledge of developing novel antiviral drug strategies for RNA viruses pathogenic to pathogenic to humans and animals.

5.4 Methods

Agro-constructs and Agro-infiltration

The construction and characteristic features of an agro-construct of STMV (pSTMV; Fig. 7.1A) have recently been described [129]. All CP variants constructed in this study are incorporated into the genetic background of pSTMV using a mega PCR approach [144]. Briefly, to construct the N-terminal CP variants (Fig. 7.1 C) 3R → A, 5K → A, 7K→A, 10R → A, 11K → A and 5R+7K → 5A+7A (refer to as 57A; Fig. 1C), STMV CPKO, STMV 13aa, STMV 13aa/A and STMV 13aa/B(Fig 5A), the CP gene harboring nucleotides (nt) 159 to 614 was first amplified with the following forward primers (For 3R → A 5'GCTATGGGGGCAGGTAAGGTTAAACC3', 5' GCTATGGGGAGAGGTGCAGTTAA ACC3', for 5K → A 5' GGGGAGAGGTAAGGTTGCACCAAACC 3'; for 7K → A 5' GGTAAGGTTAAACCAAACGCAAAATCGACG 3'; for 10R → A 5' GGTTAAACCAAAC CGTGCATCGACG 3'; 5' for CPKO, 5'CTGTTTCCAGCTTAGGGGAGAGGTTAGGTTA AACC3'; for 13aa, 5'CGTAAATCGACGTAGGACAATTCGAATG 3'; for 13aa/A5'GCTAT GGGGGCAGGTAAGGTTAAACC 3'; 13aa/B 5' CCAGCTATGGGGAGAGCAGCAGCA GCAGCAGCAGCAGCATCGACGTAGGACAATTCG 3') and a commonly shared reverse

primer (5' GGCGACTTGTCGACAGTTGC 3'; SalI site is underlined). The resulting PCR products were gel purified and used as mega primers in a PCR reaction. A region spanning the sequence of pSTMV from nt768-614 was amplified using a forward primer (5'CGCCAAGCTTGCATGCCTGCAGG3'; *Hind*III site is underlined) and each mega primer for the respective variant prepared above. The resulting PCR products were digested with *Hind*III and *Sal*I and sub-cloned into a similarly treated pSTMV. The nature of all recombinant clones was verified by DNA sequencing. Transformation of wild type and variant pSTMV agroconstructs to *Agrobacterium* strain GV3101 followed by infiltration into the abaxial side of the fully expanded *N. benthamiana* leaves was as described previously [145].

Mechanical inoculation, progeny analysis, packaging assays and EM analysis

Procedures used for mechanical inoculation of *N. bethamiana* with STMV variant virions in the presence of TMV U5, isolation of total RNA from agroinfiltrated leaves, purification of STMV virions, and Northern blot analyses are as described previously [129, 146, 147]. For electron microscopy analysis, partially purified wild type and variant STMV virions were further subjected to purification using 10-40% sucrose gradient centrifugation and spread on glow-discharged grids followed by negative staining with 1% uranyl acetate prior to examination with a JEM 1200-EX transmission electron microscope operated at 80 keV, and images were recorded digitally with a wide-angle (top mount) BioScan 600-W 1Kx1K pixel digital camera.

RT-PCR

Virion RNA from wild type and CP variants was subjected to Poly-A tailing using *E. coli* poly A polymerase according to the manufacturer's protocol (NEB). Following inactivation of poly A polymerase by heat (95°C for 2 to 3'minutes), first strand cDNA was synthesized using a reverse primer (5' GGGAGGACACAGCCAACATACGTATTTTTTTTTTT TTTTTTTTTT 3'; *Sna*B I site is underlined) and M-MulV reverse transcriptase (New England Biolabs). The resulting product was subjected to PCR using a forward primer (5' TACGTAAACTTACCAATCAAAAG 3') and a reverse primer (5'GGGAGGACACAGC CAACATACGTA3'; SnaB I site is underlined). PCR products were analyzed by agarose gel (1%) electrophoresis then visualized and photographed under UV light following ethidium bromide staining.

Co-immunoprecipitation assay (Co-IP)

Total protein extracts were prepared from healthy and agroinfiltrated N. *benthamiana* leaves as described previously [148]. Isolated proteins were incubated with anti-126 kDa protein antibody at 1:100 dilution (kindly provided by Rick Nelson) for 8 h at 4 °C. Then, a 30-μl aliquot of protein G-agarose beads (Santa Cruz Biotechnology, USA) was added to each tube, followed by incubation for 2 h at room temperature. The immune complexes were then precipitated by centrifugation for 1 min at 10,000 × g and washed three times in 1ml of phosphate-buffered saline (0.1M NaCl, 90 mM sodium phosphate [pH 7.0]). For RNase A treatment, the precipitated proteins were treated with RNase A (50 μg/ml) for 2 h at 25 °C [144]. The resulting precipitated proteins were eluted from the beads by boiling

in SDS-PAGE sample buffer for 3 min. Equal volumes of protein samples were analyzed by SDS-PAGE, followed by immuno blot analysis with anti-126-kDa and anti-STMV CP antibodies.

In-vitro translation of CP

N. benthamiana leaves were co-infiltrated with agrotransformants of STMV wt or CP variants 3A or 5A or 7A and TBSV P19, a suppressor of RNA silencing [149]. Total RNA isolated from each of these infiltrated leaves was used to produce their respective CPs using a wheat germ in vitro translation kit (Promega Corporation) and subjected to Western, Northwestern and Far-Northwestern blot analyses.

Western, Northwestern and Far-Northwestern analyses

Western blot analysis was performed as described previously [129]. For Northwestern analysis [134], wild type or desired variant CPs of STMV were resolved by 10% or 12% SDS-PAGE and transferred to a nitrocellulose membrane. Proteins immobilized on the nitrocellulose membrane were re-natured overnight in a buffer containing 15'mM HEPES (pH 8.0), 10 mM KCl, 10% glycerol, and 1 mM dithiothreitol at 4 °C. Membranes were then hybridized with ^{32}P-STMV RNA for 1 h at room temperature in re-naturing buffer containing 2 mg/ml of yeast tRNA. The membrane was washed twice with re-naturing buffer at room temperature to remove any unbound RNA, followed by autoradiography. For Far-Northwestern analysis [134], after performing Northwestern analysis as described above, hybridization of the membrane was performed with a mixture containing ^{32}P-labelled STMV RNA transcripts and *in vitro* translated STMV CP.

ACKNOWLEDGMENTS

We wish thank John Lindbo for the pJL 36 construct, Barbara Baker for the pRP construct and Richard Nelson for anti-126kDa RdRp antibody. This work is supported in part by an RSAP grant from UCR (A.L.N.R and D.M) and the National Science Foundation Grant No. DMR-1310687 (R.Z).

Chapter 6

Role of genome in the formation of conical retroviral shells

6.1 Introduction

Human immunodeficiency virus (HIV) has, not surprisingly, attracted significant attention from the wider scientific community due its association with the AIDS pandemic [150]. The lifecycle of HIV involves the budding of the non-infectious immature virion that subsequently transforms into an infectious, mature virus [11, 12, 151]. Like many other viruses, the mature HIV particle consists of a protein shell that surrounds its genetic materials, enveloped by a bilayer lipid membrane. The shell of the immature virion is roughly spherical and is built up from a large number of multi-domain protein called Gag. When the virus matures, the Gag protein is cleaved into three structural domains called matrix (MA), capsid (CA) and nucleocapsid (NC) [11, 12, 151]. The MA protein remains

attached to the membrane envelope, while the positively charged NC proteins bind to the genome. The CA protein self assembles around the genome/NC complexes to form predominantly conical capsid shells, but cylindrical and irregular shapes have also been observed[3]. The physical processes that control the unique shapes of the mature capsid are not well understood.

Several *in vitro* experiments have shown that CA proteins can assemble spontaneously in solution to form cylindrical and conical shells in the absence of genome[152, 153, 154, 155, 156, 14, 157, 13, 158, 159, 17, 160]. In 1999, Ganser *et al.* [14] used a recombinant CA-NC fusion proteins to test the impact of specific genome sequences on the formation of conical capsids by mixing the CA-NC protein not only with the native HIV genome (a 1400 nucleotide template) but also with other viral and non-viral RNAs. They found that the HIV genome is not specifically required for the formation of conical capsids. In fact, their studies show that under high-salt concentrations both conical and cylindrical shells form, confirming that the genome is not required for the cone formation. Still, according to their experiments, the majority of the capsids that form are cylindrical, in contrast to HIV capsids assembled *in vivo* [14]. We note that in the experiments of Ganser *et al.* RNA interacts with the shell, due to the attractive interaction between the genome and the positive charges on the recombinant CA-NC fusion proteins. On the other hand, in *in vivo* experiments the protease cleaves the link between NC and CA and as such the shell is solely built from CA and the positively charged NC proteins are attracted to RNA and condenses it. It is well-known that there is no attractive interaction between RNA and CA proteins [161, 16].

There are a number of theoretical and computer simulation studies that investigate factors that contribute to the formation of conical capsids. The focus of prior work is on the elastic energy of the capsids, mostly in the absence of genome, and also on role of kinetic pathways towards assembly. In 2005, Nguyen, Gelbart and Bruinsma did the first equilibrium studies of HIV capsids and compared the elastic energy of conical and cylindrical shells [162]. In their work, the presence of the genome was implicitly taken into account by putting constraints on the shells. They found that the conical capsid constitutes an elastic minimum energy structure only under the condition of fixed area, volume and spanning length. The volume constraint mimics the RNA being confined inside the capsid. The spanning length constraint is justified because *in vivo* the height of cylindrical and conical HIV capsids is fixed as a result of the enclosing membrane (the envelope). It is worth mentioning that cylinders and cones with the same area and volume also have the same height, that is, to lowest order in the cone aperture angle.

Despite intense research, including the aforementioned studies, the predominance of conical over cylindrical capsids remains poorly understood. In particular, no theoretical studies have explicitly taken into account the impact of the genome on the formation of retroviral capsids. In this paper, using a simple mean-field model, we obtain the free energy of confinement of a chain confined in conical and cylindrical capsids. Due to the absence of sufficient information on the CA-CA interaction and the free energies associated with it, we focus entirely on role of genome-shell interaction in the formation of capsids. We obtain the concentration profile of the model genome, a linear polymer, inside the capsid and consider the cases in which the genome interacts attractively with the capsid mirroring

the experiments of Ganser *et al.*, and also consider *in vivo* conditions in which the genome presumably does not interact with the capsid wall built from CA proteins.

We find for a genome that does not interact attractively with the shell, the free energy cost of confining this genome in a conical capsid is smaller than that associated with confining it in a cylindrical one, presuming they have the same height and area mimicking the impact of the envelope and fixing the number of Gag proteins that make up the shell. In conical capsids the genome is primarily located in the base of the cone, consistent with the experimental findings of Refs. [161, 16, 163, 164]. However, if the genome interacts attractively with the capsid wall, as is the case in the experiments of Ganser *et al.* [14], we find that with the same amount of genomic material and the same number of capsid proteins, the genome confinement free energy is slightly larger for a cone-shaped than that of a cylindrical capsid. This may explain why cones are less favorable than cylinders in *in vitro* HIV-1 experiments.

In what follows, we first present our mean-field model, which allows us to calculate the free energy of a linear genome in both cylindrical and conical capsids. We then numerically obtain the free energy of confinement of genome in the absence of genome-capsid attractive interaction but with excluded volume interaction. Next, we solve free energy of confinement in cylindrical and conical shells analytically in the absence of excluded-volume interactions. Finally, we calculate the free energy of a genome confined in a capsid that interacts attractively with the capsid. We summarize our work and present our main conclusions in the last section.

6.2 Model

Mean-field theory has been extensively used to obtain the free energy of chains confined in spherical shells[71, 165, 62, 166, 98, 63, 61, 50, 27, 73, 117, 167, 168]. Since the focus of this paper is on HIV particles, we employ Edwards-de Gennes-Lifshitz theory to calculate the free energy of a linear polymer confined in a cylindrical and conical shell [169, 170, 60]. In the ground-state approximation of long chains, it reads

$$\beta\Delta F = \int dV \left(\frac{a^2}{6}(\nabla\psi)^2 + \frac{1}{2}\upsilon\psi^4 \right) \tag{6.1}$$

where a is the Kuhn length, υ is the excluded volume and $\beta = 1/k_B T$ the reciprocal thermal energy with T the temperature and k_B the Boltzmann constant. The quantity ψ indicates the polymer density field and ψ^2 represents the monomer density at positions r, the radial distance from the center of the capsid, and h, the distance along the height of the capsid. To obtain the genome profile inside the shell, we will make contour plots of ψ^2 as a function of r and h. The gradient term in Eq. 6.1 is equal to zero for a constant ψ and thus it is associated with the entropic cost due to the non-uniform chain distribution. The second term is the energy penalty related to the excluded volume interaction. Hence, our reference free energy is that of a uniformly distributed polymer. In our closed system the number of encapsulated monomers, N, is fixed implying that

$$N = \int dV \psi^2 = constant. \tag{6.2}$$

Minimizing the free energy given in Eq. 6.1 with respect to the field ψ, subject to the above constraint, gives the following Euler-Lagrange differential equation for the density

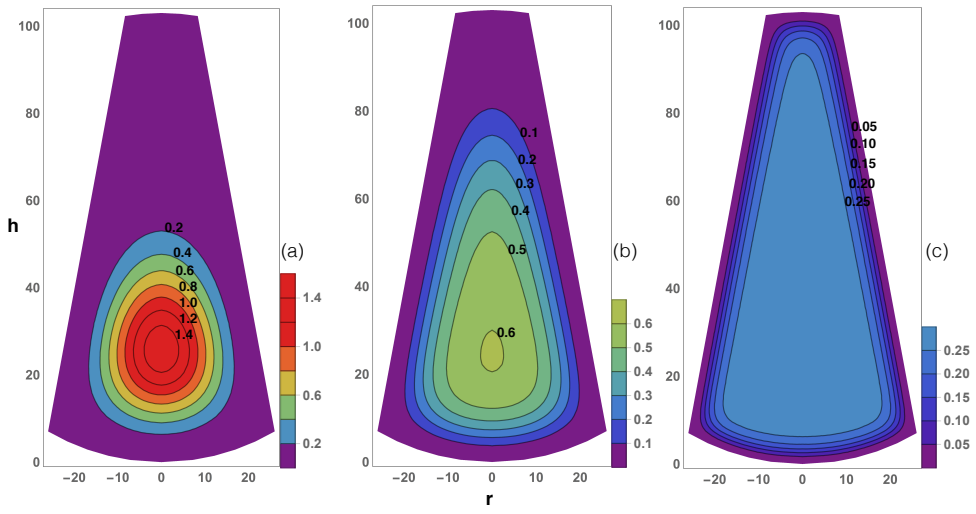


Figure 6.1: Contour plot of the genome density profile (ψ^2 in units of $1/nm^3$) for excluded volume $\nu = 0$ (a), $\nu = 0.01\ nm^3$ (b) and $\nu = 0.1\ nm^3$ (c). The total number of monomers is $N = 18000$. The other parameters are $h = 103\ nm$, $R_b = 26\ nm$, $\alpha = 21^\circ$ and $a = 1\ nm$. As the excluded volume increases, the polymer spreads more evenly over the cavity.

profile inside the capsid

$$\frac{a^2}{6}\nabla^2\psi = -\lambda\psi + \upsilon\psi^3, \tag{6.3}$$

where λ is a Langrange multiplier that will be fixed by the condition of the conservation of mass.

Considering there is no known interaction between the genome and capsid proteins (CA) in HIV-1 virions, we employ the Dirichlet boundary condition in which the density field at the surface $\psi\big|_S = 0$ [169].

To obtain the density field ψ for a polymer trapped in conical and cylindrical geometries, we numerically find the solutions of the non-linear differential equation, Eq. 6.3, using a 2D Finite Element Method [171]. More specifically, for a given monomer number, N, we vary the Lagrange multiplier λ and self-consistently solve Eqs. 6.2 and 6.3 till we

obtain the total number of monomers equal to the one reported in the experiments and thus find the corresponding ψ. Inserting the density field ψ back into Eq. 6.1, we can calculate the relevant free energies as a function of the different system parameters such as the length of the polymer and the dimensions and shape of the cavity, see Fig. 6.2.

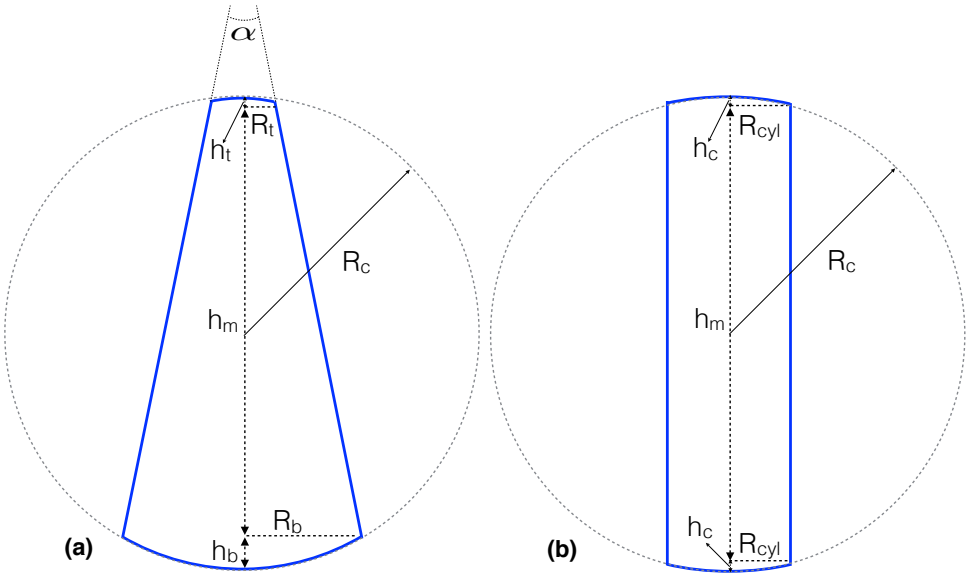


Figure 6.2: A conical capsid (blue solid lines) inside a membrane (gray dashed sphere). It consists of the lateral surface of a truncated cone, along with two spherical caps. The spherical caps both follow the surrounding membranes, and have radii of R_c. The top and bottom radii of the truncated cone are given as R_t and R_b, respectively, and the cone angle is given by α. The perpendicular height of the cone is h ($h = 2R_c$), whereas the top and bottom caps have heights of h_t and h_b, respectively (a). The cylindrical capsid used in the calculations shares the same geometry as the conical capsid with the conditions $R_{cyl} = R_t = R_b$ and $h_t = h_b = h_c$ (b).

The results of our numerical calculations for the conical core are presented as a contour plot of the polymer segment density in Fig. 6.1 with (a) a zero and (b, c) a non-zero excluded volume parameter, v. We set the height of the cone at $h = 103\ nm$, the radius of the base at $R_b = 26\ nm$ and presumed a cone angle of $\alpha = 21°$, consistent with

observations on HIV [14, 172, 164, 3]. See also Fig. 6.2 (a). The quantities $h_b = 7.05\ nm$ and $h_t = 0.68\ nm$ can then be easily obtained as a function of h, R_b and α. The Kuhn length a is $1\ nm$ in all our calculations [83], and the total number of monomers is chosen as $N = 18000$, which is the approximate number of nucleotides in two copies of the HIV genome carried in the capsid [173]. The genome profile in Fig. 6.1(a) with $\nu = 0$ is condensed towards the base of the cone similar to the genome profile in conical cores observed in experiments [161, 16, 163, 164]. Note that due to the electrostatic interaction between the positively charged cleaved NC proteins and the negatively charged genome the solution can in some sense be considered near the θ conditions. For simplicity, thus we set the excluded volume interaction to zero, $\nu = 0$. If $\nu > 0$, the chain is not as condensed at the base but is distributed more uniformly along the cone as can be seen in Fig. 6.1 (b) and (c). As illustrated in the figure, the chain locates further from the base of the cone and distributes more uniformly along the cone as ν increases.

