

UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

Physical Principles of Virus Capsid Assembly and Genome Packaging

Permalink

<https://escholarship.org/uc/item/82k1x3m5>

Author

Dong, Yinan

Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Physical Principles of Virus Capsid Assembly and Genome Packaging

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

Doctor of Philosophy

in

Physics

by

Yinan Dong

September 2023

Dissertation Committee:

Dr. Roya Zandi, Chairperson

Dr. John P. Barton

Dr. Thomas E. Kuhlman

Copyright by
Yinan Dong
2023

The Dissertation of Yinan Dong is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I would like to extend my heartfelt gratitude to my advisor, Dr. Roya Zandi, for her unwavering support and invaluable guidance throughout my PhD journey. Her mentorship has played a crucial role in shaping both my research and personal growth. Without her mentorship, I would not have reached this point. Her genuine concern extends beyond my research, encompassing my daily lives as well. I deeply appreciate having her as my advisor.

I am also deeply thankful to Prof. Alex Travasset, whose excellent guidance and valuable suggestions significantly contributed to the success of my projects. His mentorship has been invaluable, and I have gained extensive knowledge from his expertise.

My sincere appreciation goes to my committee members, Dr. John P. Barton and Dr. Thomas E. Kuhlman, for their expertise and valuable feedback during my oral exam and final defense.

Special thanks go to Siyu Li, whose assistance has greatly accelerated the progress of my research. I have learned a wealth of knowledge and techniques from working with her.

Furthermore, I am grateful to my labmates, Sanaz Panahandeh, Alireza Ramezani, Amin Safdari, for their friendship and constructive discussions, which enriched my research experience.

In times of pandemic, I am especially thankful to my roommates, Yunxiao Li, Mengyuan Xi, Yawei Qin, and Nan Hua, whose companionship and support helped me navigate through challenging times.

Lastly, my deepest appreciation goes to my parents and family for their unwavering support and understanding throughout this entire journey. Their encouragement has been a constant source of strength for me.

Chapter 2 is mainly a version of Y. Dong, S. Li, and R. Zandi, Effect of the Charge Distribution of Virus Coat Proteins on the Length of Packaged RNAs, *Phys Rev E* 102, 062423 (2020).

Chapter 3 is a version of Y. Dong, R. Zandi, and A. Travasset, Exact Solution for Elastic Networks on Curved Surfaces, *Phys. Rev. Lett.* 129, 088001 (2022).

A small portion of chapter 2 is a version of L. Marichal, L. Gargowitsch, R. L. Rubim, C. Sizun, K. Kra, S. Bressanelli, Y. Dong, S. Panahandeh, R. Zandi, and G. Tresset, Relationships between RNA Topology and Nucleocapsid Structure in a Model Icosahedral Virus, *Biophys. J.* 120, 3925 (2021).

To my parents for all the support.

ABSTRACT OF THE DISSERTATION

Physical Principles of Virus Capsid Assembly and Genome Packaging

by

Yinan Dong

Doctor of Philosophy, Graduate Program in Physics
University of California, Riverside, September 2023
Dr. Roya Zandi, Chairperson

Single-stranded RNA viruses efficiently encapsulate their genome into a protein shell called the capsid. Understanding the physical principles underlying the formation of virus capsid assembly and genome packaging is of great interest because of their potential applications in blocking viral infections and various areas of bio-nanotechnology, such as drug delivery and gene therapy.

The first part of the thesis investigates the encapsidation of single-stranded RNAs into virus capsid. Electrostatic interactions between the positive charges in the capsid protein's N-terminal tail and the negatively charged genome have been postulated as the main driving force for virus assembly. Recent experimental results indicate that the N-terminal tails with the same number of charges and same lengths package different amounts of RNA, which reveals that electrostatics alone cannot explain all the observed outcomes of the RNA self-assembly experiments. Using a mean-field theory, we show that the combined effect of genome configurational entropy and electrostatic interaction can explain to some extent the amount of packaged RNA with mutant proteins where the location and number

of charges on the tails are altered and shed light on many experimental results relevant to BMV assembly.

The second part focuses on understanding the physical principles of virus capsid assembly, specifically for the difficult cases of the assembly of nonspherical structures such as Human Immunodeficiency Viruses (HIV). For HIV shells, while there are often 5 defects at the smaller and 7 at the larger caps, defect positions vary from one HIV structure to another. Currently, there is no clear understanding of what determines the position of the defects as the surfaces with non-zero Gaussian curvature such as the conical shell of HIV grow. To tackle this issue, in this thesis, we take the first step and solve an intermediate problem of characterizing the structure of an elastic network constrained to lie on a frozen curved surface by continuum elasticity theory. We provide an exact solution to this problem without resorting to any approximation in terms of geometric quantities.