We now compare the free energy of a genome confined in a cylindrical shell with that in a conical one as explained above (Fig. 6.1), assuming that the number of subunits are the same in both cylindrical and conical structures, and that the height of the cylinders and cones are also the same. This is reasonable because the height of the core of mature HIV-1 virions is defined by the presence of an enclosing membrane, as already advertised. Under these constraints, the radius of the cylinder becomes $R_{cyl} = 17.9\ nm$ and the height of the spherical caps $h_c = 3.22\ nm$, see Fig. 6.2 (b). Figure 6.3(a) and (b, c) show the density profiles of the genome in a cylindrical core with and without monomer-monomer excluded-volume interactions, respectively.

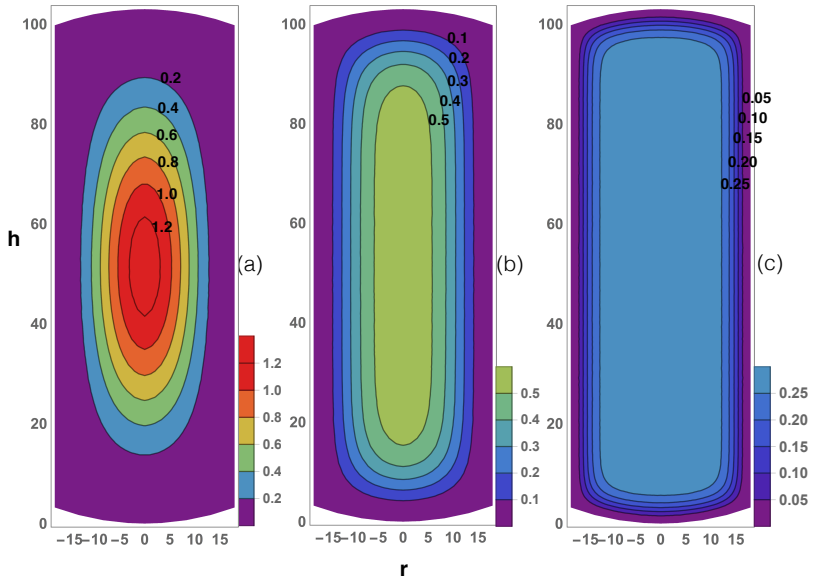


Figure 6.3: Contour plot of the genome density profile (ψ^2 in units of $1/nm^3$) for excluded volume $\nu = 0$ (a), $\nu = 0.01\ nm^3$ (b) and $\nu = 0.1\ nm^3$ (c). The total number of monomers is $N = 18000$. Other parameters are $h = 103\ nm$, $R_{cyl} = 17.9\ nm$ and $a = 1\ nm$. As the excluded volume increases, polymer spreads more evenly over the volume of the cavity.

Table 6.1 summarizes the free energy values in units of k_BT with and without excluded-volume interaction for both geometries. We find that for the same number of monomers and the same surface area, the latter imposing equal numbers of capsid proteins in the shells, the genome confinement free energy for the conical cavity is lower than that for the cylindrical cavity. As presented in the table, the difference is rather substantial. In part, this lowering of the confinement free energy is due to the lower overall density of segments in the cone, the volume of which is larger than a cylinder with the same surface area. We note here that while for simplicity we employed linear chains for all the calculations in this paper, we also studied the impact of branching (following the techniques used in Ref. [73]) on the final results. Including the secondary structures of RNA makes the chain more

Table 6.1: Confinement free energy for cylindrical and cone-shaped cavities. The total number of monomers is $N = 18000$. Other parameters take the values $h = 103\ nm$, $R_b = 26\ nm$, $\alpha = 21°$, $R_{cyl} = 17.9\ nm$ and $a = 1\ nm$.

Δ F (k_BT)		
	Conical	Cylindrical
$\nu = 0$	45	57
$\nu = 0.01\ nm^3$	85	93
$\nu = 0.1\ nm^3$	301	314

compact and lower the free energy values for both cylindrical and conical shells, but do not change the conclusion of the paper.

6.3 Gaussian chain

We now compare our numerical findings, presented in the previous section, with some analytical results, which we were able to find in certain limits. Indeed, it is possible to solve the Euler Lagrange equation (Eq. 6.3) to obtain analytical predictions for the polymer free energy and the polymer concentration profile in cylindrical and conical shells, at least if we treat the polymer as a Gaussian chain by setting $\upsilon = 0$ and by assuming that the polymer does not interact with the capsid other than that it acts as a confining wall. Equation 6.3 can be solved exactly for a purely cylindrical shell, so without the caps, and for a conical structure that has a sharp tip. Both model capsids are illustrated in Fig. 6.4.

The fact that the cone of Fig. 6.4 does not quite look like the truncated cone of HIV capsids is not problematic, because we found numerically in Fig. 6.1(a) that the density of the genome for $\upsilon = 0$ to be concentrated in the bottom part of the shell. Hence, the

exact shape of the tip should be irrelevant to our calculations, and this is confirmed below. Similarly, the exact shape of the cylindrical cap does not effect our calculations.

For a cylindrical shell we obviously use cylindrical coordinate $\mathbf{x} = \{\rho, \phi, z\}$. For a cylinder of height h and radius R_{cyl} (Fig. 6.4(a)), the Dirichlet boundary conditions are

$$\psi(\rho, \phi, z = 0) = \psi(\rho, \phi, z = h) = 0, \tag{6.4}$$

$$\psi(\rho = R_{cyl}, \phi, z) = 0. \tag{6.5}$$

Setting $6/a^2\lambda = \mu$, the eigenfunctions and eigenvalues of Eq. 6.3 with $v = 0$ can be found analytically. Expressed in the "quantum" number n for the z direction, m for the ϕ direction and q for the ρ direction, we find the eigenvalues and eigenfunctions

$$\mu_{n,m,q} = \frac{\pi^2}{h^2} n^2 + \frac{\lambda_{m,q}^2}{R_{cyl}^2}, \tag{6.6}$$

and

$$\psi(\mathbf{x}) = \sum_{n,m,q} \alpha_{n,m,q} J_m\left(\frac{\lambda_{m,q}\rho}{R_{cyl}}\right) \sin\left(\frac{\pi}{h} n z\right) e^{im\phi}, \tag{6.7}$$

respectively. Here, $\alpha_{n,m,q}$ is the normalization factor which can be found from the constraint, Eq. 6.2, and $\lambda_{m,q}$ is the q^{th} zero of the m^{th} Bessel function of the first kind J_m.

By inserting these expressions in Eq. 6.1, we calculate the free energy in the ground state, for long chains, so $\sqrt{N}a \gg R_{cyl}$, and we can keep only the first term in the q sum. In calculating the free energy we see that all the terms besides $m = 0$ integrate to zero. We also assume that the height of cylinder is very large implying $\sqrt{N}a \ll h$, and thus we can ignore the n^2 term in the eigenvalues. We find the free energy of a gaussian chain confined in a cylinder obeys

$$\beta\mathcal{F}_{cyl} \approx \frac{Na^2}{6} \frac{\lambda_{01}^2}{R_{cyl}^2}, \tag{6.8}$$

with $\lambda_{01} \approx 2.4$. Even if $\sqrt{N}a \gg h$, this expression holds, at least if $h \gg R_{cyl}$. The genome density profile inside the cylinder can be obtained from Eq. 6.7 and a plot of the density profile is presented in Fig. 6.5(a).

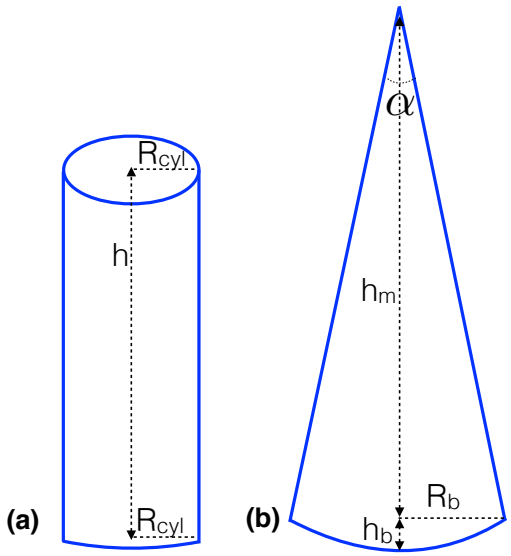


Figure 6.4: Geometries used for the analytical calculations involving the confinement of a gaussian chain into a cylinder (a) and a cone with a sharp tip and rounded bottom (b). R_{cyl} is the radius of the cylinder and R_b is the bottom radius of the cone with opening angle α. The perpendicular height of the cone is $h = h_m + h_b$ whereas the bottom cap has a height of h_b. The perpendicular height of the cylinder is h.

For a polymer confined in a conical shell with a cut sphere base, depicted in Fig. 6.4(b), it makes sense to use spherical coordinates $\mathbf{x} = \{r, \theta, \phi\}$. If α is the opening angle of the cone and R_b the radius of the spherical cap, the Dirichlet boundary conditions are

$$\psi(r, \theta = \alpha/2, \phi) = 0, \tag{6.9}$$

$$\psi(r = R_b/\sin\alpha/2, \theta, \phi) = 0. \tag{6.10}$$

The relevant quantum numbers now become q for the r direction, l for the θ direction and m for the ϕ direction. The eigenfunctions are given by

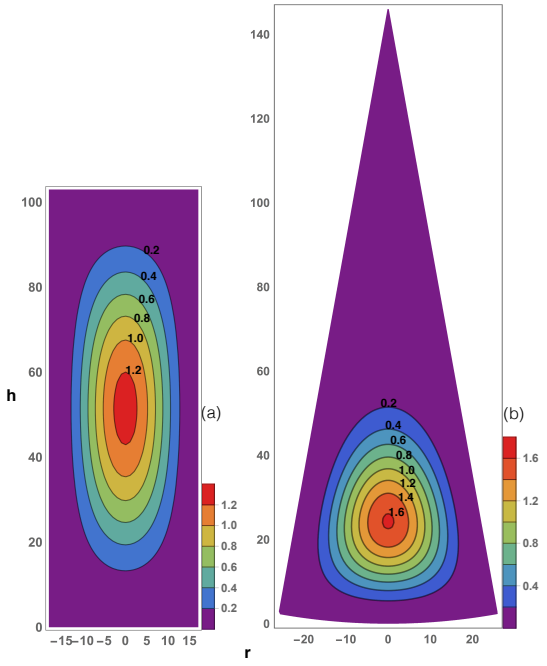


Figure 6.5: Contour plot of the genome density profile (ψ^2 in units of $1/nm^3$) plotted from the eigenfunctions obtained from Eq. 6.7 and Eq. 6.11 for a cylindrical (a) and conical (b) capsid. Other parameters take the values $R_b = 26\ nm$, $\alpha \approx 21°$, $R_{cyl} = 17.9\ nm$, $a = 1\ nm$, $h_b = 2.3\ nm$, $h = 147\ nm$ for the cone and $h = 103\ nm$ for the cylinder.

$$\psi(\mathbf{x}) = \sum_{m,l,q} \alpha_{m,l,q}\, j_{\nu(l,m)}(\sqrt{\mu}_{m,l,q} r) P^m_{\nu(l,m)}(\cos\theta) e^{im\phi} \tag{6.11}$$

where $\alpha_{m,l,q}$ is the normalization factor which can be found from the constraint, Eq. 6.2, $j_{\nu(l,m)}$ is a spherical Bessel function, $P^m_{\nu(l,m)}$ is the associated Legendre polynomial and the eigenvalues are given by

$$\mu_{m,l,q} = \frac{\lambda^2_{\nu(l,m)+\frac{1}{2},q} \sin^2 \alpha/2}{R_b^2} \tag{6.12}$$

with $\nu(l,m)$ the l^{th} zero of the Legendre polynomial, for which

$$P^m_{\nu(l,m)}(\cos\alpha/2) = 0 \tag{6.13}$$

holds, and where $\lambda_{\nu(l,m)+\frac{1}{2},q}$ is the q^{th} zero of the Bessel function of the first kind $J_{\nu(l,m)+\frac{1}{2}}$. Plots of a spherical Bessel function with $m = 0$ also indicate that the concentration of genome is maximum at the axis of the cone, consistent with numerical results for $\upsilon = 0$. Figure 6.5(b) shows the location of the chain inside the conical shell, which is consistent with our numerical results (see Fig. 6.1(a)).

To obtain the free energy of the gaussian chain confined into a conical shell, we insert Eq. 6.11 into Eq. 6.1. In performing the integral, only the m = 0 term survives. Since the theory applies in the ground-state limit, which demands that $\sqrt{N}a \gg R_b/\sin\alpha/2$, we consider only the $l = 1$ and $q = 1$ state. We find the free energy of a gaussian chain confined in a conical shell to obey

$$\beta\mathcal{F}_{cone} \approx \frac{Na^2}{6}\frac{\lambda_{\nu(1,0)+\frac{1}{2},1}\sin^2\alpha/2}{R_b^2}, \tag{6.14}$$

with $\nu(1,0)$ calculated from Eq. 6.13 for a given α to obtain the value of the Bessel function zero.

In order to estimate the difference between the free energies of a gaussian chain trapped in a cylinder with respect to that trapped in a cone, we set all the relevant parameters to the values used in the numerical calculations of previous sections, *i.e.*, the total length of the genome $N = 18000$, cone angle $\alpha \approx 21^\circ$ and the radius of the base of the cone $R_b = 26\ nm$, typical for HIV particles [14, 172, 164, 3]. From this we obtain $h = 147\ nm$ and the height of the spherical cap $h_b = 2.3\ nm$. We note that even though we set the

height of the truncated cone in the previous sections to $h = 103$, we expect that for the Gaussian chain the precise form and height of the shell are not hugely important as the genome is condensed at the bottom of the shell. For the cylindrical capsid we set radius of the capsid at $R_{cyl} = 17.9\ nm$ and its height at $h = 103\ nm$ as we did in our numerical calculations.

Based on the analytical calculation, we find that the free energy of a chain confined in the cylindrical capsid is equal to 54 k_BT and that in the conical capsid is equal to 47 k_BT. These values are very close to the confinement free energy for the cylinder of 57 k_BT and for the cone of 45 k_BT that we obtained numerically. Once again, this could explain the predominance of conical shells compared to cylindrical ones as observed in *in vivo* experiments.

6.4 Capsid-genome Interaction

As noted in the introduction, using the recombinant CA-NC fusion proteins and HIV-1 RNA as well as other types of RNAs, Ganser *et al.* obtained a mixture of cones and cylinders in their *in vitro* self-assembly studies [14]. We now include attractive interaction between NC proteins in the shell and RNA in the calculations. The free energy becomes [60, 71, 117]

$$\beta\Delta F_{int} = -\gamma\beta a^3 \int dS\psi^2 + \beta\Delta F, \tag{6.15}$$

where γ is the interaction energy between the genome and the inner surface of the CA-NC complex per unit area and ΔF is the polymer confinement free energy, see Eq. 6.1. Minimizing Eq. 6.15 with respect to the field ψ, subject to the constraint given in Eq. 6.2,

produces the same Euler-Lagrange differential equation as given in Eq. 6.3, but subject to the following boundary conditions obtained from the minimization of Eq. 6.15 with respect to the the field ψ on the surface,

$$\left(\hat{n} \cdot \nabla\psi = \kappa\psi\right)\Big|_S \tag{6.16}$$

with $\kappa^{-1} = 1/6a\beta\gamma$ a length representing the strength of interaction of the capsid wall with the monomers. If the length κ^{-1} is larger than the Edwards correlation length ($\zeta_E = a/\sqrt{3\nu\phi_0}$) with $\phi_0 = N/V$ the monomer number density, the excluded volume interaction between the monomers overcomes the monomer attraction to the wall and we are in the weak adsorption limit. Conversely, in the strong adsorption limit, κ^{-1} is smaller than the correlation length ζ_E [60].

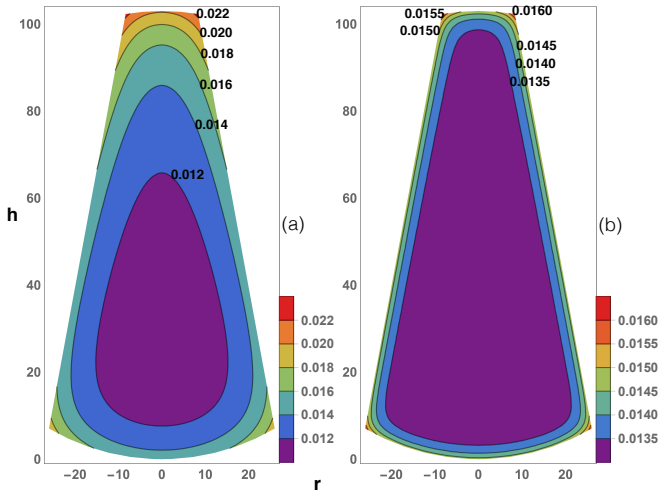


Figure 6.6: Contour plot of the genome density profile (ψ^2 in units of $1/nm^3$) confined into a conical capsid with an attractive inner wall for excluded volume $\nu = 0.1$ (a) and $\nu = 1.0\ nm^3$ (b). The total number of monomers is $N = 1400$. Other parameters are $h = 103\ nm$, $R_b = 26\ nm$, $\alpha = 21°$, $a = 1\ nm$ and $\kappa = 10^{-4} nm^{-1}$. The polymer profile is denser at the corners of the capsids, having the maximum density at the tip of the cone.

For the weak adsorption regime, corresponding in practice to a high salt concentration mimicking the experiments of Ganser *et al.* [14], κ^{-1} becomes larger than the Edwards correlation length. Figures 6.6 and 6.7 show the genome density profiles in the conical and cylindrical geometries, respectively. For all plots the total number of monomers is fixed at $N = 1400$, which corresponds to the HIV-1 RNA template used in the experiments of Ganser *et al.* [14]. The figures correspond to the weak adsorption regime with $\kappa^{-1} \gg \zeta_E$.

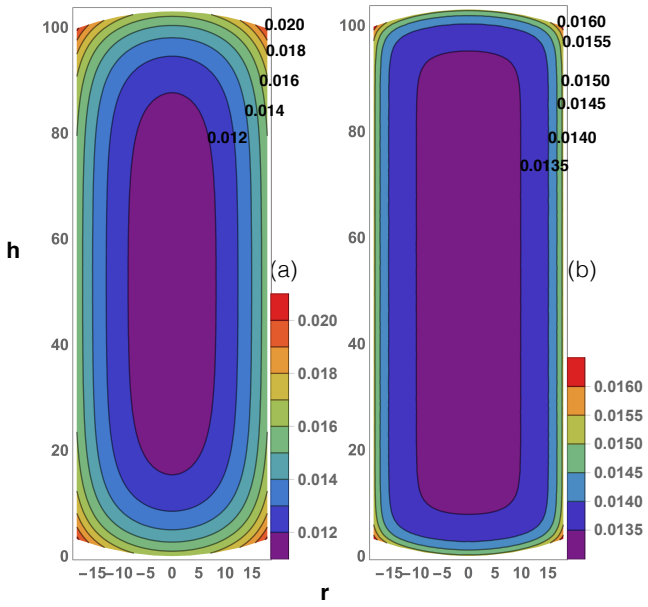


Figure 6.7: Contour plot of the genome density profile (ψ^2 in units of $1/nm^3$) confined into a cylindrical capsid with an attractive inner wall for excluded volume $\nu = 0.1$ (a) and $\nu = 1.0\ nm^3$ (b). The total number of monomers is $N = 1400$. Other parameters are $h = 103\ nm$, $R_b = 26\ nm$, $\alpha = 21°$, $R_{cyl} = 17.9\ nm$, $a = 1\ nm$ and $\kappa = 10^{-4} nm^{-1}$. The polymer profile is denser at the corners of the capsids, having the maximum density at the corners of the cylinder.

As illustrated in the contour plots for the polymer density profile in Figs. 6.6 and 6.7, the genome covers the wall completely, and the density is higher near the rims of the cones and cylinders. This is not surprising, of course. Unlike the case studied in Sect.

Model, the free energy of the genome confined in a cylindrical core, $\Delta F_{cyl} = -0.79\ k_BT$, is more negative than that of the conical core, $\Delta F_{cone} = -0.77\ k_BT$, albeit by only a rather small amount for the excluded volume $\nu = 0.1$. Since in the experiments of Ganser *et al.*, NC proteins are in the capsid wall and do not condense the genome, we expect a higher excluded volume interaction. For an excluded volume $\nu = 1.0$, the difference between the free energies increase, *e.g.* $\Delta F_{cyl} = -6.62\ k_BT$ and $\Delta F_{cone} = -6.43\ k_BT$. This would suggest, if the binding energies of the proteins are the same in both types of capsid, and kinetic effects are unimportant, then a slightly larger number of cylinders forms compared to cones, based on our calculations. Interestingly, the ratio of cylinders to cones formed in the *in vitro* experiments of Ganser *et al.* was 3/2, whereas the ratio of cylinders to cones found in *in vivo* studies is very small.

6.5 Conclusions

As noted in the introduction, CA proteins assemble spontaneously *in vitro* to form tubular arrays or conical structures with a geometry similar to that of the mature HIV shells, even in the absence of genome [14, 17, 160]. However, Ganser *et al.* showed that recombinant CA-NC proteins can assemble *in vitro* around any RNA to form a mixture of conical and cylindrical capsids but most structures are tubular in contrast to *in vivo* structures in which most capsids form conical structures. More recently, using Electron Cryotomography, Woodward *et al.* monitored the maturation intermediates of HIV particles *in vivo* and found that most cylindrical capsids do not encapsulate RNA [16]. In fact, they found a condensed form of genome sitting next to but outside the cylindrical shells. These

experiments suggest that the presence of encapsulated genome could promote the formation of conical capsids.

Several theoretical and numerical studies investigated the formation of conical structures in the absence of genome and of membrane. It seems that in the absence of genome, the formation of a conical shape could be the result of irreversible steps in the growth of an elastic sheet, and connected to the dynamics of formation of pentamers during the growth process [174, 175, 176, 177]. However, the focus of this paper is solely on the impact of genome on the assembly of conical and cylindrical capsids.

Using a simple mean-field theory, we explicitly calculated the encapsulation free energy assembly of both cylindrical and conical structures. It is well-known that the interaction between the positively charged NC domain and negatively charged RNA is responsible for the encapsulation of genome in the immature HIV virus and that the CA lattice of the mature HIV does not interact with RNA. The fact that RNA stays inside the capsid despite the absence of interaction with the CA proteins is not yet well-understood. It has been suggested that while the viral protease cleaves the CA-NC link during maturation, incomplete cleaved links could be the reason why the RNA remains encapsulated in the mature capsid [161, 16].

It is important to note that while the interaction between CA proteins is the driving force for Gag assembly in both the mature and immature hexagonal lattices, the free energy associated with the CA-CA interaction is thought to be weak, and the free energy of the CA-CA interaction in conical and cylindrical capsids has not yet been determined experimentally. To this end, in this paper, we only focus on the contribution of genome

confinement free energy, assuming that the free energies due to CA-CA interaction between cylindrical and conical capsids are not considerable as they both appear in the *in vitro* studies in which no genome was present.

One of the important results of this paper is that while the free energy due to confinement of genome is lower for conical capsid than a cylindrical one in *in vivo* experiments, where interactions between the genome and the capsid are believed to be negligible, the opposite is true if there is interaction with the wall. We emphasize that to obtain these results we employed the parameters associated with the height, radius and cone angle given in Ref. [172]. Since these parameters are not exactly the same as reported by different groups, we checked the robustness of our results using the range of parameters obtained in different experiments. We find that our results do not depend on the exact parameters reported by one group and are robust [14, 172, 164, 3]. Furthermore, the HIV conical capsids occasionally have larger apex angles than those studied above. To this end, we compare the free energy of the encapsidated genome by a cylinder vs. a cone with a larger apex angle than 21° while keeping the area of both structures the same. Our findings turn out not to qualitatively depend on the apex angle and our conclusions remain the same.

Understanding the factors that contribute to the formation of conical capsids can play an important role in the development of anti-viral drugs and nano-containers for gene therapy.

Acknowledgments

The authors thank William Gelbart for many useful and productive discussions over many years on the physics of virus assembly, including the work presented in this paper. This work was supported by the National Science Foundation through Grant No. DMR-13-10687 (R.Z.).

Chapter 7

In vitro Protease Cleavage and Computer Simulations Reveal the HIV-1 Capsid Maturation Pathway

7.1 Introduction

Formation of the infectious Human Immunodeficiency Virus (HIV-1) particle occurs via two processes: the assembly of spherical immature particles that are non-infectious, as the virus buds out at the plasma membrane, followed by maturation of the viral core [12]. During maturation, the viral protease (PR) cleaves the Gag polyprotein into its constituents: matrix (MA), capsid (CA), nucleocapsid (NC), and p6, thereby also releasing the SP1 and SP2 peptides [178]. The interaction between the positively charged NC domain and negatively charged RNA [179], in particular the 5' UTR, is responsible for the en-

capsidation of the RNA genome within particles. Protein-protein interactions between CA domains are the driving force for Gag assembly in the immature hexagonal lattice [180, 181], as well as for CA assembly in the mature capsid [158, 182, 160]. Previous computer simulations have revealed key features of CA self-assembly into conical mature HIV-1 capsid [160, 183, 184, 185, 186]. HIV-1 maturation occurs in multiple stages [187]. Following the first cleavage between SP1 and NC, the NC-RNA complex condenses into a dense material. Subsequent cleavage at the MA-CA junction liberates MA and frees CA-SP1 from membrane attachment. The slowest cleavage is the release of SP1 from the C-terminus of CA [187, 188, 189]. Proteolytic maturation is essential for infectivity, and PR inhibitors are a key element of current antiretroviral therapies [190]. A potent maturation inhibitor, Bevirimat (BVM), blocks CA-SP1 cleavage and prevents formation of the mature conical capsid [191, 192, 193, 194]. Structural analyses of the Gag lattice in mutant viruses that have impaired cleavage of Gag at specific sites suggest that processing is ordered and that the RNA/protein complex (RNP) may maintain a link with the remaining Gag lattice after cleavage [195].

While the architectures of immature and mature virions are well characterized [181, 160, 196, 197], the pathway of maturation and the morphological transition is not well understood. Recent studies have led to two distinct, competing models for the transformation of immature spherical virion to mature virions with conical cores, namely the disassembly/reassembly model and the displacive model [180, 195, 19, 18, 17, 16]. In the disassembly/reassembly model, the immature lattice disassembles following PR cleavage, generating a pool of soluble CA molecules from which a mature capsid assembles *de novo*.

Several previous studies provide support for this model. Firstly, based on early cryo-electron tomography (cryoET) observations of 1) virus-like particles (VLPs) with multiple capsids, 2) mature capsids with tip closing defects [3], and 3) a correlation between the capsid length and the membrane diameter [196], models of capsid growth, from either the "narrow end" [3] or the "wide end" [196] of the capsid, have been proposed. Secondly, structural studies revealed distinct lattice spacings and different interaction interfaces between the immature and the mature CA lattices, suggesting the transition would involve a complete disassembly of the immature lattice [180, 198]. More recently, cryoET of maturation inhibitor BVM-treated virions displayed a shell that resembles the CA layer of the immature Gag shell but is less complete [193]. Lastly, cryoET of budded viral particles comprising immature, maturation-intermediate, and mature core morphologies suggested that the core assembly pathway involves the formation of a CA sheet that associates with the condensed RNP complex and further polymerizes to produce the mature conical core [16].

By contrast, the displacive transformation model postulates that maturation involves a direct, non-diffusional remodeling of the immature Gag lattice to the mature capsid lattice [19, 17]. In this model, the immature lattice does not disassemble to form the mature one. Evidence for a displacive model includes: 1) PR cleavage of *in vitro* assembled CA-SP1-NC tubular assemblies resulted in conversion to mature CA tubes without disassembly [17]; 2) mutant particles with a cleavage defect at the CA-SP1 site have thin-walled spheroidal shells with lattices displacively transformed into a mature-like lattices [18]; and 3) recent cryoET observation of multiple, normal-sized cores within a large membrane en-

closure, which argues against *de novo* nucleation and assembly model and suggests a rolling sheet process for CA lattice transformation to conical capsid [19].

To examine the sequence of structural changes that take place during maturation, we established novel *in vitro* cleavage systems that mimic the maturation process, by digesting purified PR-deficient virions and *in vitro* assembled Gag VLPs using recombinant HIV-1 PR. We further investigated the effects of BVM and Gag cleavage mutants on the *in vitro* maturation process. Using computer simulation, we revealed the impact of membrane and genome on mature capsid formation. Integrating our biochemical and structural findings from the *in vitro* maturation experiments with computer modeling and simulations, we conclude that the HIV-1 maturation pathway is neither simply displacive nor exclusively *de novo* reassembly, but a sequential combination of both displacive and disassembly/reassembly processes.

7.2 Results

In vitro maturation by HIV-1 PR cleavage

To study the structural transitions occurring during HIV-1 maturation, we first established a novel *in vitro* system to mimic the PR-driven HIV-1 maturation process by digesting purified PR-deficient immature virions with recombinant HIV-1 PR. The viral membranes of immature particles were permeabilized with Triton X-100 to allow recombinant HIV-1 PR access to the viral structural polyproteins. Processing of $Pr55^{gag}$ by PR was accurate and efficient in the reaction (Fig. 7.1). Gel analysis of the PR-treated particles revealed release of MA, CA, and NC in a time-dependent manner, consistent with the

process seen in native virion maturation [187]. The PR cleavage was also pH dependent, most efficient at pH 6.0, at which almost all polyproteins were cleaved into CA in 2 hrs at $37°C$ (Fig. 7.1 A), whereas the majority of CA-SP1 remained intact at pH 7.5. Furthermore, cleavage at the CA-SP1 junction was inhibited by the maturation inhibitor BVM (Fig. 7.1 B), indicating that the immature particles are of the correct structure for binding the inhibitor [199]. Moreover, the MA-CA cleavage site was apparently partially protected by BVM (Fig. 7.1 B), an effect that was not previously observed during HIV-1 maturation [200, 201]. The overall efficient and accurate cleavage of $Pr55^{gag}$ by PR, observed in this system, suggested that this system could prove useful for mimicking the process of HIV-1 maturation *in vitro*.

In parallel, we established a unique system for *in vitro* maturation of assembled Gag VLPs, using entirely purified recombinant protein components. Immature VLPs were assembled from recombinant Gag protein lacking p6 and most of the MA regions ($\Delta MA_{15-100}p6$) together with RNA. We tested the cleavage efficiency with various HIV-1 PR concentrations (Fig. 7.1 C). Under the experimental conditions at pH 6.0, the cleavage process was completed within 2 hrs in the presence of 3.3 μM PR (Fig. 7.1 C), agreeing well with the results obtained from immature virions (Fig. 7.1 A). As with the immature virions, the process was also time-dependent (Fig. 7.1 D); cleavage between CA-SP1 was slow, and all CA was released within 2 hrs in the presence of 3.3 μM PR. Intriguingly, the effect of BVM was pH-dependent, with a strong effect at pH 7.4 but a marginal effect at pH 6.0 (Fig. 1E). The CA-SP1 cleavage process was nearly complete at pH 6.0, whereas the majority of CA-SP1 remained intact at pH 7.4, despite the presence of a high

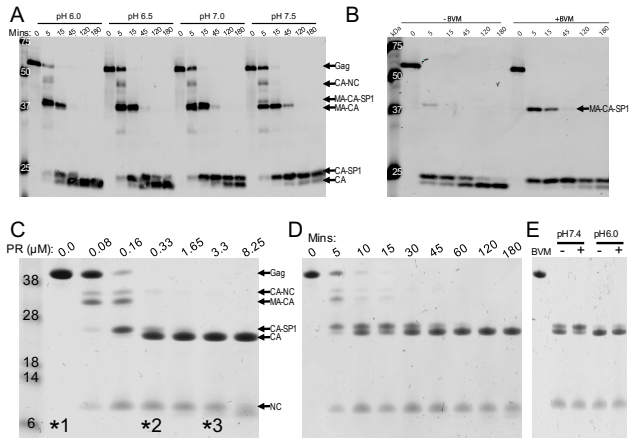


Figure 7.1: *In vitro* maturation of HIV-1 by protease (PR) cleavage. (A&B) HIV-1 PR treatment (2μM recombinant PR) of PR-inactive HIV-1 immature virions at different pHs and for different amounts of time (minutes), up to 2 hrs, at 37°C (A) and in the presence or absence of 2μM bevirimat (BVM) at pH 7.4, for up to 2 hrs, at 37°C (B). Protein products were immunoblotted with a CA-specific antibody and labeled accordingly. (C-E) Cleavage of in vitro assembled Gag VLPs at pH 6.0, at different PR concentrations for 2 hrs (C), different times (minutes) of incubation with 3.3 μM PR (D), and in the presence or absence of 40μM BVM at pH 7.4 and pH 6.0 with 3.3 μM PR for 2 hrs (E). Protein products were visualized by Coomassie blue staining and labeled accordingly. Molecular weight markers were labeled on the left. Samples labeled with "*" were subject to cryoEM analysis, shown in Fig. 2.

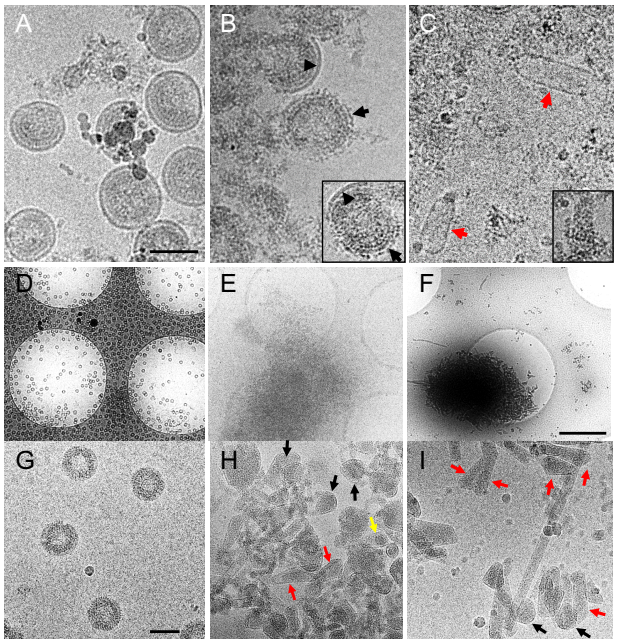


Figure 7.2: CryoEM analysis of HIV *in vitro* maturation by PR cleavage. (A-C) CryoEM images of PR-deficient HIV-1 immature virions before Triton X-100 treatment (A), after Triton X-100 treatment without HIV-1 PR (B), and after Triton X-100 treatment but with PR at 2μM, pH 6 for 2hrs (C). Intact and perforated virial membranes are labeled with black arrowheads and black arrows in B, respectively. The inset in (B) shows one virion in which part of the membrane was still intact and part was punctured. Ordered lattice appears in a conical shaped structure (red arrow). (D-I) CryoEM images of *in vitro* maturation of Gag VLPs. Low (D-F) and high (G-I) magnification cryoEM images show Gag VLPs morphology changes from spheres (D&G) to cones and tubes after 0.33μM (E&H) or 3.3μM (F&I) PR treatment for 2 hours. Scale bars, 100 nm in A, C and G, 1 μm in D.

concentration of BVM (40 μM). The same pH-dependent BVM activity pattern was also observed with immature virions. It should be noted that the efficiency of CA-SP1 cleavage was also reduced at pH 7.4 compared to pH 6.0 (Fig. 7.1 E), as seen with immature virions (Fig. 7.1 A). Taken together, these data indicate that both *in vitro* PR cleavage systems recapitulate the HIV-1 maturation process, providing useful tools for characterizing effects of maturation inhibitors.

CryoEM analysis of *in vitro* maturation process

To investigate the structural changes upon HIV-1 PR cleavage, we carried out cryoEM analysis of the *in vitro* maturation process described above. The PR-deficient immature virions appeared as membrane-bound spheres of variable sizes, about 100 nm in diameter (Fig. 7.2 A). The Gag lattices, lining beneath the membrane, were clearly distinguishable. Limited detergent treatment of immature virions resulted in nonhomogeneous membrane perforation of immature virions (Fig. 7.2 B). Within a single virion, part of the viral membrane became punctured (Fig. 7.2 B and inset, black arrow), while other areas of viral membrane remained intact (Fig. 7.2 B and inset, black arrowhead). In fact, most of immature particles were broken or fragmented upon mild detergent treatment (Fig. 7.2 B). Nevertheless, we observed some conical or tubular shaped mature-like HIV-1 cores (Fig. 7.2 C), with an apparent lattice structure (Fig. 7.2 C, inset), formed only upon in situ PR cleavage, suggesting a successful viral maturation process in the *in vitro* system.

Can such a transition from immature Gag lattice to a mature capsid take place without any host cell factors or other viral proteins? To address this question, we employed our all recombinant and reconstituted *in vitro* maturation system and carried out cryoEM analysis of PR-treated Gag VLPs from the same samples analyzed biochemically (Fig. 7.1 C&D, marked with "*"). In the presence of yeast tRNA (10% w/w), purified recombinant Gag polyprotein assembled very efficiently into spherical particles about 100 nm in diameter with well-ordered surface lattices (Fig. 7.2 D&G). The particles were stable and remained intact after 2 hrs incubation with digestion buffer (Fig. 7.1 C "*1", Fig. 7.2 D&G). Upon completion of PR digestion (Fig. 7.1 C "*3"), large clusters, comprising numerous cleavage

products, were present (Fig. 7.1 F), in stark contrast to the entirely monodispersed initial Gag VLPs (Fig. 7.1 C "*1", Fig. 7.2 D). Remarkably, the clustered cleavage products were either conical or tubular shaped assemblies, resembling the mature-like structure (Fig. 7.2 I, red arrows). A few hybrid entities (mix of immature Gag lattice and mature CA lattice) were occasionally observed (Fig. 7.2 I, black arrows), suggesting incomplete conversion. We also imaged samples processed with a lower PR concentration (Fig. 7.1 C "*2"). Not surprisingly, we found fewer cones and tubes and more incompletely cleaved particles (Fig. 7.2 E&H), compared to the sample that underwent cleavage to completion. Intriguingly, the majority of cleaved products appeared to be derived from discrete Gag spheres, some with clear condensed RNP inside (Fig. 7.2 H, yellow arrow), whereas a small number of long tubes (Fig. 7.2 I) resulted from growth or fusion. These data suggest that the transition from immature Gag sphere to mature-like conical CA structure can occur without host cell factors, viral membrane or viral proteins other than Gag and PR. More importantly, individual immature particles seem to have been directly converted into cones or tubes. Purified CA protein can assemble into cones and tubes *in vitro*, but this requires high protein concentration (¿80 μM) and very high salt (¿1M NaCl) []. Given the relatively low CA concentration (37 μM) and very low salt (no added NaCl) in our cleavage reactions, *de novo* assembly of cleaved CA into these structures is unlikely. Taken together, these results support, at least in part, a displacive maturation pathway rather than the disassembly and reassembly process [19, 17].