Contents

List of Figures	xi
List of Tables	xv
1 Introduction	1
1.1 Structures of viruses	1
1.2 Virus self-assembly	4
1.3 Overview	5
2 Effect of the charge distribution of virus coat proteins on the length of packaged RNAs and RNA secondary structure on genome packaging	7
2.1 Abstract	7
2.2 Introduction	8
2.3 Methods	13
2.3.1 N-terminal tails	17
2.4 Results	19
2.4.1 A capsid with 60 tails ($T=1$)	19
2.4.2 A capsid with 180 tails ($T=3$)	23
2.4.3 Relationships between RNA topology and nucleocapsid structure in a model icosahedral virus	29
2.5 Discussion	31
3 Exact Solution for elastic networks on curved surfaces	36
3.1 Abstract	36
3.2 Introduction	37
3.3 Methods	40
3.3.1 Summary of the methods	40
3.3.2 Explicit formulas for the different quantities	46
3.3.3 Particularization to surface revolution	47
3.3.4 Boundary condition	50
3.3.5 Free energy normalization	52
3.3.6 About units	53

3.3.7	Connection with linear elasticity theory	54
3.3.8	Theory of defects, inverse Laplacian square	63
3.4	Results	64
3.5	Discussion	65
4	Conclusions	75

List of Figures

1.1	Cryo microscopy images of (a) tobacco mosaic virus (TMV) [55] (b) cowpea chlorotic mottlevirus [5] (c) human immunodeficiency viruses (HIV) [57] (d) bacteriophage P22 [145]	3
2.1	(a) A $T = 3$ icosahedral shell with 180 protein subunits. The darker (blue) color shows the pentamers. The structure is similar to the BMV capsid. (b) The interior of a $T = 3$ viral shell with N-terminal domains (pink tails) extended towards the center of the capsid. Each N-terminal domain contains eight positive charges, not shown in the figure. The structure in (a) is reproduced using UCSF Chimera packages (http://www.rbvi.ucsf.edu/chimera).	10
2.2	(a) Schematic of the sequences of N-terminal tails of six mutants used in the experiments of Ni <i>et al.</i> [99]. The mutants are denoted by 2HA ₇ , 2H ₇ , 2HA ₁₅ , 2H ₁₅ , 4S and 4R. The triangles denote the location of the insertions. For 2HA ₇ and 2HA ₁₅ , eight neutral amino acids are inserted into the N-terminal. For 2H ₇ and 2H ₁₅ , four neutral and four positive amino acids (with boldface and underlined) are inserted. The four positive amino acids are two lysines (K) and two arginines (R), leading to the increased length of N-terminal regions and also 720 additional positive charges per capsid. For 4S and 4R, the length of N-terminals remains the same. In the case of 4S four neutral amino acids (MAAA) are replaced with another four neutral amino acids and for 4R mutants, four neutral amino acids (MAAA) are replaced with four positively charged arginines (R). (b) Spectroscopic analysis of the number of nucleotides per virion.	12
2.3	(a) The white circles indicate the locations of N-terminals on a $T = 3$ capsid. (b) 3D view of inside of a $T = 3$ capsid with 180 protruded regions representing N-terminals. There are eight positive charges on each cylinder (N-terminal tail) in a wild-type BMV capsid. The positive charges are not shown in the figure.	18

2.4	Genome density profile inside a $T = 1$ capsid as a function of the distance from the capsid center. The solid lines in the figure show the profiles along N-terminal tails, but the dashed graphs correspond to the direction without N-terminal tails (inset). (a) The plot illustrates the profile when the distance between the two charges is 0.2 nm with the total number of monomers $N = 658$. (b) The plot corresponds to the profile when the distance between the two charges is 1.4 nm with $N = 680$. See Fig. 2.5(a) for a schematic view of charge distributions. The length of the tail is 4 nm , and the size of each charged region is 0.2 nm . The polymers are branched with $f_b = 3.86$. The other parameters are salt concentration $\mu = 500 \text{ mM}$, the capsid radius $R = 9 \text{ nm}$, and the total charge on N-terminals $Q_c = 120$	19
2.5	(a) Schematic of an N-terminal tail. The distance between two positive charges along the N-terminal domain increases from bottom to top. Each yellow rectangle is 0.2 nm and denotes one positive charged amino acid. The smallest distance between the two charges is 0.2 nm . From the shortest to the longest distance, we examine seven different cases. The largest distance between the two charges is 2.8 nm . The charge on the right side is next to the wall and its position is fixed. (b) Optimal length of RNA encapsulated as a function of the distance between two charges for a capsid with radius $R = 9 \text{ nm}$, the tail length 4 nm and salt concentration $\mu = 500 \text{ mM}$. RNA is modeled as an annealed branched polymer.	22
2.6	Encapsulation free energy as a function of monomer number, N . The dashed line corresponds to the case in which the distance between the two charges is short ($d = 0.2 \text{ nm}$) and the solid curve to when the distance between the charges is a little bit longer ($d = 1.4 \text{ nm}$), See Fig. 2.5(a) for a schematic of two charge distributions. The other parameters are the capsid radius $R = 9 \text{ nm}$, the tail length 4 nm , salt concentration $\mu = 500 \text{ mM}$ and the total positive charge on the capsid is $Q_c = 120$. RNA is modeled as an annealed branched polymer and its fugacity is $f_b = 3.86$. The optimal number of packaged monomers for $d = 0.2 \text{ nm}$ is 658 while for $d = 1.4 \text{ nm}$ is 680.	23
2.7	Table of seven charge distributions along N-terminals where each yellow rectangle represents a positively charged amino acid and blue triangles neutral ones. The table includes the optimal encapsulation free energy of the RNA confined into a spherical shell, the optimal length of encapsulated RNA, percent change (theory) of optimal length compared to the wild-type BMV and the percent change (experiment) from Fig. 2.2. The salt concentration is 500 mM . The radius of the capsid is 12 nm . For wild-type, 2HA ₇ , 2HA ₁₅ and 4S, the total charge on capsid is $Q_c = 1440$ but for 2H ₇ , 2H ₁₅ and 4R is $Q_c = 2160$. The tail length for wild-type, 4S, and 4R are 5 nm while for 2HA ₇ , 2HA ₁₅ , 2H ₇ , and 2H ₁₅ are 6.5 nm . The Debye length λ_D for 500 mM is 0.438 nm	24

2.8	The genome density profile vs. r , the distance from the center of the capsid for four different charge distributions along the N-terminal domains: (a) wild-type (WT), (b) 2HA ₁₅ (M1), (c) 2HA ₁₅ (M2) and (d) 2HA ₁₅ . The first column in Figs. 2.7 and 2.9 show the schematics of N-terminal tails for each case. The peaks in the RNA profiles correspond to the position of positive charges along the N-terminal tails. As the distance between the charges located in the middle of N-terminal tails increases, the density of genome between the two peaks goes lower. However, the amount of RNA between two peaks due to the entropic contribution and the range of electrostatic interaction does not drop to zero in case of 2HA ₁₅ (M1) (b). The genome density between the two peaks becomes smaller for 2HA ₁₅ (M2) (c) and becomes almost zero for 2HA ₁₅	25
2.9	Table of six different charge distributions along N-terminals. As before each yellow rectangle represents an amino acid with a positive charge and blue rectangles represent neutral amino acids. The table includes the optimal length of encapsulated RNA for three different salt concentrations, 500 <i>mM</i> , 300 <i>mM</i> and 100 <i>mM</i> . The distance between the fourth positive charge (the fourth yellow rectangle) and the fifth positive charge from top to bottom is 0.2 <i>nm</i> , 0.6 <i>nm</i> , 1.0 <i>nm</i> , 1.4 <i>nm</i> , 1.8 <i>nm</i> and 2.2 <i>nm</i> . The percent change (theory) of the optimal length of encapsulated RNA for each mutant relative to the RNA encapsulated by mutant 2HA ₇ (M) is also presented in the table. The capsid radius is 12 <i>nm</i> and the tail length is 6.5 <i>nm</i> with total charges on the capsid $Q_c = 1440$. Debye length is $\lambda_D = 0.979$ <i>nm</i> for $\mu = 100$ <i>mM</i> , $\lambda_D = 0.565$ <i>nm</i> for $\mu = 300$ <i>mM</i> and $\lambda_D = 0.438$ <i>nm</i> for 500 <i>mM</i>	26
2.10	The encapsidation free energy as a function of monomer numbers for the mutants presented in Fig. 2.7. 2HA ₇ and 4S have the same free energy as wild-type. The additional length inserted in 2HA ₇ does not have a huge impact on the optimal encapsidated genome because it does not modify the distance between the charges along the N-terminal tails, see Fig. 2.7. The capsid radius is 12 <i>nm</i> and the tail length is 6.5 <i>nm</i> with total charges on the capsid $Q_c = 1440$. The salt concentration is 500 <i>mM</i>	27
2.11	Schematic of two N-terminal tails with different charge distributions. As before each yellow rectangle represents an amino acid with a positive charge and effective size of $d = 0.2$ <i>nm</i> and blue rectangles represent neutral amino acids but with the same size. The table includes the effective encapsulation free energy of the RNA confined into a spherical shell, the optimal length of encapsulated RNA, and the percent change (theory) of optimal length of packaged RNA with respect to the first charge distribution. The salt concentration is $\mu = 500$ <i>mM</i> , and tail length is 4.5 <i>nm</i> for both cases. The total charge on the capsid is $Q_c = 1440$. When the charges are distributed more evenly, the optimal length of the encapsidated genome increases. For the first line of the table, the distance between yellow rectangles is either zero or 0.2 <i>nm</i> , while for the second one is either 0.2 <i>nm</i> or 0.4 <i>nm</i>	29