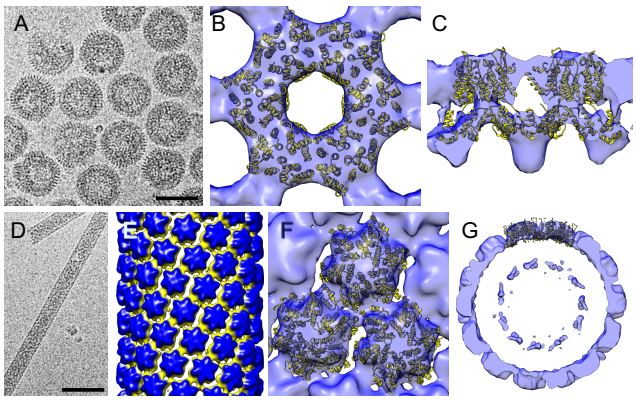


Figure 7.3: 3D reconstruction of Gag VLPs before and after PR cleavage. (A-C) The cryoEM image (A) and the subtomogram averaged 3D map (B and C) of Gag VLPs before PR cleavage. The resulting map is contoured at 1.5σ, with immature CA lattice (PDB code 4USN, yellow) docked into the density map, viewed along (B) and perpendicular to (C) the radial direction. (D&E) The cryoEM image and 3D reconstruction of the tubular structures present after PR cleavage of VLPs. The resulting map is contoured at 1.5σ, and colored radially from yellow to blue. (F&G) Fitting of mature CA hexamer (PDB code 3J34, yellow) into the density map, viewed from tube surface (F) and along the tube axis (G). The map is contoured at 1.5σ. The mature CA hexamer, not the immature, docks well into the EM density map. The inner layer density (seen in G) is likely the cleaved NC protein. Scale bars, 100 nm.

CryoEM structures of Gag VLPs and maturation product

To confirm that the assembled spherical Gag VLPs do indeed have an immature lattice whereas the final cleavage products have mature-like structures, we obtained 3D density maps of these assemblies by cryoEM. Using cryoET and sub-tomogram averaging [180, 181, 202], we obtained a 3D density map of Gag VLPs (Fig. 7.3 A) to 18.4 Å resolution by gold-standard Fourier shell correlation (FSC). As shown in Fig. 7.3 B&C, at this resolution, the averaged density map clearly displayed an immature configuration, overlapping well with the previously published immature HIV-1 structure 5. Docking of the immature structural model 4USN [181] into our density map exhibited a very good fit (Fig. 7.3

B&C), with a cross correlation coefficient of 0.81. Therefore, the assembled Gag VLPs used for *in vitro* maturation are indeed in the immature lattice arrangement.

Upon PR cleavage of Gag VLPs to completion (i.e. all Gag molecules converted to CA, Fig. 7.1 C "*3"), we observed a small fraction of long tubes that were amenable for structural analysis (Fig. 7.3 D), along with numerous conical shaped structures (Fig. 7.2 F&I). Combining the segments from the tubes with the same helical symmetry and using a single particle real space helical refinement approach [160, 203], we obtained a 3D reconstruction, at 22 Å resolution, of the tubular products that resulted from PR cleavage (Fig 7.3 E-G). In contrast to the undigested precursor, the tube density map showed an entirely a mature-like conformation (Fig. 7.3 E), with hexameric CA subunits arranged in helical arrays, as seen in mature CA assemblies [160]. The mature CA structural model 3J34 [160] fitted into the density map very well, further confirming that the tubes have a mature CA conformation. In contrast, docking the mature hexamer 3J34 into the undigested (immature) map and the immature hexamer 4USN into the digested (mature) map resulted in very poor fitting, with cross correlation coefficient values of 0.39 and 0.4, respectively (Supplementary Fig. 7.8). More interestingly, unlike our previous *in vitro* CA tubular assembly, additional densities, likely corresponding to the released NC and RNA, appeared inside these tubes. The existence of such density within the tubes clearly suggests that these mature-like CA tubes are not formed by *de novo* assembly of completely cleaved CA monomers, but instead through some maturation intermediates that maintain released RNPs inside the structure.

CA association following Gag cleavage

We observed only a few mature cores upon *in vitro* cleavage of PR-defective immature virions (Fig. 7.2 C), a surprising result considering the large number of starting immature particles. One plausible explanation for this paucity of cores is dissociation of CA molecules from the particles upon cleavage. To test this, particles were pelleted after PR digestion and particle-associated CA was detected by immunoblotting. As shown in Fig. 7.4 A, a significant amount (more than 1/3) of CA and CA-SP1 remained pelletable, i.e. associated with the particles, during the *in vitro* maturation process, after NC and MA had been cleaved off. Given that the viral membranes were permeabilized to allow PR to enter viral particles, which would also allow free CA to leave, the particle-associated CA may have arisen from either direct conversion from immature lattice without dissociation, or via a disassembly/reassembly process. If cleaved CA molecules dissociated from the lattice and reassembled into mature capsid, one might expect from this *de novo* assembly model that capsid-destabilizing mutations and capsid assembly inhibitors would affect the reassembly process and thus inhibit mature capsid formation. We thus investigated the effect of capsid-destabilizing mutation K203A, which does not affect particle production but abolishes HIV-1 infectivity, on CA association following Gag cleavage. As shown in Fig. 7.4 B, the K203A CA mutation did not decrease the level of particle-associated CA upon PR treatment. Similarly, capsid assembly inhibitor CAP-1 did not inhibit the *in vitro* maturation process (Fig. 7.4 C). These studies suggest that maturation could occur via conversion of the immature lattice to the mature lattice without requiring dissociation of the CA subunits.

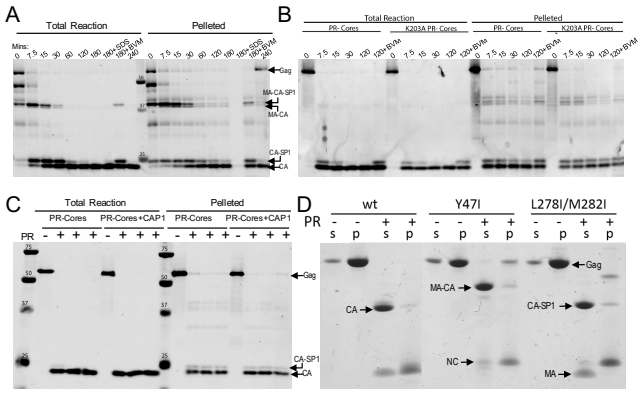


Figure 7.4: CA association following Gag cleavage by PR. (A) Immunoblot analysis of PR-treated immature virions in the presence of bevirimat (BVM) where indicated. PR-negative immature virions were digested with 1 μM PR at 37°C for varying amounts of time, up to 4 hrs. Particles were pelleted after digestion and CA association with core was monitored by immunoblotting with CA-specific mouse monoclonal antibody. (B) Effect of the capsid-destabilizing mutation K203A on CA association following Gag cleavage. Purified cores from PR-defective particles with wild type (wt) or K203A mutant CA were treated and analyzed as in (A). (C) Effect of CAP-1 on CA association during *in vitro* maturation. Cores from PR-deficient HIV-1 particles were treated as in (A) in the presence of 20 μM capsid assembly inhibitor CAP-1 for 2.5 hrs, then analyzed as in (A). (D) Effect of cleavage mutants on CA association during *in vitro* maturation by PR digestion of Gag VLPs. Supernatant (s) and pellet (p) of cleavage products following 3.3 μM PR digestion at 37°C for 2hrs, analyzed by SDS-PAGE gel and visualized by Coomassie Blue staining. The corresponding cleavage products are labeled.

On the other hand, *in vitro* cleavage of Gag VLPs yielded very little associated CA in pellet (Fig. 7.4 D, wt), in contrast to the PR-deficient immature particles (Fig. 7.4 A). In fact, more than 95% of the CA dissociated from the particles and was released into solution (Fig. 7.4 D, wt), consistent with the observed infrequent occurrence of the clustered cleavage products in Fig. 7.2 F&I, seemingly suggestive of a disassembly model. To further investigate the process of *in vitro* PR cleavage of Gag VLPs, we purified and assembled Gag VLPs carrying mutations preventing cleavage at the MA-CA (Y47I mutation) and CA-SP1 (L278I/M282I mutations) junctions [204]. Like the corresponding wild type Gag protein, both mutants efficiently assembled into spherical particles. PR treatment of Gag mutant VLPs yielded the expected cleavage products, MA-CA and CA-SP1, with more than 95% of MA-CA and CA-SP1 in the soluble fraction, as seen for CA in the wt VLPs. On the contrary, the NC products released from the wild type Gag protein and both Gag mutants were found in the pellet fraction. These studies suggest that the cleavage products, CA, MA-CA and CA-SP1, can dissociate from the particles after removing NC and RNPs.

Role of membrane and genome in the formation of mature core

The conversion of spherical immature particles into conical and tubular CA assemblies and the observed dissociation of CA molecules from particles upon cleavage, appear to challenge both the entirely *de novo* assembly model and the simple displacive model. We used computer modeling and simulation to explore the underlying physical mechanisms and constraints that are relevant to capsid assembly.

Building upon our previous simulation work [112, 174, 175, 176], we used a simple model of viral capsid growth to first investigate the physical impact of the virion membrane

on the formation of cylindrical or conical shells. We modeled the membrane as a spherical case, which sets a constraint for the shell growth such that the subunits remain within a sphere of radius, r_{mem}. In the model, the membrane-subunit interaction is described as a potential for an interior soft wall, $\sum_i (d_i^m - r_{mem})^2$, if $r_{mem} < d_i^m$, with d_i^m the distance between the i^{th} vertex of the subunit and the center of the sphere, and the potential vanishes for subunits within the radius. As shown in Fig. 7.5, the membrane restricts the growth of the otherwise extended incomplete shell (Fig. 7.5 A) and induces local stresses causing formation of the pentamers necessary for the assembly of closed conical or tubular shells (Fig. 7.5 B). More interestingly, the membrane could exert a force on the growing sheet that breaks the symmetry and causes the shell to grow as a cone rather than a cylinder (Supplementary Movie 1). Statistical analysis of 100 simulations showed that the frequency of forming a conical shell increased to 51% when constrained in the membrane compared to 38% without the membrane (Supplementary Table 1) suggesting that the virion membrane promotes conical shell formation.

To explore the impact of the genome (or RNP) on shell formation, we modeled the genome as a soft ball, due to steric interactions, with the interaction between the genome and the shell presented through the potential, $\sum_i (d_i^g - r_g)^2$, if $r_g > d_i^g$, with r_g the radius of gyration of the genome (soft ball) and d_i^g the distance between the i^{th} vertex of the subunit and the center of the genome. This potential keeps the subunits outside of the ball. If $r_g < d_i^g$, no interaction occurs between the genome and protein subunits (CA).

We considered the impact of the genome by employing two different models. In the first model, we assumed that a few subunits are attached to the genome but the complexes of

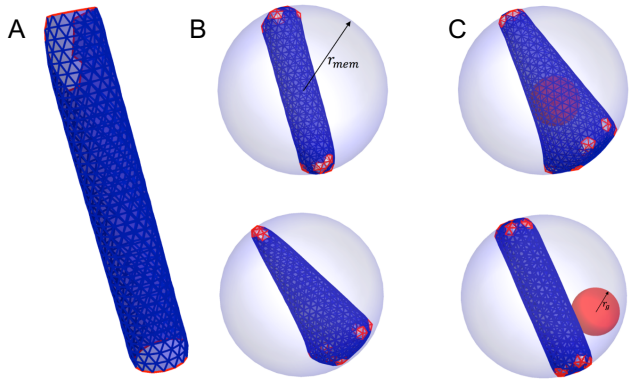


Figure 7.5: Role of membrane and genome in the formation of mature core. (A) Formation of open tube in free environment. (B) The presence of membrane (the gray sphere) limits the size of the cylinder. Membrane can also facilitate formation of pentamers resulting into the formation of cones. (C) During the assembly process, if the genome (ball) remains attached to a few subunits, a conical capsid forms. Otherwise, a cylindrical shell assembles with genome remaining outside, see the movie in the supplementary material.

the genome-subunits are free to move inside the membrane. If the genome remains attached to the growing sheet during the initial stages of the growth process, the conical shell forms with the genome inside (Fig. 7.5 C, Supplementary Movie 2). If the genome detaches from the growing shell early in the growth process, a cylindrical shell forms with the genome outside (Fig. 7.5 C, Supplementary Movie 3). As shown in Supplementary Table 1, the presence of the genome increased the frequency of cone appearance to 84%, similar to the experimentally observed cone to tube ratio of 20:1 [13].

In the second model, we kept the genome-subunit complexes attached to the membrane during the growth process until about one-sixth of the capsid had formed, at which point, we released the genome-subunit complex. The simulations show that the hexagonal sheet continues to grow asymmetrically along one side of the shell until it meets the membrane, which exerts a force on the shell and makes it roll and grow back towards the

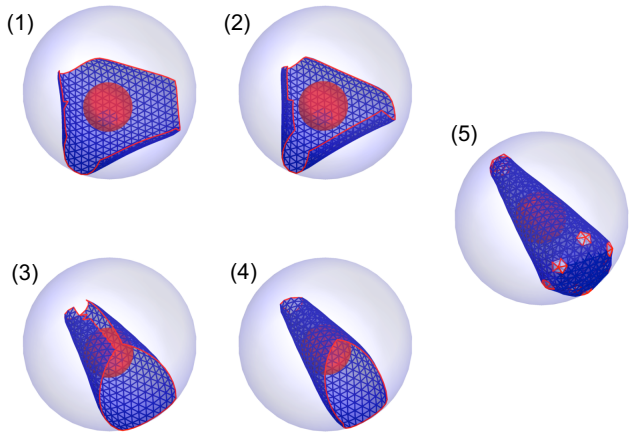


Figure 7.6: Formation of a conical capsid through a combination of displacive and disassembly models. Initially a lattice formed from 600 hundred subunits detaches from the membrane. The lattice undergoes maturation (step 1-5), i.e., the mechanical properties and the spontaneous curvature of lattice transforms from an immature capsid to a mature one.

RNP (Supplementary Fig. 7.9, Supplementary Movie 4). Similar to the first model, if the genome is attached to the growing shell, more conical capsids with genome inside occur, while, if the genome detaches, cylindrical capsids form with the genome outside. In both models, the initial attachment of the genome to the growing shell is important for the formation of a conical shell.

Formation of a conical capsid through a sequential combination of displacive and disassembly modes

We further investigated capsid formation from a partially formed shell while in an immature state at the time of release from the membrane. Based on the displacive model, we considered that the sheet of capsid proteins retains the same connectivity as Gag proteins in the immature virion. We assumed that the subunits in the immature sheet, containing as

many as 600 subunits, undergo some conformational changes without disassembly resulting in their preferred dihedral angle and flexibility similar to the mature shells. The simulation shows a large sheet of subunits attached to the surface of the spherical membrane and growing. The sheet then detaches from the membrane and subsequently rolls around the genome. Finally the sheet grows into a closed conical shell in a process identical to the *de novo* assembly model described above (Fig. 7.6, Supplementary Movie 5). The outcome of these simulations depended upon the size of the initial immature sheet. For large sheets with around 600 subunits or larger, we observed a decrease in the efficiency of cone formation and an increase in the occurrence of defective cones and cylinders. The sheets with a smaller number of subunits in the initial immature lattice have a much higher probability to grow to cones without defects. When we decreased the number of subunits in the initial "immature" lattice to 400, the lattice rolls around the genome but is not large enough for the two edges to merge immediately. Addition of free CA subunits to the growing sheet results in the formation of cones with almost no defects (Supplementary Fig. 7.10 and Supplementary Movie 6). In fact, simulation results did not support a direct conversion from the complete immature lattice, as suggested by a simple displacive model. Thus, we conclude that the most efficient way to form conical shells is a sequential combination of displacive rearrangement and growth. Starting with a partially formed immature sheet speeds up growth, because the entire lattice does not need to disassemble and reassemble; however, a diffusive growth phase appears to be necessary to minimize assembly defects.

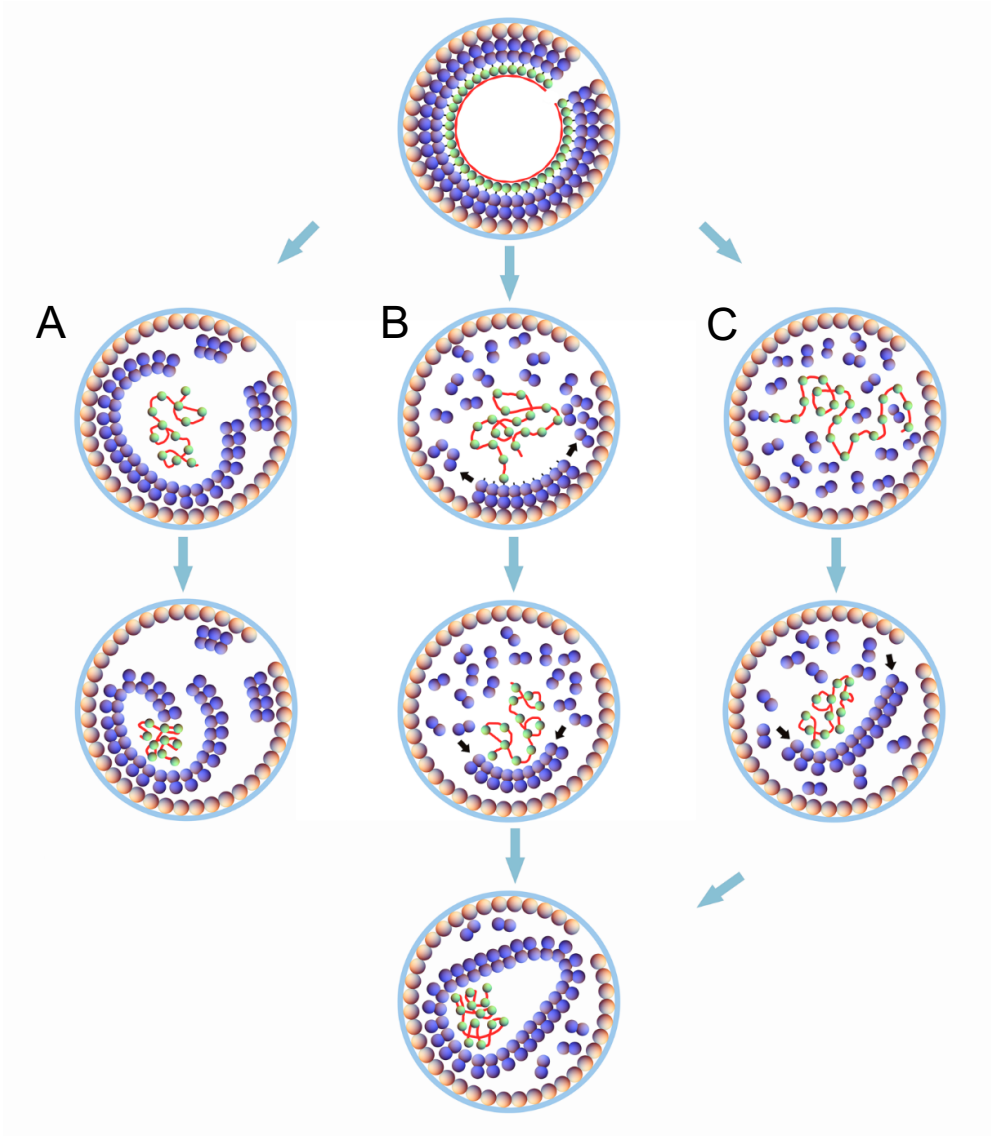


Figure 7.7: A model for different HIV-1 maturation pathways. Top figure shows an immature shell composed of a membrane (blue circle), MA proteins (orange balls), CA proteins (purple balls), NC proteins (green balls) and RNA (the red line). Column A shows the fully displacive model in which the detached immature lattice undergoes maturation. Column B illustrates the sequential combination of displacive and assembly model. A portion of the immature lattice undergoes maturation followed by addition of CA proteins to the growing edge. Column C indicates the disassembly and *de novo* reassembly model. The immature lattice completely disassembles and then the disassembled CA proteins assemble *de novo*.

7.3 Discussion

Previous studies have led to two competing HIV-1 maturation models, the *de novo* disassembly and reassembly model and the lattice displacive model, for the conversion of immature Gag lattice to mature conical capsid. To distinguish these HIV-1 capsid maturation models, we devised novel *in vitro* PR cleavage systems to mimic HIV-1 maturation and monitored the maturation process with biochemical and structural analysis, as well as computer simulations. The results presented in this study challenge both existing models for their limitations and further suggest a sequential process of initiation via displacive remodeling followed by growth via soluble subunit addition for HIV-1 capsid assembly, as illustrated in a schematic presentation (Fig. 7.7).

Our results suggest that the *de novo* assembly model cannot be the sole pathway, given that individual Gag VLPs can be converted into conical shaped structures by PR cleavage *in vitro* under conditions that are known to not support assembly of soluble CA molecules into conical or tubular structures. Consistent with a recent cryoET study with large membranes enclosing multiple capsids [19], the size of Gag spheres, rather than the viral envelope membrane, which is absent in our *in vitro* system, determines the size of the mature capsid. These data are in conflict with the complete *de novo* disassembly/reassembly model which predicts that the mature core begins to grow at its narrow or wide end, from nucleation of CA units on one face of the membrane, and stops growing when it reaches the boundary of the membrane on the opposite side [196, 3]. Furthermore, our study, as well as two recent cryoET analyses [19, 16], have identified the formation of CA sheet-like structures as an intermediate during the maturation process. While Woodward et al. [16]

interpreted this as a *de novo* CA assembly product, data from our and Frank's work [19] suggest that these initial CA sheets form through a displacive process. In addition, HIV-1 particles with functional MA-CA cleavage but defective CA-SP1 cleavage (CA5 particles) exhibit a layer morphology that is consistent with the structure of the mature CA lattice [187, 193, 195], indicating that the CA-SP1 layer is formed upon release from the membrane and dissociation of most CA from the lattice prior to capsid assembly is unlikely. Moreover, in our *in vitro* PR cleavage assays, capsid-destabilizing mutations and capsid assembly inhibitors had little effect on CA association, further arguing against the simple *de novo* reassembly model.

On the other hand, our results also argue against an entirely displacive model, as suggested by Frank *et al* [19], given our observations that a major fraction of CA and CA-SP1 are indeed dissociated from the Gag lattice during *in vitro* PR cleavage of Gag VLPs without membrane. Formation of long CA tubes in our *in vitro* system also supports capsid growth from diffusive addition of subunits. If the capsid is formed exclusively by the non-diffusional phase transition of immature to mature CA lattice, one would expect to find that the stress and deformations inherent in this process result in defects in the mature CA lattice. In fact, our computer simulations failed to yield a closed mature core when employing only the non-diffusional phase transition (Supplementary Movie 7), as the transformation of an elastic sheet to a closed conical shell requires disassembly of many subunits at the growing edge and inclusion of pentamers in the "right" position for the shell to close properly. Therefore, based on our experimental data and computer simulation

results, we propose a new combinatory pathway for HIV-1 capsid maturation, which begins with non-diffusional transition and followed by the CA dissociation and growth.

This sequential displacive rearrangement and growth model is consistent with nearly all the previous experimental observations. From the energetic point of view, this model would be kinetically favored, as an assembly pathway involving both displacive and *de novo* modes is advantageous as a part of the shell is already built and does not need to reassemble and this makes the process more efficient. The model further implicates that the viral membrane, MA and the RNP likely guide the core assembly pathway [16, 205]. Interestingly, a recent cryoET study revealed that viral particles with the NC-SP2 domains replaced by a leucine zipper lacked condensed RNPs and contained an increased proportion of aberrant core morphologies [206]. Consistent with our maturation model, a role for HIV-1 integrase (IN) in initiating core morphogenesis and viral RNP incorporation into the mature core has been recently suggested, and IN has emerged as a maturation determinant [207, 208]. The effect of IN may explain the apparent paradox that *in vitro* digestion of PR- virions results in 1/3 of CA being pelletable, while this was only 5% for *in vitro* assembled particles, owing to the presence of Pol in PR- virions and perhaps additional virion-associated host factors. Our model also explains the observed potent trans-dominant effect of uncleavable Gag on viral infectivity, where 4% of Gag with MA-CA linker resulted in a 50% decrease in infectivity [188, 209, 210]. The presence of even a few such Gag subunits that remain attached to the membrane would preclude CA lattice sheet detachment, thus inhibiting capsid formation and leading to a disproportionate drop in infectivity. Furthermore, in our *in vitro* maturation system, the maturation inhibitor BVM not only

blocked cleavage of CA-SP1, but also partially protected the MA-CA cleavage site. The immature lattice-stabilizing effect of BVM has been previously observed in cryoET studies [193] and is reflected in our new combinatory capsid maturation pathway model. Our model also explains the predominance of conical vs. cylindrical structures.We find that any asymmetry developed in the growing lattice due to interaction withthe membrane or genome (Fig. 7.5), or due to the shape of initial immature lattice creates conical capsids, as opposed to cylindrical shells (Supplementary Fig. 7.11).

HIV-1 and other retroviruses have optimized the feat of packaging their genome and employing the accessible pathways to maturation. While many aspects of the HIV-1 maturation process remain to be explored, our combined biochemical, structural and computational study allows us to integrate our new findings with many previous observations into a new model for HIV-1 maturation. Our model of initial displacive rearrangement from immature Gag followed by growth to the mature capsid, reconciles most of the current experimental evidence and provides a new pathway to elucidate novel therapeutic targets for preventing virion maturation and subsequent infectivity. Moreover, the novel *in vitro* maturation systems that we established in this study will be useful for many researchers to dissect the mechanisms of retrovirus maturation and to test the effects of maturation inhibitors on this process.

7.4 Methods

Protein expression and purification

The cDNA encoding gag polyprotein, $Pr55^{gag}$ was obtained from the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH [211]. To generate Gag ($\Delta MA_{15-100}\Delta p6$), plasmid Prr400 encoding corresponding Gag regions was subcloned into pET21 (EMD chemicals, Inc. San Diego, CA) using NdeI and XhoI sites. Proteins were expressed in*E. coli*, Rosetta 2 (DE3), cultured in Luria-Bertani media or modified minimal medium, and induced with 0.5mM IPTG at 23°C for 16h. All the mutants were constructed from pET21 Prr400 with site-direct mutagenesis.

Cell pellets were collected and resuspended in Lysis buffer (25mM Tris, pH7.5, 0.5M NaCl, 1μM ZnSO4, and 10mM 2-Mercaptoethanol) and broken with a microfludizer. Subsequently, 0.11 total volume of 10% polyethyleneimine (pH8.0) was added to precipitate nucleic acids. The lysate was then centrifuged at 35,000 g for 30 min, and the pellet discarded. To the supernatant was added ammonium sulfate to 60% saturation, on ice, with stirring for 2hr. The solution was then centrifuged at 10,000 g for 30 min. The pellet was resuspended in lysis buffer and dialyzed overnight against dialysis buffer (25mM Tris, pH7.5, 50mM NaCl, 1μM $ZnSO_4$, and 10mM 2-Mercaptoethanol). The supernatant was loaded to ion-exchange chromatography (MonoS 10/100 GL column, GE Healthcare, Piscataway, NJ), using 50mM-1M NaCl in 25mM Tris, pH7.5, 1μM $ZnSO_4$, and 10mM 2-Mercaptoethanol. The final purification step for all proteins comprised size-exclusion chromatography (Hi-Load Superdex 200 26/60 column, GE Healthcare, Piscataway, NJ) in 25mM Tris, pH7.5, 0.15M NaCl, 1μM $ZnSO_4$, and 10mM 2-Mercaptoethanol. Proteins

were concentrated to 4mg/mL using Amicon concentrators (Millipore, Billerica, MA), flash-frozen with liquid nitrogen and stored at $-80°$C.

In vitro Gag VLP assembly

Protein at 4mg/ml in storage buffer was mixed with yeast tRNA at a nucleic acid/protein ratio of 10% (wt/wt), and slowly diluted to 1.5 mg/ml by dropwise addition of 50 mM sodium acetate (pH6.0), 100 μM $ZnSO_4$, and 5 mM DTT or 50 mM Tris-HCl (pH7.4), 100 μM $ZnSO_4$, and 5 mM DTT. The mixture was then incubated at 10°C overnight.

Isolation of PR-deficient immature virions

Immature HIV-1 cores were isolated from particles produced by transfection of 293T cells with the PR-inactive HIV-1 proviral clone R9.PR-, as previously described [212]. Supernatants from transfected 293T cells were filtered through a 0.45 μm pore-size filter to remove cellular debris, and particles were pelleted by ultracentrifugation (120,000 xg for 3 hr at 4°C using Beckman rotor SW32Ti) through a cushion of 20% sucrose in PBS buffer. The viral pellet was resuspended in 250 μL of STE buffer (10 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH 7.4) and incubated on ice for 2 hr prior to ultracentrifugation (120,000 g for 16 hr at 4°C) through a 0.5 ml layer of 1% Triton X-100 into a 10 ml linear gradient of 30% to 70% sucrose in STE buffer. Immature cores were recovered from fractions 8 and 9 of the gradient, corresponding to the known density of retroviral cores. Aliquots of the immature core-containing fractions were flash frozen in liquid nitrogen and thawed once for

use in PR digestion assays. For cryoEM studies, the immature cores were prepared for use without freezing.

In vitro cleavage of PR-deficient immature virions and Gag VLPs

Frozen aliquots of purified cores from PR-deficient particles were thawed, and 50-100 ng of cores was added to 50 μL reactions containing 100mM NaOAc, 100mM NaCl, 1mM EDTA, 1mM DTT, at the indicated pH and concentrations of recombinant HIV-1 PR (kind gift from Dr. Celia Schiffer, University of Massachusetts). Reactions were incubated at 37°C and stopped by the addition of 20 μM of HIV-1 PR inhibitor (Crixivan). A fraction (66%) of each reaction was removed, diluted to 200 μl in reaction buffer lacking PR, and pelleted by centrifugation (45,000 rpm for 30 min at 4°C in a Beckman TLA-55 rotor). Pellets were dissolved and subjected to 12% SDS-PAGE and immunoblotting with CA-specific mouse monoclonal antibody 183-H12-5C. For cryoEM studies, PR digestion reactions were performed in 200 μl volumes containing 1500 ng of immature cores. For the cryoEM studies shown in Figure 7.2 panels A-C, PR-deficient HIV-1 particles were concentrated by pelleting through a 20% sucrose cushion without detergent exposure, resuspended in STE buffer, and used in PR digestion reactions in the presence and absence of Triton X-100.

PR digestion experiments with recombinant VLPs were performed by addition of various concentrations of PR to the Gag assembly mixture, and incubated at 37°C for 2hr. For kinetic analysis of Gag cleavage, 3.3 μM of HIV-1 PR was added to the Gag assembly mixture and incubated at 37°C; at different time points, 25 μl of the digestion reaction mixtures was taken out and put on ice to stop the reaction, and then subjected to SDS-

PAGE and cryoEM analysis. The cleavage products were separated by NuPAGE Novex 4-12% Bis-Tris gel (Invitrogen) and visualized by Coomassie blue staining.

Incubation with maturation inhibitor

Gag assemblies at different pH values (pH7.4 and pH6.0) were incubated with 40 μM BVM (kind gift of Chin-Ho Chen, Duke University) and 3.3 μM HIV-1 PR at 37°C for 2hr. Then 0.5 μl of the whole sample was mixed with loading buffer for SDS-PAGE analysis. Leftover was processed for cryoEM analysis. 2 μl samples were mixed with loading buffer and separated by NuPAGE Novex 4-12% Bis-Tris gel (Invitrogen) and visualized by Coomassie blue staining.

CryoEM specimen preparation and data collection

Gag assemblies with or without PR cleavage were applied (4 μl) to the carbon side of a glow discharged perforated Quantifoil grid (Quantifoil Micro Tools, Jena, Germany). The grids were then manually blotted with a filter paper from the backside, to remove the excess fluid, and plunge-frozen in liquid ethane using a home-made manual gravity plunger. For cryoEM imaging, the frozen grids were loaded into a cryo-holder (Gatan Inc., Pleasanton, CA) and inserted into a Tecnai F20 transmission electron microscope (FEI, Inc., Hillsboro, OR) and imaged with a 2k2k charge-coupled device camera (Gatan). Low dose (10$\sim$20 $e^-/Å^2$) projection images were collected at a nominal magnification of 50,000 with a pixel size of 4.52 Å and underfocus values of approximately 5.0 μm.

3D reconstruction of helical tubes

Well-ordered long straight tubes from Gag cleavage samples were Fourier transformed and indexed for helical symmetry. Two tubes belonging to the same helical family (-14, 11) were included in the final reconstruction. In total, 333 segments of 130130 were boxed out for image processing, and 3D helical refinement and reconstruction were carried out as previously described [203]. During the refinement, helical symmetry and contrast transfer function correction were applied. The UCSF Chimera package was employed for 3D visualization and isosurface rendering [213]. The resolution of the final 3D reconstruction was estimated to be 28/22 Å at the 0.5/0.143 thresholds respectively, from the gold standard Fourier shell correlation (FSC) curve.

Cryo-electron tomography and sub-tomogram averaging of VLPs

Gag VLP assemblies (4 μl) were applied to the carbon side of glow discharged perforated R2/2 Quantifoil grids and quickly mixed with 1 el of a 10 nm fiducial gold bead solution before plunge-freezing using a manual gravity plunger. Tomography tilt series were collected at a nominal magnification of 59,000 (effective pixel size of 1.9 Å) on a Tecnai G2 Polara electron microscope operated at 300 kV. A series of images were recorded on an FEI Falcon II direct electron detector by tilting the specimen from -45 to 55 in increments of 3 tilt angles. Altogether 34 images were collected in one tilt series with a dose of 30 $e^-/Å^2$. Images were recorded at a defocus value of 4-5 μm using FEI batch tomography software.

Eight tilt series were corrected for the phase inversions due to the Contrast Transfer Function (CTF) using a tile based approach, and subsequently aligned and reconstructed

using weighted-back-projection in IMOD [214]. We selected 28 Gag VLPs for further processing from the resulting tomograms. To extract sub-tomograms corresponding to the Gag lattice, initial positions respective to a Cartesian grid defined by each tomogram, were approximated by using a template matching algorithm implemented in Matlab with a reference derived from the previous published immature structure low-pass filtered to 6nm (EMD-2706) [181]. This resolution as well as a coarse angular search were chosen to eliminate any statistical correlation of high resolution information between half data sets in later image processing steps. Following template matching and sub-tomogram extraction, which we limited to the top scoring 300 sub-tomograms from each VLP, the data were randomly split into two groups of 4,200 sub-tomograms each, which were processed independently for all subsequent steps. Each sub-tomogram was iteratively aligned to the average from its respective half-set by using standard cross-correlation based methods. After eliminating the lowest scoring sub-tomograms, the remaining half-sets had 2,000 Gag-hexamers each, with the Fourier Shell Correlation (FSC) between them reaching 23.2/18.4A at the 0.5/0.143 thresholds respectively.

Computer model building

Computer simulations were performed based on a model that we had previously developed and employed to successfully explain the *in vitro* formation of conical structures in the absence of membrane and genome as well as the defective structures observed in the *in vivo* experiments [176]. Simulations based upon these previous works were designed to test the impact of the genome and the virion membrane on the formation of conical shells and to investigate the different maturation pathways.

The building blocks of both mature and immature shells were modeled as triangular subunits because they can pack to form hexagonal sheets in flat space [174, 175, 176] similar to the Gag and CA molecules. The difference between immature and mature capsids in our model is presented in the mechanical properties of the subunits forming the triangular lattice. It is modeled as an elastic sheet that has an associated bending energy resulting from deforming it away from its preferred curvature, and a stretching energy that comes from distorting the shape of the triangles in the mesh. The bending energy is modeled as simple torsional springs between all neighboring triangles, $E_b = \sum_{<ij>} k_b(1 - \cos(\theta_{ij} - \theta_0))$ with the sum over all neighboring triangles $< i, j >$, k_b a torsional spring constant. The quantity θ_{ij} is the angle between the normal vectors, and θ_0 is the equilibrium angle between the triangles. We can relate R_0 to the spontaneous radius of curvature for the capsid sheet as $\sin(\theta_0/2) = (12R_0^2/b_0^2 - 3)^{-1/2}$. The stretching energy corresponds to a network of springs, $E_s = \sum_i \sum_{a=1}^{3} \frac{k_s}{2}(b_i^a - b_0)^2$ with the sum over all subunits i in the mesh, b_i^a is the length of the a^th bond, and b_0 the equilibrium bond length.

The equilibrium shape of the capsid sheet is found by allowing the mesh to relax to its minimum energy configuration. This is done numerically using a non-linear conjugate gradient method [215]. The minimum of the total energy of the mesh depends on only two dimensionless constants, the ratio of the linear and torsional spring constants also called the Foppl von Karman number $\tilde{\gamma} = \frac{b_0^2 k_s}{k_b}$, and the equilibrium angle between the triangles θ_0. The shape of the final shells formed depends upon the values of $\tilde{\gamma}$ and R_0, and previous work has explored the parameter space and mapped out the types of shells formed [174, 175]. The results we are presenting in this paper corresponds to one single $\tilde{\gamma}$ and R_0 where we

obtain cylindrical and conical shells rather than spherical or irregular [176, 14]. These two dimensionless parameters define the material properties of the elastic sheet. In the next section we briefly explain our growth model.

Computer simulations: Nucleation and growth

The assembly of a shell was modeled through the sequential, irreversible addition of a triangular subunit to an incomplete shell (Supplementary Fig. 7.12). After each assembly step, the shell is relaxed, assuming that the mean time between additions is much larger than the elastic relaxation time. This corresponds to a physical situation in which each new subunit will be added to the optimal position, *i.e.*, the number of newly formed bonds is maximized and the elastic energy is minimized. Due to the number of bonds formed after addition of each subunit, we consider the process is irreversible [174, 175], once a pentamer or hexamer formed, they can no longer dissociate.