2.12	Branch density profile versus r , the distance from the center of the capsid. The dashed black line and red solid line correspond to the branch density profiles for C2 and RF2, respectively. The dot-dashed blue line (inset graph) is the nucleotide density profile for C2 and RF2, which is almost the same for both. The capsid radius is $R = 12$ nm, N-terminal tails length 4 nm, and the salt concentration 100 mM. The total number of positive charges on the capsid interior is 1800.	31
3.1	(a) 3D cryo-EM reconstruction of HSV [152]; (b) EM image of HIV surrounded with the lipid envelope [57]. The result of the computer simulations of (c) an HSV [83] and (d) an HIV with genome and membrane [100]. (e) The reference space consists of a flat surface without a disclination or (g) with a disclination at the center. The actual space is any manifold endowed with its natural metric. As an example, we consider (f) a spheroid($\beta = 1$) and (h) a sombrero($\beta = 3$).	38
3.2	Lattice reconstruction for both the spheroid and sombrero. The lengths R_0 , ρ_0 and r_0 are given in units of the lattice constant a_L . The plots under spheroid and sombrero surfaces indicate the corresponding mapping function from the reference to actual space. $\frac{\rho_0}{R_0} = 0.7$ for each plot. The number in the parenthesis in each figure denote $(\beta, q_i, \nu_p, \hat{\tau})$	45
3.3	Free energy for a spheroid and sombrero without ($q_i = 0$) and with ($q_i = 1$) a disclination at the center at $\frac{\tau}{YR_0} \equiv \hat{\tau} = 0$ (zero line tension) and at fixed Poisson ratio $\nu_p = 0.3$. The solid line corresponds to the exact results while the dashed line denotes the analytical results within linear elasticity. The three different colors represent different values of $\beta = 1, 2, 3$, indicating the magnitude of $f(r)$	68
3.4	Free energy (Eq. 3.1) for both the spheroid and sombrero at finite line tension $\frac{\tau}{YR_0} = \hat{\tau} = 0.05$. The solid line corresponds to the exact results while the dashed line denotes the analytical results within linear elasticity.	69
3.5	The elastic energy (see Eq. 3.1) for the spheroid at $\frac{\tau}{YR_0} = \hat{\tau} = 0.05$ for two different values of the Poisson ratio $\nu_p = 0.3$ and 0.8 . The solid lines correspond to the exact results while the dashed lines denote analytical results within linear elasticity.	70
3.6	Free energy Eq. 1 for a spheroid and sombrero at $\hat{\tau} = 0.01$. The solid lines correspond to the exact results, while the dashed lines denote analytical results within linear elasticity.	71
3.7	Free energy Eq. 1 for the spheroid at different Poisson ratios $\nu_p = 0.3$ and 0.8 . The solid lines correspond to the exact results while the dashed lines denote analytical results within linear elasticity.	72
3.8	Solution Eq. B4 and Eq. 3.60 for the spheroid and sombrero relatively small ρ_0 compared with FIG. 3.9. The solid lines correspond to the exact result while the dashed lines denote analytical results within linear elasticity. . . .	73
3.9	Solution Eq. B4 and Eq. 3.60 for the spheroid and sombrero with relatively large ρ_0 compared with FIG. 3.8. The solid lines correspond to the exact results while the dashed lines denote analytical results within linear elasticity. . . .	74

List of Tables

3.1	Transition points at which the surface with ($q_i = 1$) and without ($q_i = 0$) a disclination have the same energy. Here $\theta = \frac{r}{R_0}$ and θ_{AH} are the predictions reported in Ref. [1].	64
-----	--	----

Chapter 1

Introduction

1.1 Structures of viruses

Viruses, the simplest biological organisms, primarily consist of a genome, which can be either RNA or DNA, enclosed within a protective protein capsid. Capsids act as protein shells safeguarding the genome from the harsh external environment. Some viruses, like the retrovirus HIV-1, may contain a lipid envelope that provides protection to their capsids.

The generalized life cycle of viruses includes adsorption, entry, uncoating, replication, assembly, and release. Adsorption during which the virus attaches to the host cell membrane. Different viruses employ distinct entry strategies. Most viruses without envelopes enter cells through endocytosis. Upon entry, the capsid undergoes disassembly during the uncoating process, releasing its genome. The released genome utilizes the host cell's machinery to replicate itself and produce capsid proteins (CPs). Subsequently, the newly generated CPs assemble around the genomes, forming new virions that exit the cell

to infect other cells. The assembly of CPs into the viral shell has garnered considerable attention over the past two decades, and significant progress has been made in understanding the physics underlying this process through experimental and theoretical studies [68, 147, 31, 50].

Viruses are classified based on various factors. From an architectural standpoint, viruses can exhibit cylindrical shape (e.g., tobacco mosaic virus, as shown in Fig. 1.1a), spherical shape (e.g., cowpea chlorotic mottle virus, as depicted in Fig. 1.1b), or conical shape (e.g., human immunodeficiency viruses, as seen in Fig. 1.1c). In terms of genome type, viruses are distinguished as single-stranded (ss) RNA, double-stranded (ds) RNA, single-stranded DNA (ssDNA), or double-stranded DNA (dsDNA). Other classification factors include the host organism, the utilization of reverse transcriptase during replication, and the presence of an envelope.

Most spherical viruses exhibit the remarkable property of adopting icosahedral symmetry (IO). Within the icosahedral structure, proteins are arranged in pentagonal and hexagonal units, with pentamers evenly distributed on the capsid surface. An intriguing aspect of viruses is their ability to spontaneously assemble into highly symmetric IO structures, both in vivo and, in some cases, in vitro. A significant milestone was achieved in 1967 when Bancroft et al. successfully reconstituted the first spherical virus, CCMV, in vitro. It is noteworthy that although most spherical viruses vary in size, ranging from 20nm to 100nm in diameter, they possess icosahedral symmetry. Triangulation number, calculated based on the Casper and Klug model [26], provides a method to categorize icosahedral structures with different numbers of subunits. The triangulation number T is defined as

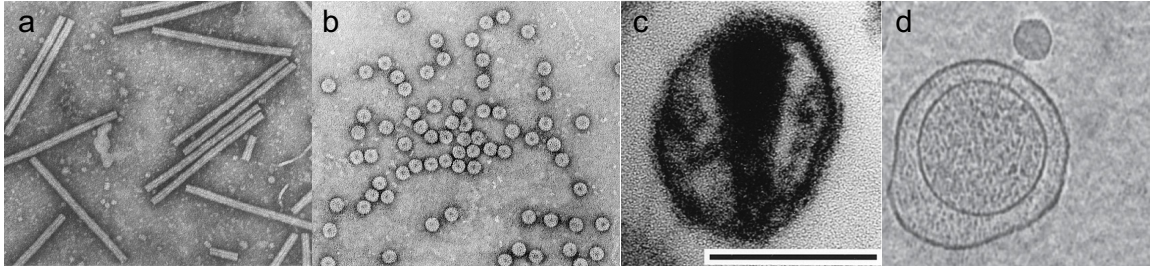


Figure 1.1: Cryo microscopy images of (a) tobacco mosaic virus (TMV) [55] (b) cowpea chlorotic mottle virus [5] (c) human immunodeficiency viruses (HIV) [57] (d) bacteriophage P22 [145]

follows:

$$T = h^2 + hk + k^2. \quad (1.1)$$

the variables h and k represent the number of steps connecting adjacent pentamers along the principal vectors of a hexagonal lattice. The total count of protein subunits in an icosahedral capsid is 60 times the triangulation number (T). Notably, regardless of the capsid's size, there are always 12 pentamers present, while the remaining protein subunits arrange themselves as hexamers. The values of h and k , being positive integers, determine specific triangulation numbers (e.g., $T = 1, 3, 4, 7$, etc.). It is noteworthy that the formation of highly symmetric structures, including icosahedral architectures, is not limited to viruses alone, as various biological structures such as clathrin vesicles also undergo self-assembly using protein subunits [32, 40].

1.2 Virus self-assembly

Numerous studies have demonstrated the spontaneous self-assembly of small single-stranded RNA (ssRNA) viruses in vitro, taking place outside of living cells in solutions containing virus coat protein subunits and the viral genome. During this process, the capsid proteins encapsulate the genome (RNA or DNA) to form a stable and protective shell. Remarkably, virus coat proteins exhibit the ability to co-assemble with various cargos, including RNAs from unrelated viruses, synthetic polyanions, and negatively charged nanoparticles [132, 76, 69]. This encapsidation property relies on the electrostatic interaction between the positively charged protein subunits and the negatively charged cargos [125, 117, 99, 89, 153, 127]. Leveraging this encapsidation feature, viruses have found diverse applications in bio-nanotechnology, such as gene therapy and drug delivery.

Both experimental and theoretical investigations have contributed to our understanding of the physical factors that govern the efficient assembly and stability of small virus particles. In-vitro experiments have elucidated the role of pH and salt concentration in the assembly of viruses like CCMV [58]. Furthermore, X-ray scattering techniques have been employed to unravel the self-assembly pathways of viral capsids encapsulating RNA genomes or other polyelectrolytes [34]. Extensive theoretical research has been conducted to unravel the mysteries surrounding the life cycle of viruses. Elasticity theory, for instance, has shed light on viruses entry into a cell and the architectural properties of viral capsids [44, 150, 149]. Polymer physics has been widely applied to understand genome configurations during the co-assembly process [109, 52, 124].

1.3 Overview

Through both experimental and theoretical investigations, our understanding of the physical mechanisms governing the virus self-assembly has significantly advanced. However, there are still many questions that need to be answered. How does the charge distribution on the N-terminal tails of capsid protein impact the optimal number of nucleotides packaged? What determines the defect positions of viruses with non-spherical structures like Human Immunodeficiency Viruses (HIV)? To answer these questions, our study employs field theory and elasticity theory to investigate the intricate process of virus assembly. Through this interdisciplinary approach, we aim to uncover key insights into the fundamental principles driving virus capsid assembly and genome packaging, advancing our knowledge for innovative strategies in viral infection prevention and gene therapy applications.

Electrostatic interactions between the positive charges in the capsid protein's N-terminal tails and the negatively charged genome have been postulated as the main driving force for virus assembly. In Chapter 2, a mean-field theory is employed to demonstrate that the combined effect of genome configurational entropy and electrostatics can explain to some extent the amount of the packaged RNA with mutant proteins, where alterations are made to the locations and number of charges on the N-terminal tails. We have been able to show that the specific locations and spacing of charges along the N-terminal tails play a crucial role in determining the amount of packaged RNA, effectively providing a qualitative explanation for the experimental findings reported by Ni *et al.* [99]. Furthermore, we demonstrate a reduction in free energy for RNA segments exhibiting a higher number of branch points.

The simplicity of virus structures reflects a fundamental connection between the elasticity theory [77] and the theory of defects [98, 29] built upon topology and differential geometry. The task of characterizing the structure of an elastic network confined to a fixed curved surface arises in various scientific disciplines and has been addressed by diverse methodologies, most notably, extending linear elasticity or through effective defect interaction models. In Chapter 3, we present a novel solution to this problem by employing exact nonlinear elasticity, eliminating the need for approximations based on geometric quantities. This allows us to consider previously challenging or unfeasible factors, such as finite line tension, explicit dependence on the Poisson ratio, and the determination of particle positions throughout the lattice. Furthermore, we discuss the implications of our findings for the characterization of virus assembly.