During the growth process, if there is a location with 5 triangles attached, two exposed edges can either join to create a pentamer or a new triangle can be inserted to form a hexamer (Supplementary Fig. 7.12a). As the shell grows and gradually starts to roll, the two distance edges of the shell can become close together. If the distance between the edges is very small, they merge to mimic the hydrophobic interaction between the subunits at the edge (Supplementary Movie 1).

Acknowledgements

We thank Dr. Celia Schiffer, University of Massachusetts for the purified recombinant HIV-1 PR protein and Chin-Ho Chen (Duke University), for BVM. We thank Dr.

Xiaofeng Fu for help with cryoEM imaging and Dr. Teresa Brosenitsch for critical reading of the manuscript. Monoclonal antibody 193-H12-5C was obtained from Dr. Bruce Chesebro via the NIH AIDS Research and Reference Reagent Program. This work was supported by the National Institutes of Health NIGMS Grant P50GM082251, the Office of the Director Grant S10OD019995, and by the National Science Foundation through Grant No. DMR-1310687.

Supplementary figures

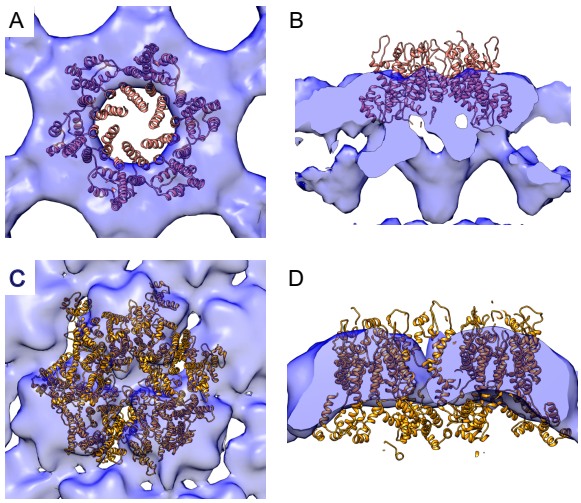


Figure 7.8: (A&B) Fitting of mature CA hexamer (PDB code 3J34, orange) into the Gag sphere density map, viewed along (A) and perpendicular (B) to the radial direction. (C&D) Fitting of immature CA lattice (PDB code 4USN, orange) into the tube density map, viewed from tube surface (C) and along the tube axis (D). The maps are contoured at 1.5σ.

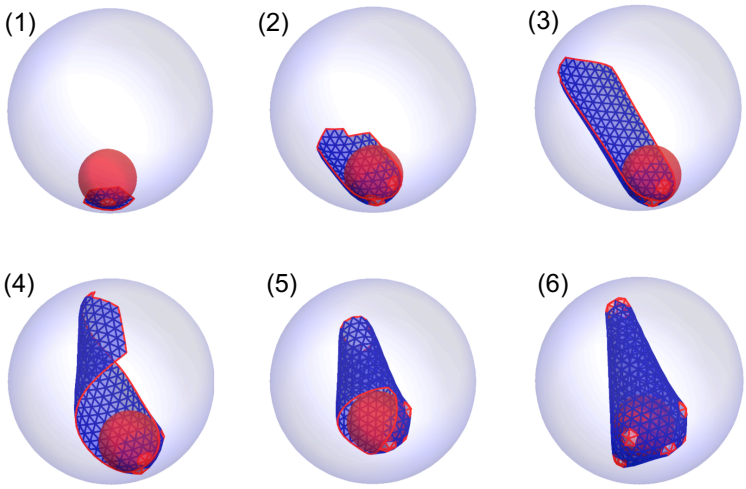


Figure 7.9: Formation of a conical capsid. The genome and a few subunits remain attached to the membrane at the beginning of the growth process.

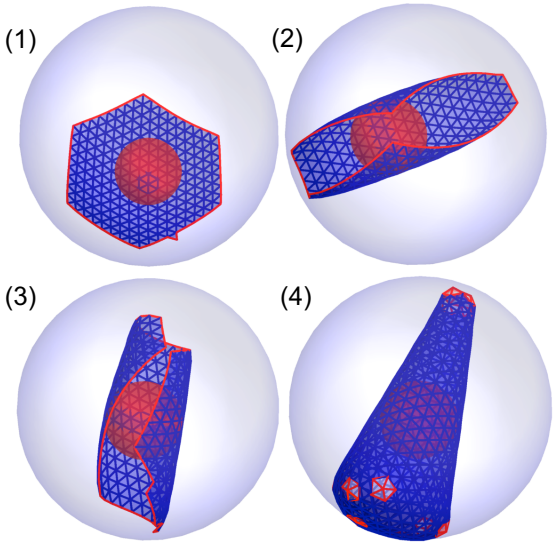


Figure 7.10: Snapshots of formation of a conical capsid through a combination of displacive and disassembly models. Initially a lattice formed from 400 hundred subunits detaches from the membrane. The lattice undergoes maturation, i.e., the mechanical properties and the spontaneous curvature of lattice transforms from an immature capsid to a mature one. The orange ball illustrates the genome.

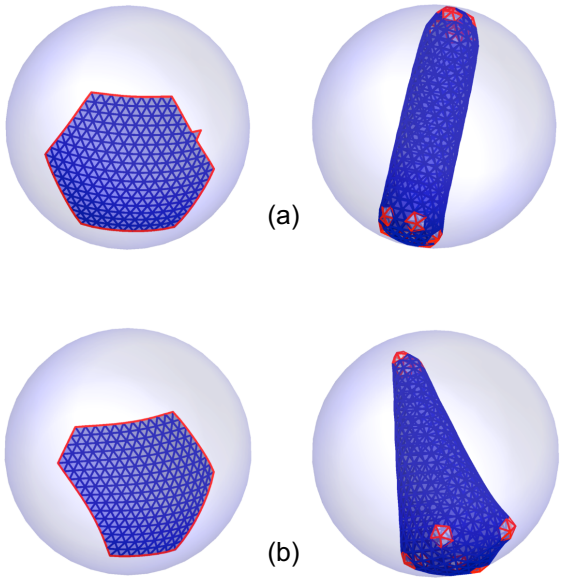


Figure 7.11: The role of initial condition and symmetry in the final structure of capsids. If the initial lattice has hexagonal symmetry, a cylindrical shell forms (a); otherwise a conical capsid forms (b).

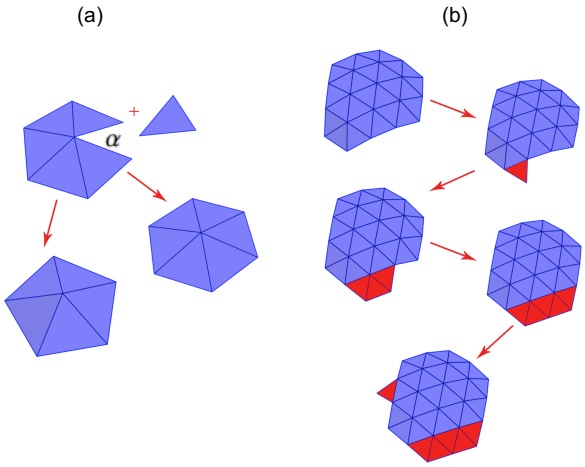


Figure 7.12: (a) Formation of a pentamer or hexamer depends on the opening angle. Five assembled subunits can either become a pentamer or a new subunit can be added to them to become a hexamer. (b) The step by step growth of a shell by addition of triangular subunits.

Chapter 8

Conclusions

In summary, we showed that the secondary structure of RNA is one of the key features of virions that makes assembly more efficient allowing viral RNAs to out-compete cellular RNAs during replication in infected host cells. RNA topology combined with the electrostatic interactions can offer an explanation to the overcharging observed in RNA viruses. Furthermore, the secondary structure of RNA clearly effects the osmotic pressure of the capsid resulting into more stability. The osmotic pressure becomes more negative for a virus particle with a branched polymer compared to the one with a linear chain. There have been several analytical models that include the electrostatics (non-specific) interactions which is believed to be the driving force of the virus assembly and important results have been reported. Combining the secondary structure of the RNA with electrostatic interactions in our model is an important step in realism of the modeling of virus assembly. However, the proposed model cannot include the specific interactions between the stem-loops of the secondary-structure of RNA and capsid domains. Further studies on

both specific and non-specific interactions could elucidate the mechanism of viral assembly and advance the development of antiviral drugs.

Furthermore, in this dissertation, we explored the role of the genome and membrane in the formation of conical shells and examined the HIV maturation pathways. Our results showed that the confinement free energy for a homopolymeric model genome confined in a conical capsid is lower than that in a cylindrical capsid, at least when the genome does not interact with the capsid, which seems to be the case in *in vivo* experiments. Conversely, the confinement free energy for the cylinder is lower than for a conical capsid if the genome is attracted to the capsid proteins as the *in vitro* experiments. This might be one of the reasons why conical shells are observed dominantly *in vivo*. We obtained the genome density profiles in both geometries and quite interestingly genome became condensed at the bottom of the cone similar to the genome profile in conical cores observed in experiments. In the last chapter of this dissertation, we simulated the HIV capsid growth using an irreversible coarse-grained model. We showed that presence of membrane or genome, or an asymmetric initial immature lattice shape breaks the local hexagonal symmetry, causing the formation of conical capsids. Additionally, to obtain a conical shape, viruses not only use *de novo* pathways but also take advantage of a different pathway, combination of *de novo* and *displacive* assembly, which we proposed here as a new HIV maturation model. In particular, we modeled both membrane and genome as soft balls. The membrane is a spherical boundary around the capsid and the genome takes a spherical shape condensing towards the bottom of the capsid. Further investigations, such as simulating the genome as a polymer, could improve the modeling and might advance our fundamental understanding

of HIV life-cycle. Understanding the structural mechanisms involved in the formation of HIV is a problem of central scientific interest to design of effective antiviral drugs.

Bibliography

[1] S. J. Flint, L. W. Enquist, R. M. Krug, V. R. Racaniello, and A. M. Skalka, *Principles of Virology: Molecular Biology, Pathogenesis, an Control*. Washington: ASM Press, 2000.

[2] R. F. Bruinsma, "Physics of RNA and viral assembly," *Euro. Phys. J. E*, vol. 19, p. 303, 2006.

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

[4] J. B. Bancroft and E. Hiebert, "Formation of an infectious nucleoprotein from protein and nucleic acid isolated from a small spherical virus," *Virology*, vol. 32, p. 354, 1967.

[5] J. B. Bancroft, E. Hiebert, and C. E. Bracker, "The effects of various polyanions on shell formation of some spherical viruses," *Virology*, vol. 39, p. 924, 1969.

[6] R. F. Garmann, M. Comas-Garcia, M. S. T. Koay, J. J. L. M. Cornelissen, C. M. Knobler, and W. M. Gelbart, "Role of electrostatics in the assembly pathway of a single-stranded rna virus," *J. Virol.*, vol. 88, p. 10472, 2014.

[7] N. Patel, E. Dykeman, R. Coutts, G. Lomonossoff, D. Rowlands, S. Phillips, N. Ranson, R. Twarock, R. Tuma, and P. Stockley, "Revealing the density of encoded functions in a viral RNA," *Proc. Natl. Acad. Sci.*, vol. 112, p. 2227, 2015.

[8] E. C. Dykeman, P. G. Stockley, and R. Twarock, "Packaging signals in two single-stranded rna viruses imply a conserved assembly mechanism and geometry of the packaged genome," *J. Mol. Biol.*, vol. 425, p. 3235, 2013.

[9] M. Comas-Garcia, R. D. Cadena-Nava, A. L. N. Rao, C. M. Knobler, and W. M. Gelbart, "In Vitro Quantification of the Relative Packaging Efficiencies of Single-Stranded RNA Molecules by Viral Capsid Protein," *J. Virol.*, vol. 86, p. 12271, 2012.

[10] R. F. Garmann, M. Comas-Garcia, C. M. Knobler, and W. M. Gelbart, "Physical principles in the self-assembly of a simple spherical virus," *Acc. Chem. Res.*, vol. 49, p. 48, 2016.

[11] B. K. Ganser-Pornillos, M. Yeager, and W. I. Sundquist, "The structural biology of hiv assembly," *Curr. Opin. Struct. Biol.*, vol. 18, p. 203, 2008.

[12] W. I. Sundquist and H.-G. Kraeusslich, "Hiv-1 assembly, budding, and maturation," *Cold Spring Harb. Perspect. Med.*, vol. 2, p. a006924, 2012.

[13] S. Li, C. P. Hill, W. I. Sundquist, and J. T. Finch, "Image reconstructions of helical assemblies of the hiv-1 ca protein," *Nature*, vol. 407, p. 409, 2000.

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

[15] J. A. Briggs, K. Grnewald, B. Glass, F. Frster, H.-G. Krusslich, and S. D. Fuller, "The mechanism of hiv-1 core assembly: Insights from three-dimensional reconstructions of authentic virions," *Structure*, vol. 14, p. 15, 2006.

[16] C. L. Woodward, S. N. Chenga, and G. J. Jensen, "Electron cryotomography studies of maturing hiv-1 particles reveal the assembly pathway of the viral core," *J. Virol.*, vol. 89, p. 1267, 2015.

[17] X. Meng, G. Zhao, E. Yufenyuy, D. Ke, J. Ning, M. DeLucia, J. Ahn, A. M. Gronenborn, C. Aiken, and P. Zhang, "Protease cleavage leads to formation of mature trimer interface in hiv-1 capsid," *PLoS Pathog.*, vol. 8, p. e1002886, 08 2012.

[18] P. W. Keller, R. K. Huang, M. R. England, K. Waki, N. Cheng, J. B. Heymann, R. C. Craven, E. O. Freed, and A. C. Steven, "A two-pronged structural analysis of retroviral maturation indicates that core formation proceeds by a disassembly-reassembly pathway rather than a displacive transition.," *J. Virol.*, vol. 87, p. 13655, 2013.

[19] G. A. Frank, K. Narayan, J. Bess, Julian W., G. Q. Del Prete, X. Wu, A. Moran, L. M. Hartnell, L. A. Earl, J. D. Lifson, and S. Subramaniam, "Maturation of the hiv-1 core by a non-diffusional phase transition," *Nature Communications*, vol. 6, p. 5854, 2015.

[20] P.-G. de Gennes, "Statistics of branching and hairpin helices for the dat copolymer," *Biopolymers*, vol. 6, p. 715, 1968.

[21] A. Y. Grosberg, A. M. Gutin, and E. I. Shakhnovich, "Conformational Entropy of a Branched Polymer," *Macromolecules*, vol. 28, p. 3718, 1995.

[22] P. Brion and E. Westhof, "Hierarchy and dynamics of rna folding," *Annu. Rev. Biophys. Biomol. Struct.*, vol. 26, p. 113, 1997.

[23] R. Gutell, B. Weiser, C. R. Woesse, and H. F. P. Noeller, "Comparative anatomy of 16-s-like ribosomal rna," *Nucl. Acid Res. Mol. Biol.*, vol. 32, p. 155, 1985.

[24] M. Daoud and A. Lapp, "Branched polymers and gels," *J. Phys.: Condens. Matter.*, vol. 2, p. 4021, 1990.

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

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

[27] P. van der Schoot and R. Zandi, "Impact of the topology of viral rnas on their encapsulation by virus coat proteins," *J Biol Phys*, vol. 39, no. 2, p. 289, 2013.

[28] P. Higgs, "Rna secondary structure: physical and computational aspects," *Quarterly Reviews of Biophysics*, vol. 33, p. 199, 2000.

[29] J. D. Perlmutter, C. Qiao, and M. F. Hagan, "Viral genome structures are optimal for capsid assembly," *eLife*, vol. 2, p. e00632, 2013.

[30] A. Gopal, Z. H. Zhou, C. M. Knobler, and W. M. Gelbart, "Visualizing large rna molecules in solution," *RNA*, vol. 18, p. 284, 2012.

[31] R. F. Garmann, A. Gopal, S. S. Athavale, C. M. Knobler, W. M. Gelbart, and S. C. Harvey, "Visualizing the global secondary structure of a viral rna genome with cryo-electron microscopy," *RNA*, vol. 21, p. 877, 2015.

[32] P.-G. de Gennes, "Some conformation problems for long macromolecules," *Rep. Prog. Phys.*, vol. 32, p. 187, 1969.

[33] P.-G. de Gennes, "Polymers at an interface. 1. Adsorption and depletion layers," *Macromolecules*, vol. 14, p. 1637, 1981.

[34] P.-G. de Gennes, "Polymers at an interface. 2. Interaction between 2 plates carrying adsorbed polymer layers," *Macromolecules*, vol. 15, p. 492, 1982.

[35] P.-G. de Gennes and P. Pincus, "Scaling theory of polymer adsorption : proximal exponent," *J. Physique Lett*, vol. 44, p. L241, 1983.

[36] I. S. Jones and P. Richmond, "Effects of Excluded Volume on the Conformation of Adsorbed Polymers," *J. Chem. Soc., Faraday Trans. II*, vol. 73, p. 1062, 1977.

[37] G. J. Fleer and J. Scheutjens, "Adsorption of interacting oligomers and polymers at an interface," *Adv. Colloid Interface Sci.*, vol. 16, p. 341, 1982.

[38] P. A. Pincus, C. J. Sandroff, and T. A. Witten, "Polymer adsorption on colloidal particles," *J. Physique*, vol. 45, p. 725, 1984.

[39] R. Ober, L. Paz, C. Taupin, P. Pincus, and S. Boileau, "Study of the Surface Tension of Polymer Solutions: Theory and Experiments. 1. Good Solvent Conditions ," *Macromolecules*, vol. 16, p. 50, 1983.

[40] E. Eisenriegler, K. Kremer, and K. Binder, "Adsorption of polymer chains at surfaces: Scaling and Monte Carlo analyses," *J. Chem. Phys.*, vol. 77, p. 6296, 1982.

[41] J. M. di Meglio, R. Ober, L. Paz, C. Taupin, P. Pincus, and S. Boileau, "Study of the surface tension of polymer solutions: theory and experiments in theta solvent conditions ," *J. Physique*, vol. 44, p. 1035, 1983.

[42] D. Hone, H. Ji, and P. A. Pincus, "Polymer adsorption on rough surfaces. 1. ideal long chain," *Macromolecules*, vol. 20, p. 2543, 1987.

[43] R. G. Winkler and A. G. Cherstvy, "Strong and Weak Polyelectrolyte Adsorption onto Oppositely Charged Curved Surfaces," *Adv Polym Sci.*, vol. 255, p. 1, 2014.

[44] M. A. Carignano and I. Szleifer, "Structure and thermodynamics of grafted three-arm branched polymer layers," *Macromolecules*, vol. 27, p. 702, 1994.

[45] J. Sun, C. DuFort, M.-C. Daniel, A. Murali, C. Chen, K. Gopinath, B. Stein, M. De, V. M. Rotello, A. Holzenburg, C. C. Kao, and B. Dragnea, "Core-controlled polymorphism in virus-like particles," *PNAS.*, vol. 104, p. 1354, 2007.

[46] R. K. DeLong, C. M. Reynolds, Y. Malcolm, A. Schaeffer, T. Severs, and A. Wanekaya, "Functionalized gold nanoparticles for the binding, stabilization, and delivery of therapeutic dna, rna, and other biological macromolecules.," *Nanotechnol. Sci. Appl.*, vol. 3, p. 53, 2010.