Chapter 2

Effect of the charge distribution of virus coat proteins on the length of packaged RNAs and RNA secondary structure on genome packaging

2.1 Abstract

Single-stranded RNA viruses efficiently encapsulate their genome into a protein shell called the capsid. Electrostatic interactions between the positive charges in the capsid protein's N-terminal tail and the negatively charged genome have been postulated as the

main driving force for virus assembly. Recent experimental results indicate that the N-terminal tail with the same number of charges and same lengths packages different amounts of RNA, which reveals that electrostatics alone cannot explain all the observed outcomes of the RNA self-assembly experiments. Using a mean-field theory, we show that the combined effect of genome configurational entropy and electrostatics can explain to some extent the amount of packaged RNA with mutant proteins where the location and number of charges on the tails are altered. We also we show that for simple icosahedral single-stranded RNA viruses, the branched topology due to the RNA secondary structure is thought to lower the free energy required to complete a virion. Understanding the factors contributing to the virus assembly could promote the attempt to block viral infections or to build capsids for gene therapy applications.

2.2 Introduction

Viruses have optimized the feat of packaging of their negatively charged genomes into a protein shell called the capsid, often built from a large number of one or a few different kinds of protein subunits [4]. Under many *in vitro* conditions, coat proteins of several single-stranded RNA (ssRNA) viruses can spontaneously encapsulate all types of anionic cargos including their native genome, linear polymers, and heterologous and nonviral RNAs[36, 12, 13, 69, 100]. The capsid proteins of several RNA viruses contain an unstructured positively charged N-terminal domain that extends toward the center of the capsid and interacts with the viral genome, see Fig. 2.1[23]. Although the specific sequence of the viral RNA plays an important role in packaging [111, 131], it is now well-established that the electrostatic

interaction between N-terminal tails and RNA is the main driving force for the formation of viral particles and their stability [133, 80, 147].

Self-assembly studies of various ssRNA viruses have revealed that the amount of RNA packaged depends directly on the number of positive charges on the N-terminal tails of capsid proteins. Many experiments show that mutant virions with less positive charges on N-terminal domain encapsidate lower amounts of RNA and mutants with increased positive charges package more [127, 99]. For example, the experimental studies of Sivanandam *et al.* show that the deletions of even one single positively charged residue of the Satellite Tobacco Mosaic Virus (STMV) N-terminal domain results in the formation of virus particles with a reduced amount of viral RNAs [127]. Belyi and Muthukumar as well as Hu *et al.* [8, 134] also examined the relation between the total number of positive charges in the tails and the length of the encapsidated RNA in various viruses and found a strong relation between them.

Of particular interest is the self-assembly experiments of Ni *et al.* [99] who specifically focused on Brome Mosaic Virus (BMV) and systematically investigated the role of electrostatics on the amount of RNA packaged [99]. The N-terminal domain of BMV capsid proteins is composed of 26 residues, eight of which are positively charged. The genome of BMV consists of four RNA molecules: RNA1 (3.2 kb), RNA2 (2.9 kb), RNA3 (2.1 kb), and RNA4 (0.9 kb). While RNA3 and RNA4 co-assemble together in one capsid, RNA1 and RNA2 are each encapsidated separately. Quite interestingly, the total length of encapsidated genome is more or less the same in each capsid. The BMV capsids of these three types are virtually identical, *i.e.*, have $T = 3$ icosahedral structures consisting of 180 copies of

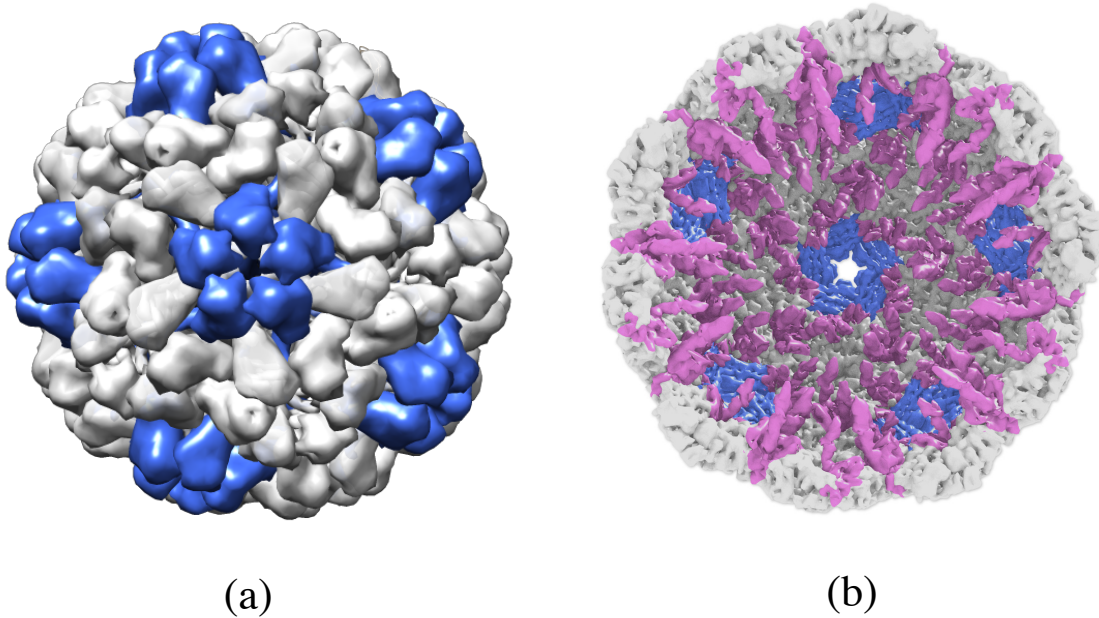


Figure 2.1: (a) A $T = 3$ icosahedral shell with 180 protein subunits. The darker (blue) color shows the pentamers. The structure is similar to the BMV capsid. (b) The interior of a $T = 3$ viral shell with N-terminal domains (pink tails) extended towards the center of the capsid. Each N-terminal domain contains eight positive charges, not shown in the figure. The structure in (a) is reproduced using UCSF Chimera packages (<http://www.rbvi.ucsf.edu/chimera>).

the same protein with the same mechanical properties [150], see Fig. 2.1. We note that the structural index T , introduced by Casper and Klug, defines the number of protein subunits in viral shells, which is 60 times the T number [26]. Thus $T = 1$ and $T = 3$ capsids have 60 and 180 protein subunits, respectively.

To gain more insight into the effect of electrostatic interactions, Ni *et al.* [99] made several mutants to increase the number of charges on N-terminal domains. A summary of their experimental results is presented in Fig. 2.2. In one case, they inserted eight

residues including four positively-charged ones after residue 15 (2H₁₅). They also examined the impact of the length of N-terminal without adding more positive charges but by introducing six alanines and two threonines, which are neutral (2HA₁₅). To examine if the position of the insertions has an impact on the amount of packaged RNA, they repeated the aforementioned experiments but introduced insertions after residue 7 and constructed 2H₇ and 2HA₇. Furthermore, to exclusively examine the effect of the increasing charges while keeping the length of the N-terminal tail the same as the wild-type one, they replaced four uncharged residues along the tail with four arginines (4R), each containing one positive charge. They found that in all cases, the structure of capsids was almost the same even though the amount of encapsidated RNA was different.

The spectroscopic analysis of the experiments of Ni *et al.* [99] reveals that as the number of charges on the N-terminal increases, the higher amount of nucleotides per capsid is packaged [99]. Nevertheless, it appears that the amount of encapsidated RNA increase does depend on other factors than the number of positive charges on the N-terminals. While the experiments clearly indicate that electrostatics plays a major role in RNA packaging, it is not obvious whether electrostatics can explain all the effects observed in Fig 2.2. Many theoretical and experimental studies have already shown that the length of packaged RNA increases with the number of charges in N-terminal tails [80, 99, 127], but how the amount of RNA encapsidated depends on the distribution and location of charges on the N-terminals have remained elusive.

The level of branching depends upon RNA sequence, and viruses have evolved to have highly branched and compact RNAs [65, 139]. Compactness has been shown to increase

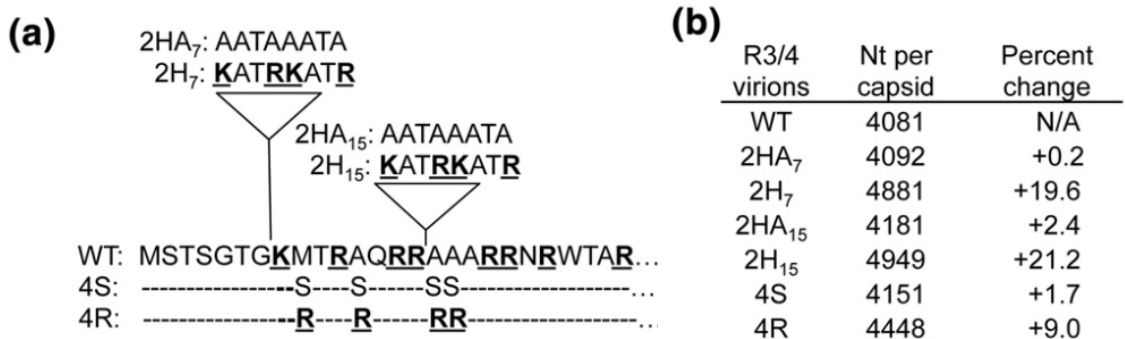


Figure 2.2: (a) Schematic of the sequences of N-terminal tails of six mutants used in the experiments of Ni *et al.*[99]. The mutants are denoted by 2HA₇, 2H₇, 2HA₁₅, 2H₁₅, 4S and 4R. The triangles denote the location of the insertions. For 2HA₇ and 2HA₁₅, eight neutral amino acids are inserted into the N-terminal. For 2H₇ and 2H₁₅, four neutral and four positive amino acids (with boldface and underlined) are inserted. The four positive amino acids are two lysines (K) and two arginines (R), leading to the increased length of N-terminal regions and also 720 additional positive charges per capsid. For 4S and 4R, the length of N-terminals remains the same. In the case of 4S four neutral amino acids (MAAA) are replaced with another four neutral amino acids and for 4R mutants, four neutral amino acids (MAAA) are replaced with four positively charged arginines (R). (b) Spectroscopic analysis of the number of nucleotides per virion.

the packaging efficiency both in *in vitro* experiments [38] and in simulation studies [126, 81]. Moreover, increased compactness seems to allow longer RNA molecules to be packaged by viral CPs [112, 53, 81]. However, the relationship between compactness and selectivity is not straightforward since it was shown in some conditions that linear RNAs could be packaged more efficiently than branched RNAs of the same length [11]. An explanation could be that a high level of branching can lead to an increased stiffness that may be detrimental for packaging selectivity [45]. The connection between RNA structure and packaging selectivity is then far from being thoroughly understood.