[47] B. J. Johnson, B. J. Melde, M. A. Dinderman, and B. Lin, "Stabilization of rna through absorption by functionalized mesoporous silicate nanospheres," *PLOS ONE*, vol. 7, p. e50356, 2012.

[48] L. H. Rothberg, "Detection of specific sequences in rna using defferential adsorption of a single-stranded oligonucleotides on gold nanoparticles," *Anal Chem*, vol. 77, p. 6229, 2005.

[49] C. Chen, Z. Daniel, M.-C.and Quinkert, M. De, B. Stein, V. Bowman, P. Chipman, V. Rotello, C. C. Kao, and B. Dragnea, "Nanoparticle-Templated Assembly of Viral Protein Cages," *Nano Lett.*, vol. 6, p. 611, 2006.

[50] A. Siber, R. Zandi, and R. Podgornik, "Thermodynamics of nanospheres encapsulated in virus capsids," *Phys. Rev. E*, vol. 81, p. 051919, 2010.

[51] Y. Ding, Z. Jiang, K. Saha, C. Kim, S. Kim, R. Landis, and V. M. Rotello, "Gold nanoparticles for nucleic acid delivery," *Molecular Therapy*, vol. 22, p. 1075, 2014.

[52] W. Cai, T. Gao, H. Hong, and J. Sun, "Applications in cancer nanotechnology," *Nanotechnol Sci Appl*, vol. 1, p. 17, 2008.

[53] P. M. Tiwari, V. Komal, V. A. Dennis, and R. S. Singh, "Functionalized Gold Nanoparticles and Their Biomedical Applications," *Nanomaterials*, vol. 1, p. 31, 2011.

[54] A. Y. Grosberg, "Disordered polymers," *Physics-Uspekhi*, vol. 40, p. 125, 1997.

[55] Y. Hu, R. Zandi, A. Anavitarte, C. M. Knobler, and W. M. Gelbart, "Packaging of a polymer by a viral capsid: the interplay between polymer length and capsid size," *Biophys. J.*, vol. 94, p. 1428, 2008.

[56] R. D. Cadena-Nava, Y. F. Hu, R. F. Garmann, B. Ng, A. N. Zelikin, C. M. Knobler, and W. M. Gelbart, "Exploiting Fluorescent Polymers To Probe the Self-Assembly of Virus-like Particles," *J. Phys. Chem. B*, vol. 115, p. 2386, 2011.

[57] G. Gaspari and J. Rudnick, "n-vector model in the limit $n \to 0$ and the statistics of linear polymer systems: A ginsburg-landau theory," *Phys. Rev. B*, vol. 33, p. 3295, 1986.

[58] J. F. Joanny, "Polyelectrolyte adsorption and charge inversion," *Eur. Phys. J. B*, vol. 9, p. 117, 1999.

[59] I. Borukhov, D. Andelman, and H. Orland, "Random polyelectrolytes and polyampholytes in solution," *Euro. Phys. J. B*, vol. 5, p. 869, 1998.

[60] H. Ji and D. Hone, "Polymer Adsorption on Rough Surfaces. 2. Good Solvent Conditions," *Macromolecules*, vol. 21, p. 2600, 1988.

[61] R. Zandi and P. van der Schoot, "Size Regulation of ss-RNA Viruses," *Biophys. J.*, vol. 96, p. 9, 2009.

[62] P. van der Schoot and R. Zandi, "Kinetic Theory of Virus Capsid Assembly," *Phys. Biol.*, vol. 4, p. 296, 2007.

[63] A. Siber and R. Podgornik, "Nonspecific Interactions in Spontaneous Assembly of Empty Versus Functional Single-Stranded RNA Viruses," *Phys. Rev. E*, vol. 78, p. 051915, 2008.

[64] R. Zandi, J. Rudnick, and R. Golestanian, "Probing polyelectrolyte elasticity using radial distribution function," *Phys. Rev. E*, vol. 67, p. 021803, Feb 2003.

[65] R. Zandi and J. Rudnick, "Constraints, histones, and the 30-nm spiral," *Phys. Rev. E*, vol. 64, p. 051918, Oct 2001.

[66] P.-G. de Gennes, "Exponents for the excluded volume problem as derived by the wilson method," *Phys. Lett. A*, vol. 38, p. 339, 1972.

[67] T. C. Lubensky and J. Isaacson, "Statistic of Lattice Animals and Dilute Branched Polymers," *Phys. Rev. A*, vol. 20, p. 2130, 1979.

[68] K. Elleuch, F. Lequeux, and P. Pfeuty, "Crosslink Effects on Equilibrium Polymers," *J. Phys. I France*, vol. 5, p. 465, 1995.

[69] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics*, vol. 73 of *International Series of Monographs on Physics*. Oxford: Oxford Science Publications, 1986.

[70] S. F. Edwards, "The statistical mechanics of polymers with excluded volume," *Proc. Phys. Soc.*, vol. 85, p. 613, 1965.

[71] P. van der Schoot and R. Bruinsma, "Electrostatics and the Assembly of an RNA Virus," *Phys. Rev. E*, vol. 71, p. 061928, 2005.

[72] A. L. Bozic, A. Siber, and R. Podgornik, "How simple can a model of an empty viral capsid be? Charge distributions in viral capsids," *J. Biol. Phys.*, vol. 38, p. 657, 2012.

[73] G. Erdemci-Tandogan, J. Wagner, P. van der Schoot, R. Podgornik, and R. Zandi, "Rna topology remolds electrostatic stabilization of viruses," *Phys. Rev. E*, vol. 89, p. 032707, 2014.

[74] A. G. Zilman and S. A. Safran, "Thermodynamics and structure of self-assembled networks," *Phys. Rev. E*, vol. 66, p. 051107, 2002.

[75] J. B. Bancroft, "The self-assembly of spherical plant viruses," *Adv. Virus Res.*, vol. 16, p. 99, 1970.

[76] F. D. Sikkema, M. Comellas-Aragones, R. G. Fokkink, B. J. M. Verduin, J. Cornelissen, and R. J. M. Nolte, "Monodisperse polymer-virus hybrid nanoparticles," *Org. Biomol. Chem.*, vol. 5, p. 54, 2007.

[77] Y. P. Ren, S. M. Wong, and L. Y. Lim, "In vitro-reassembled plant virus-like particles for loading of polyacids," *J. Gen. Virol.*, vol. 87, p. 2749, 2006.

[78] P. Ni, Z. Wang, X. Ma, N. Das, P. Sokol, W. Chiu, B. Dragnea, M. Hagan, and C. C. Kao, "An examination of the electrostatic interactions between the n-terminal tail of the brome mosaic virus coat protein and encapsidated rnas," *J. Mol. Biol*, vol. 419, p. 284, 2012.

[79] A. Losdorfer Bozic, A. Siber, and R. Podgornik, "Statistical analysis of sizes and shapes of virus capsids and their resulting elastic properties," *J. Biol. Phys.*, vol. 39, p. 215, 2013.

[80] M. C. Daniel, I. B. Tsvetkova, Z. T. Quinkert, A. Murali, M. De, V. M. Rotello, C. C. Kao, and B. Dragnea, "Role of Surface Charge Density in Nanoparticle-Templated Assembly of Bromovirus Protein Cages," *ACS Nano*, vol. 4, p. 3853, 2010.

[81] A. Zlotnick, R. Aldrich, J. M. Johnson, P. Ceres, and M. J. Young, "Mechanism of Capsid Assembly for an Icosahedral Plant Virus," *Virology*, vol. 277, p. 450, 2000.

[82] H. K. Lin, P. van der Schoot, and R. Zandi, "Impact of charge variation on the encapsulation of nanoparticles by virus coat proteins," *Phys. Biol.*, vol. 9, p. 066004, 2012.

[83] A. M. Yoffe, P. Prinsen, A. Gopal, C. M. Knobler, W. M. Gelbart, and A. Ben-Shaul, "Predicting the sizes of large RNA molecules," *PNAS*, vol. 105, p. 16153, 2008.

[84] F. Li Tai, W. M. Gelbart, and A. Ben-Shaul, "The size of RNA as an ideal branched polymer," *J. Chem. Phys.*, vol. 135, p. 155105, 2011.

[85] H. Tao, Z. Rui, and B. I. Shklovskii, "Electrostatic theory of viral self-assembly," *Physica A*, vol. 387, p. 3059, 2008.

[86] V. A. Belyi and M. Muthukumar, "Electrostatic origin of the genome packing in viruses," *PNAS*, vol. 103, p. 17174, 2006.

[87] C. L. Ting, J. Z. Wu, and Z. G. Wang, "Thermodynamic basis for the genome to capsid charge relationship in viral encapsidation," *PNAS*, vol. 108, p. 16986, 2011.

[88] O. M. Elrad and M. F. Hagan, "Encapsulation of a Polymer by an Icosahedral Virus," *Phys. Biol.*, vol. 7, p. 045003, 2010.

[89] M. F. Hagan, "Modeling Viral Capsid Assembly," *Adv. Chem. Phys.*, vol. 155, p. 1, 2014.

[90] A. Borodavka, R. Tuma, and P. G. Stockley, "Evidence that viral RNAs have evolved for efficient, two-stage packaging," *PNAS*, vol. 109, p. 15769, 2012.

[91] P. G. Stockley, R. Twarock, S. E. Bakker, A. M. Barker, A. Borodavka, E. Dykeman, R. J. Ford, A. R. Pearson, S. E. V. Phillips, N. A. Ranson, and R. Tuma, "Packaging signals in single-stranded RNA viruses: natures alternative to a purely electrostatic assembly mechanism," *J. Biol. Phys.*, vol. 39, p. 277, 2013.

[92] A. McPherson, "Micelle formation and crystallization as paradigms for virus assembly," *BioEssays*, vol. 27, p. 447, 2005.

[93] I. Borukhov, D. Andelman, and H. Orland, "Polyelectrolyte Solutions between Charged Surfaces," *Europhys. Lett.*, vol. 32, p. 499, 1995.

[94] I. Borukhov and D. Andelman, "Scaling Laws of Polyelectrolyte Adsorption," *Macromolecules*, vol. 31, p. 1665, 1998.

[95] I. Borukhov, D. Andelman, and H. Orland, "Effect of Polyelectrolyte Adsorption on Intercolloidal Forces," *J. Phys. Chem. B*, vol. 103, p. 5042, 1999.

[96] A. Shafir, D. Andelman, and R. R. Netz, "Adsorption and depletion of polyelectrolytes from charged surfaces," *J. Chem. Phys.*, vol. 119, p. 2355, 2003.

[97] T. C. Lubensky and J. Isaacson, "Statistic of Lattice Animals and Dilute Branched Polymers," *Phys. Rev. A*, vol. 20, p. 2130, 1979.

[98] S. I. Lee and T. T. Nguyen, "Radial Distribution of RNA Genomes Packaged inside Spherical Viruses," *Phys. Rev. Lett.*, vol. 100, p. 198102, 2008.

[99] P.-G. de Gennes, "Polymers a t an Interface. 2. Interaction between Two Plates Carrying Adsorbed Polymer Layers," *Macromolecules*, vol. 15, p. 492, 1982.

[100] I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer, M. Tacker, and P. Schuster, "Fast Folding and Comparison of RNA Secondary Structures," *Monatsh. Chem.*, vol. 125, p. 167, 1994.

[101] The standard deviation for the number of branch points in the ensemble of RNA secondary structures is around 2.5 for the RNA1 sequence of both BMV and CCMV.

[102] M. Rossmann and V. Rao, *Viral molecular machines*, vol. 726. Springer Science & Business Media, 2011.

[103] A. Siber, A. L. Bozic, and R. Podgornik, "Energies and pressures in viruses: contribution of nonspecific electrostatic interactions," *Phys. Chem. Chem. Phys.*, vol. 14, p. 3746, 2012.

[104] R. Kusters, H.-K. Lin, R. Zandi, I. Tsvetkova, B. Dragnea, and P. van der Schoot, "Role of charge regulation and size polydispersity in nanoparticle encapsulation by viral coat proteins," *J. Phys. Chem. B*, vol. 119, p. 1869, 2015.

[105] V. Sivanandam, D. Mathwes, R. Garmann, G. Erdemci-Tandogan, R. Zandi, and A. Rao, "Functional analysis of the n-terminal basic motif of a eukaryotic satellite rna virus capsid protein in replication and packaging," *Scientific Reports*, vol. 6, p. 26328, 2016.

[106] P. Stockley, R. Twarock, S. Bakker, A. Barker, A. Borodavka, E. Dykeman, R. Ford, A. Pearson, S. Phillips, and N. Ranson, "Packaging signals in single-stranded RNA viruses: natures alternative to a purely electrostatic assembly mechanism," *J. Biol. Phys.*, vol. 39, p. 277, 2013.

[107] N. Patel, E. Dykeman, R. Coutts, G. Lomonossoff, D. Rowlands, S. Phillips, N. Ranson, R. Twarock, R. Tuma, and P. Stockley, "Revealing the density of encoded functions in a viral RNA," *Proc. Natl. Acad. Sci.*, vol. 112, p. 2227, 2015.

[108] R. Zandi, P. van der Schoot, D. Reguera, W. Kegel, and H. Reiss, "Classical Nucleation Theory of Virus Capsids," *Biophys. J.*, vol. 90, p. 1939, 2006.

[109] G. Erdemci-Tandogan, J. Wagner, P. van der Schoot, and R. Zandi, "Role of genome in the formation of conical retroviral shells," *J. Phys. Chem. B*, vol. 120, p. 6298, 2016.

[110] L. Tubiana, A. L. Bozic, C. Micheletti, and R. Podgornik, "Synonymous Mutations Reduce Genome Compactness in Icosahedral ssRNA Viruses," *Biophysical J.*, vol. 108, p. 194, 2015.

[111] S. W. Singaram, R. F. Garmann, C. M. Knobler, W. M. Gelbart, and A. Ben-Shaul, "Role of rna branchedness in the competition for viral capsid proteins," *J. Phys. Chem. B.*, vol. 119, p. 13991, 2015.

[112] M. F. Hagan and R. Zandi, "Recent advances in coarse-grained modeling of virus assembly," *Curr. Opin. Virol.*, vol. 18, p. 36, 2016.

[113] W. H. Roos, I. L. Ivanovska, A. Evilevitch, and G. J. Wuite, "Viral capsids: Mechanical characteristics, genome packaging and delivery mechanisms," *Cell Mol Life Sci.*, vol. 64, p. 1484, 2007.

[114] L. Javidpour, A. L. Boi, A. Naji, and R. Podgornik, "Multivalent ion effects on electrostatic stability of virus-like nano-shells," *J. Chem. Phys.*, vol. 139, p. 154709, 2013.

[115] Y. Andoh, N. Yoshii, A. Yamada, K. Fujimoto, H. Kojima, K. Mizutani, A. Nakagawa, A. Nomoto, and S. Okazaki, "All-atom molecular dynamics calculation study of entire poliovirus empty capsids in solution," *J. Chem. Phys.*, vol. 141, p. 165101, 2014.

[116] A. Naji, M. Kanduˇ c, J. Forsman, and R. Podgornik, "Perspective: Coulomb fluids – weak coupling, strong coupling, in between and beyond," *J. Chem. Phys.*, vol. 139, p. 150901, 2013.

[117] J. Wagner, G. Erdemci-Tandogan, and R. Zandi, "Adsorption of annealed branched polymers on curved surfaces," *J. Phys.:Condens. Matter*, vol. 27, p. 495101, 2015.

[118] L. Lavelle, M. Gingery, M. Phillips, W. M. Gelbart, C. M. Knobler, R. D. Cadena-Nava, J. R. Vega-Acosta, L. A. Pinedo-Torres, and J. Ruiz-Garcia, "Phase Diagram of Self-assembled Viral Capsid Protein Polymorphs," *J. Phys. Chem. B*, vol. 113, p. 3813, 2009.

[119] M. Zuker, "Mfold web server for nucleic acid folding and hybridization prediction," *Nucleic Acids Res.*, vol. 31, p. 3406, 2003.

[120] www.ncbi.nlm.nih.gov.

[121] A. G. Moreira and R. R. Netz, *Electrostatic Effects in Soft Matter and Biophysics, Chapter: Field-Theoretic Approaches to Classical Charged Systems*, vol. 46 of *NATO Science Series*. Springer Netherlands, 2001.

[122] R. A. Valverde and J. A. Dodds, "Evidence for a satellite rna associated naturally with the u5 strain and experimentally with the u1 strain of tobacco mosaic virus," *J. Gen. Virol.*, vol. 67, p. 1875, 1986.

[123] T. E. Mirkov, D. M. Mathews, D. H. Du Plessis, and J. A. Dodds, "Evidence for a satellite rna associated naturally with the u5 strain and experimentally with the u1 strain of tobacco mosaic virus," *Virology*, vol. 170, p. 139, 1989.

[124] R. A. Valverde and J. A. Dodds, "Some properties of isometric virus particles which contain the satellite rna of tobacco mosaic virus," *J. Gen. Virol.*, vol. 68, p. 965, 1987.

[125] S. B. Larson, S. Koszelak, J. Day, A. Greenwood, J. A. Dodds, and A. McPherson, "Double-helical rna in satellite tobacco mosaic virus," *Nature*, vol. 361, p. 179, 1993.

[126] S. B. Larson and A. McPherson, "Satellite tobacco mosaic virus rna: structure and implications for assembly," *Curr. Opin. Struct. Biol.*, vol. 11, p. 59, 2001.

[127] G. Routh, M. N. Yassi, A. L. Rao, T. E. Mirkov, and J. A. Dodds, "Replication of wild-type and mutant clones of satellite tobacco mosaic virus in nicotiana benthamiana protoplasts," *J. Gen. Virol.*, vol. 78, p. 1271, 1997.

[128] G. Routh, J. A. Dodds, L. Fitzmaurice, and T. E. Mirkov, "Characterization of deletion and frameshi mutants of satellite tobacco mosaic virus," *Virology*, vol. 212, p. 121, 1995.

[129] V. Sivanandam, D. Mathews, and A. L. Rao, "Properties of satellite tobacco mosaic virus phenotypes expressed in the presence and absence of helper virus," *Virology*, vol. 483, p. 163, 2015.

[130] S. B. Larson, J. Day, A. Greenwood, and A. McPherson, "Refined structure of satellite tobacco mosaic virus at 1.8a resolution," *J. Mol. Biol.*, vol. 277, p. 37, 1998.

[131] S. Zuniga, I. Sola, J. L. Moreno, P. Sabella, J. Plana-Duran, and L. Enjuanes, "Coronavirus nucleocapsid protein is an rna chaperone," *Virology*, vol. 357, p. 215, 2007.

[132] P. Ni and C. Cheng Kao, "Non-encapsidation activities of the capsid proteins of positive-strand rna viruses," *Virology*, vol. 446, p. 123, 2013.

[133] L. M. Guogas, D. J. Filman, J. M. Hogle, and L. Gehrke, "Cofolding organizes alfalfa mosaic virus rna and coat protein for replication," *Science*, vol. 306, p. 2108, 2004.

[134] V. L. Reichert, M. Choi, J. E. Petrillo, and L. Gehrke, "Alfalfa mosaic virus coat protein bridges rna and rna-dependent rna polymerase in vitro," *Virology*, vol. 364, p. 214, 2007.

[135] S. C. Keane and D. P. Giedroc, "Solution structure of mouse hepatitis virus (mhv) nsp3a and determinants of the interaction with mhv nucleocapsid (n) protein," *J. Virol.*, vol. 87, p. 3502, 2013.

[136] M. Sakata, N. Otsuki, K. Okamoto, M. Anraku, M. Nagai, M. Takeda, and Y. Mori, "Short self-interacting n-terminal region of rubella virus capsid protein is essential for cooperative actions of capsid and nonstructural p150 proteins," *J. Virol.*, vol. 88, p. 11187, 2014.

[137] K. K. Sharma, H. de Rocquigny, J. L. Darlix, J. P. Lavergne, F. Penin, J. M. Lessinger, and Y. Mely, "Analysis of the rna chaperoning activity of the hepatitis c virus core protein on the conserved 3x region of the viral genome," *Nucleic Acids Res.*, vol. 40, p. 2540, 2012.

[138] H. Wu, M. Mitra, M. N. Naufer, M. J. McCauley, R. J. Gorelick, I. Rouzina, K. Musier-Forsyth, and M. C. Williams, "Differential contribution of basic residues to hiv-1 nucleocapsid proteins nucleic acid chaperone function and retroviral replication," *Nucleic Acids Res.*, vol. 42, p. 2525, 2014.

[139] V. Yusibov and L. S. Loesch-Fries, "Functional significance of three basic n-terminal amino acids of alfalfa mosaic virus coat protein," *Virology*, vol. 242, p. 1, 1998.

[140] R. C. Olsthoorn, S. Mertens, F. T. Brederode, and J. F. Bol, "A conformational switch at the 3 end of a plant virus rna regulates viral replication," *EMBO J*, vol. 18, p. 4856, 1999.

[141] S. C. Chen and R. C. Olsthoorn, "In vitro and in vivo studies of the rna conformational switch in alfalfa mosaic virus," *J. Virol.*, vol. 84, p. 1423, 2010.

[142] M. F. Hagan, "A theory for viral capsid assembly around electrostatic cores," *J. Chem. Phys.*, vol. 130, p. 114902, 2009.

[143] J. Sun, C. DuFort, M. C. Daniel, A. Murali, C. Chen, K. Gopinath, B. Stein, M. De, V. M. Rotello, A. Holzenburg, C. C. Kao, and B. Dragnea, "Core-controlled polymorphism in virus-like particles," *PNAS*, vol. 104, p. 1354, 2007.

[144] J. K. Seo, S. J. Kwon, and A. L. Rao, "A physical interaction between viral replicase and capsid protein is required for genome-packaging specificity in an rna virus," *J. Virol.*, vol. 86, p. 6210, 2012.

[145] P. Annamalai and A. L. Rao, "Packaging of brome mosaic virus subgenomic rna is functionally coupled to replication-dependent transcription and translation of coat protein," *J. Virol.*, vol. 80, p. 10096, 2006.

[146] P. Annamalai and A. L. Rao, "Replication-independent expression of genome components and capsid protein of brome mosaic virus in planta: a functional role for viral replicase in rna packaging," *Virology*, vol. 338, p. 96, 2005.

[147] A. L. Rao, T. W. Dreher, L. E. Marsh, and T. C. Hall, "Telomeric function of the trna-like structure of brome mosaic virus rna," *PNAS*, vol. 86, p. 5335, 1989.

[148] K. Bracha-Drori, K. Shichrur, A. Katz, M. Oliva, R. Angelovici, S. Yalovsky, and N. Ohad, "Detection of protein-protein interactions in plants using bimolecular fluorescence complementation," *Plant J.*, vol. 40, p. 419, 2004.

[149] L. Lakatos, G. Szittya, D. Silhavy, and J. Burgyan, "Molecular mechanism of rna silencing suppression mediated by p19 protein of tombusviruses," *EMBO J.*, vol. 23, p. 876, 2004.

[150] A. C.S. and E. O. Freed, "Novel approaches to inhibiting hiv-1 replication," *Antiviral Res.*, vol. 85, p. 119, 2010.

[151] E. O. Freed, "Hiv-1 assembly, release and maturation," *Nat. Rev. Microbiol.*, vol. 13, p. 484, 2015.

[152] L. S. Ehrlich, B. E. Agresta, and C. A. Carter, "Assembly of recombinant human immunodeficiency virus type 1 capsid protein in vitro," *J. Virol.*, vol. 66, p. 4874, 1992.

[153] S. Campbell and V. M. Vogt, "Self-assembly in vitro of purified ca-nc proteins from rous sarcoma virus and human immunodeficiency virus type 1," *J. Virol.*, vol. 69, p. 6487, 1995.

[154] I. Gross, H. Hohenberg, and H.-G. Krausslich, "In vitro assembly properties of purified bacterially expressed capsid proteins of human immunodeficiency virus," *Eur. J. Biochem.*, vol. 249, p. 592, 1997.

[155] I. Gross, H. Hohenberg, C. Huckhagel, and H.-G. Krausslich, "N-terminal extension of human immunodeficiency virus capsid protein converts the in vitro assembly phenotype from tubular to spherical particles," *J. Virol.*, vol. 72, p. 4798, 1998.

[156] U. K. von Schwedler, T. L. Stemmler, V. Y. Klishko, S. Li, K. H. Albertine, D. R. Davis, and W. I. Sundquist, "Proteolytic refolding of the hiv-1 capsid protein amino-terminus facilitates viral core assembly," *EMBO J.*, vol. 17, p. 1555, 1998.

[157] M. Grattinger, H. Hohenberg, D. Thomas, T. Wilk, B. Muller, and H. G. Krausslich, "In vitro assembly properties of wild-type and cyclophilin-binding defective human immunodeficiency virus capsid proteins in the presence and absence of cyclophilin a," *Virology*, vol. 257, p. 247, 1999.

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

[159] I. J. Byeon, X. Meng, J. Jung, G. Zhao, Y. R, J. Ahn, J. Shi, J. Concel, C. Aiken, P. Zhang, and A. M. Gronenborn, "Structural convergence between cryo-em and nmr reveals intersubunit interactions critical for hiv-1 capsid function," *Cell*, vol. 139, p. 780, 2009.

[160] G. Zhao, J. R. Perilla, E. L. Yufenyuy, X. Meng, B. Chen, J. Ning, J. Ahn, A. M. Gronenborn, K. Schulten, C. Aiken, and P. Zhang, "Mature hiv-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics," *Nature*, vol. 497, p. 643, 2013.

[161] A. de Marco, A.-M. Heuser, B. Glass, H.-G. Krausslich, B. Muller, and J. A. G. Briggs, "Role of the sp2 domain and its proteolytic cleavage in hiv-1 structural maturation and infectivity," *J. Virol.*, vol. 86, p. 13708, 2012.

[162] T. T. Nguyen, R. F. Bruinsma, and W. M. Gelbart, "Elasticity Theory and Shape Transitions of Viral Shells," *Phys. Rev. E*, vol. 72, p. 051923, 2005.

[163] J. Fontana, K. A. Jurado, N. Cheng, N. L. Ly, J. R. Fuchs, R. J. Gorelick, A. N. Engelman, and A. C. Steven, "Distribution and redistribution of hiv-1 nucleocapsid protein in immature, mature, and integrase-inhibited virions: A role for integrase in maturation," *J. Virol.*, vol. 89, p. 9765, 2015.

[164] J. A. Briggs, T. Wilk, R. Welker, H.-G. Krausslich, and S. D. Fuller, "Structural organization of authentic, mature hiv-1 virions and cores," *EMBO J.*, vol. 22, p. 1707, 2003.

[165] V. A. Belyi and M. Muthukumar, "Electrostatic Origin of the Genome Packing in Viruses," *PNAS*, vol. 103, p. 17174, 2006.

[166] A. Šiber and R. Podgornik, "Role of electrostatic interactions in the assembly of empty spherical viral capsids," *Phys. Rev. E*, vol. 76, p. 061906, 2007.

[167] W. S. Klug, R. F. Bruinsma, J.-P. Michel, and C. M. Knobler, "Failure of viral shells," *Phys. Rev. Lett.*, vol. 97, p. 228101, 2006.

[168] E. C. Dykeman, P. G. Stockley, and R. Twarock, "Building a viral capsid in the presence of genomic rna," *Phys. Rev. E*, vol. 87, p. 022717, 2013.

[169] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics*. New York: Oxford Science Publications, 1986.

[170] P.-G. de Gennes, *Scaling Concepts in Polymer Physics*. New York: Cornell University Press, 1979.

[171] N. Gershenfeld, *The Nature of Mathematical Modeling*. New York: Cambridge University Press, 1999.

[172] R. Welker, H. Hohenberg, U. Tessmer, C. Huckhagel, and H.-G. Krausslich, "Biochemical and structural analysis of isolated mature cores of human immunodeficiency virus type 1," *J. Virol.*, vol. 74, p. 1168, 2000.

[173] C. J. Petropoulos, *Retroviruses: Appendix 2: Retroviral Taxonomy, Protein Structure, Sequences, and Genetic Maps*. New York: Cold Spring Harbor Laboratory Press, 1997.

[174] A. Levandovsky and R. Zandi, "Nonequilibirum assembly, retroviruses, and conical structures," *PRL*, vol. 102, p. 198102, 2009.

[175] J. Wagner and R. Zandi, "The robust assembly of small symmetric nanoshells," *Biophys. J.*, vol. 109, p. 956, 2015.

[176] Z. Yu, M. J. Dobro, C. L. Woodward, A. Levandovsky, C. M. Danielson, V. Sandrin, J. Shi, C. Aiken, R. Zandi, and G. J. Jensen, "Unclosed hiv-1 capsids suggest a curled sheet model of assembly," *J Mol Biol.*, vol. 425, p. 112, 2013.

[177] N. Yu and M. F. Hagan, "Simulations of hiv capsid protein dimerization reveal the effect of chemistry and topography on the mechanism of hydrophobic protein association," *Biophys. J.*, vol. 103, p. 1363, 2012.

[178] E. Freed, "Hiv-1 gag proteins: diverse functions in the virus life cycle," *Virology*, vol. 251, p. 1, 1998.

[179] L. Ko, F. Yu, K. Huang, and C. Wang, "Hiv-1 matrix domain removal ameliorates virus assembly and processing defects incurred by positive nucleocapsid charge elimination," *FEBS Open Bio*, vol. 5, p. 283, 2015.

[180] T. A. Bharat, N. E. Davey, P. Ulbrich, J. D. Riches, A. de Marco, M. Rumlova, C. Sachse, T. Ruml, and J. A. Briggs, "Structure of the immature retroviral capsid at 8 a resolution by cryo-electron microscopy," *Nature*, vol. 487, p. 385, 2012.

[181] F. K. Schur, W. J. Hagen, M. Rumlova, T. Ruml, B. Muller, H. G. Krausslich, and J. A. Briggs, "Structure of the immature hiv-1 capsid in intact virus particles at 8.8 a resolution," *Nature*, vol. 517, p. 505, 2015.

[182] O. Pornillos, B. K. Ganser-Pornillos, and M. Yeager, "Atomic-level modelling of the hiv capsid," *Nature*, vol. 469, p. 424, 2011.

[183] B. Chen and R. Tycko, "Simulated self-assembly of the hiv-1 capsid: protein shape and native contacts are sufficient for two-dimensional lattice formation," *Biophys J.*, vol. 100, p. 3035, 2011.

[184] J. M. A. Grime, J. F. Dama, B. K. Ganser-Pornillos, C. L. Woodward, G. J. Jensen, M. Yeager, and G. A. Voth, "Coarse-grained simulation reveals key features of hiv-1 capsid self-assembly," *Nat. Commu.*, vol. 7, p. 11568, 2016.

[185] J. M. Grime and G. A. Voth, "Early stages of the hiv-1 capsid protein lattice formation," *Biophys. J.*, vol. 103, p. 1774, 2012.

[186] X. Qiao, J. Jeon, J. Weber, F. Zhu, and B. Chen, "Mechanism of polymorphism and curvature of hiv capsid assemblies probed by 3d simulations with a novel coarse grain model," *Biochim. Biophys. Acta.*, vol. 1850, p. 2353, 2015.

[187] K. Wiegers, G. Rutter, H. Kottler, U. Tessmer, H. Hohenberg, and H. G. Krausslich, "Sequential steps in human immunodeficiency virus particle maturation revealed by alterations of individual gag polyprotein cleavage sites," *J. Virol.*, vol. 72, p. 2846, 1998.

[188] S. K. Lee, J. Harris, and R. Swanstrom, "A strongly transdominant mutation in the human immunodeficiency virus type 1 gag gene defines an achilles heel in the virus life cycle," *J. Virol.*, vol. 83, p. 8536, 2009.

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

[190] S. C. Pettit, M. D. Moody, R. S. Wehbie, A. H. Kaplan, P. V. Nantermet, C. A. Klein, and R. Swanstrom, "The p2 domain of human immunodeficiency virus type 1 gag regulates sequential proteolytic processing and is required to produce fully infectious virions," *Biochim. Biophys. Acta.*, vol. 68, p. 8017, 1994.

[191] C. S. Adamson, M. Sakalian, K. Salzwedel, and E. Freed, "Polymorphisms in gag spacer peptide 1 confer varying levels of resistance to the hiv- 1 maturation inhibitor bevirimat," *Retrovirology*, vol. 7, p. 36, 2010.

[192] C. S. Adamson, K. Waki, S. D. Ablan, K. Salzwedel, and E. Freed, "Impact of human immunodeficiency virus type 1 resistance to protease inhibitors on evolution of resistance to the maturation inhibitor bevirimat (pa-457)," *J. Virol.*, vol. 83, p. 4884, 2009.

[193] P. W. Keller, C. S. Adamson, J. B. Heymann, E. O. Freed, and A. Steven, "Hiv-1 maturation inhibitor bevirimat stabilizes the immature gag lattice," *J. Virol.*, vol. 85, p. 1420, 2011.

[194] K. Waki, S. R. Durell, F. Soheilian, K. Nagashima, S. L. Butler, and E. O. Freed, "Structural and functional insights into the hiv-1 maturation inhibitor binding pocket," *PLoS Pathog.*, vol. 8, p. e1002997, 2012.

[195] A. de Marco, B. Muller, B. Glass, J. D. Riches, H. G. Krausslich, and J. A. Briggs, "Structural analysis of hiv-1 maturation using cryo-electron tomography," *PLoS Pathog.*, vol. 6, p. e1001215, 2010.

[196] J. A. G. Briggs, J. D. Riches, B. Glass, V. Bartonova, G. Zanetti, and H.-G. Krausslich, "Structure and assembly of immature hiv," *PNAS*, vol. 106, p. 11090, 2009.

[197] E. R. Wright, J. B. Schooler, H. J. Ding, C. Kieffer, C. Fillmore, W. I. Sundquist, and G. J. Jensen, "Electron cryotomography of immature hiv-1 virions reveals the structure of the ca and sp1 gag shell," *EMBO J.*, vol. 26, p. 2218, 2007.

[198] J. A. Briggs and H. G. Krausslich, "The molecular architecture of hiv," *J. Mol. Biol.*, vol. 410, p. 491, 2011.

[199] J. Zhou, L. Huang, D. L. Hachey, C. H. Chen, and C. Aiken, "Inhibition of hiv-1 maturation via drug association with the viral gag protein in immature hiv-1 particles," *J. Biol. Chem.*, vol. 280, p. 42149, 2005.

[200] F. Li, R. Goila-Gaur, K. Salzwedel, N. R. Kilgore, M. Reddick, C. Matallana, A. Castillo, D. Zoumplis, D. E. Martin, J. M. Orenstein, G. P. Allaway, E. O. Freed, and C. T. Wild, "Hiv inhibitor that disrupts core condensation by targeting a late step in gag processing," *PNAS*, vol. 100, p. 13555, 2003.

[201] J. Zhou, X. Yuan, D. Dismuke, B. M. Forshey, C. Lundquist, K. H. Lee, C. Aiken, and C. H. Chen, "Small-molecule inhibition of human immunodeficiency virus type 1 replication by specific targeting of the final step of virion maturation," *J. Virol.*, vol. 78, p. 922, 2004.

[202] C. K. Cassidy, B. A. Himes, F. J. Alvarez, J. Ma, G. Zhao, J. R. Perilla, K. Schulten, and P. Zhang, "Cryoem and computer simulations reveal a novel kinase conformational switch in bacterial chemotaxis signaling," *Elife*, vol. 4, p. e08419, 2015.

[203] X. Meng, G. Zhao, and P. Zhang, "Structure of hiv-1 capsid assemblies by cryo-electron microscopy and iterative helical real-space reconstruction," *J. Vis. Exp.*, vol. 54, p. 3041, 2011.

[204] H. G. Krausslich, R. H. Ingraham, M. T. Skoog, E. Wimmer, P. V. Pallai, and C. A. Carter, "Activity of purified biosynthetic proteinase of human immunodeficiency virus on natural substrates and synthetic peptides," *PNAS*, vol. 86, p. 807, 1989.

[205] V. L. Morton, E. C. Dykeman, N. J. Stonehouse, A. E. Ashcroft, R. Twarock, and P. G. Stockley, "The impact of viral rna on assembly pathway selection," *J. Mol. Biol.*, vol. 401, p. 298, 2010.

[206] S. Mattei, A. Flemming, M. Anders-Osswein, H. G. Krausslich, J. A. Briggs, and B. Muller, "Rna and nucleocapsid are dispensable for mature hiv-1 capsid assembly," *J. Virol.*, vol. 89, p. 9739, 2015.

[207] J. Fontana, K. A. Jurado, N. Cheng, N. L. Ly, J. R. Fuchs, R. J. Gorelick, A. N. Engelman, and A. C. Steven, "Distribution and redistribution of hiv-1 nucleocapsid protein in immature, mature, and integrase-inhibited virions: a role for integrase in maturation," *J. Virol.*, vol. 89, p. 9765, 2015.

[208] K. A. Jurado, H. Wang, A. Slaughter, L. Feng, J. J. Kessl, Y. Koh, W. Wang, A. Ballandras-Colas, P. A. Patel, J. R. Fuchs, M. Kvaratskhelia, and A. Engelman, "Allosteric integrase inhibitor potency is determined through the inhibition of hiv-1 particle maturation," *PNAS*, vol. 110, p. 8690, 2013.

[209] M. A. Checkley, B. G. Luttge, F. Soheilian, K. Nagashima, and E. Freed, "The capsid-spacer peptide 1 gag processing intermediate is a dominant-negative inhibitor of hiv-1 maturation," *Virology*, vol. 400, p. 137, 2010.

[210] B. Muller, M. Anders, and J. Reinstein, "In vitro analysis of human immunodeficiency virus particle dissociation: gag proteolytic processing influences dissociation kinetics," *PLoS One*, vol. 9, p. e99504, 2014.

[211] S. Erickson-Viitanen, J. Manfredi, P. Viitanen, D. E. Tribe, R. Tritch, C. A. r. Hutchison, L. D. D., and R. Swanstrom, "Cleavage of hiv-1 gag polyprotein synthesized in vitro: Sequential cleavage by the viral protease," *AIDS Research and Human Retroviruses*, vol. 5, p. 577, 1989.

[212] D. J. Wyma, A. Kotov, and C. Aiken, "Evidence for a stable interaction of gp41 with pr55(gag) in immature human immunodeficiency virus type 1 particles," *J. Virol.*, vol. 74, p. 9381, 2000.

[213] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, "Ucsf chimera–a visualization system for exploratory research and analysis," *J. Comput. Chem.*, vol. 25, p. 1605, 2004.

[214] J. R. Kremer, D. N. Mastronarde, and J. McIntosh, "Computer visualization of three-dimensional image data using imod," *J. Struct. Biol.*, vol. 116, p. 71, 1996.

[215] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*. New York: Cambridge University Press, 1992.