In this chapter we show that electrostatics is indeed able to explain at least to some extent for many observed effects relevant to RNA packaging. Using the mean-field

theory, we show that the charge discreteness, the location and the distance between the charges along the N-terminal tails have a huge impact on the optimal number of nucleotides packaged. Consistently with the experiments of Ni *et al.* [99] we find that the optimal amount of packaged RNA depends on the location of charges within the peptide sequence and increases non-linearly with the total number of positive charges on the capsid. We also show that free energy is lowered for the RNA segments displaying the larger number of branch points and the related experimental results can be found in [94].

The chapter is organized as follows. In the next section, we introduce the model and derive the equations that we will employ later. In Section III, we present our results corresponding to the non-uniform charge distribution along the N-terminal tails of BMV coat proteins and branched topology due to the RNA secondary structure. Section IV discusses the impact of the length and sequence of amino acids N-terminal tails on the the length of encapsidated genome, and finally, we present our conclusion and summarize our findings.

2.3 Methods

To explore the impact of N-terminal charge distribution on the length of packaged RNA, we model RNA as a negatively charged flexible polymer. Many experiments show that RNA acts effectively as a branched polymer in solution [64, 109]. Due to the relatively weak strength of RNA base-pairing, the number of branch points of RNA can easily be modified through the interaction with the positive charges of virus coat proteins. Thus, we focus on the case of annealed branched polyelectrolyte, which allows the degree of branching of

RNAs, a statistical quantity, to be modified [142]. Using the mean-field theory, we calculate the free energy of the RNA confined into a spherical shell that interacts attractively with the positive charges residing on the N-terminal domains of the capsid proteins. Under the ground-state dominance approximation [41, 82] where only the dominating contribution to the polymer partition function is considered, the free energy of the genome-capsid complex in a salt solution is [15, 124, 52, 54, 80],

$$\begin{aligned} \beta F = & \int d^3r \left[\frac{a^2}{6} |\nabla \Psi(\mathbf{r})|^2 + W[\Psi(\mathbf{r})] \right. \\ & \left. - \frac{\beta^2 e^2}{8\pi\lambda_B} |\nabla \Phi(\mathbf{r})|^2 - 2\mu \cosh [\beta e \Phi(\mathbf{r})] + \beta \tau \Phi(\mathbf{r}) \Psi^2(\mathbf{r}) \right] \\ & + \int d^2r [\beta \rho(\mathbf{r}) \Phi(\mathbf{r})]. \quad (2.1) \end{aligned}$$

where β is the inverse of temperature in the units of energy, a is the Kuhn length of the polymer, e is the elementary charge, μ is the density of monovalent salt ions, and τ is the linear charge density of chain. The Bjerrum length $\lambda_B = e^2\beta/4\pi\epsilon$ is about 0.7 nm for water at room temperature. The dielectric permittivity of the medium ϵ is assumed to be constant [71]. See Ref. [144] and the appendix of Ref. [54] for a step by step derivation of Eq. (2.1), in the absence and presence of electrostatic interactions, respectively.

The field $\Psi(\mathbf{r})$ is the monomer density field and $\Phi(\mathbf{r})$ is the electrostatic potential. The density of positive charges on the N-terminal tails of capsid proteins is denoted by $\rho(\mathbf{r})$. The first term in Eq. (2.1) is the entropic cost of deviation from a uniform chain density. The last two lines of Eq. (2.1) are associated with the electrostatic interactions

between the chain segments, the capsid and the salt ions at the level of Poisson-Boltzmann theory [15, 14, 122, 123]. The term $W[\Psi]$ represents the free energy density associated with the annealed branching of the polymer including the self repulsion of the polyelectrolyte [90, 79, 49],

$$W[\Psi] = -\frac{1}{\sqrt{a^3}}(f_e\Psi + \frac{a^3}{6}f_b\Psi^3) + \frac{1}{2}v\Psi^4, \quad (2.2)$$

where f_e and f_b are the fugacities of the end and branched points of the annealed polymer, respectively [144] and v is the effective excluded volume for each monomer. Note that the stem-loop or hair-pin configurations of RNA are counted as end points in this model. The quantity $\frac{1}{\sqrt{a^3}}f_e\Psi$ indicates the density of end points and $\frac{\sqrt{a^3}}{6}f_b\Psi^3$ the density of branch points. The expectation number of end and branched points, N_e and N_b , are related to the fugacities f_e and f_b , and can be written as

$$N_e = -\beta f_e \frac{\partial F}{\partial f_e} \quad \text{and} \quad N_b = -\beta f_b \frac{\partial F}{\partial f_b}. \quad (2.3)$$

There are two additional constraints in the system. The first one corresponds to the fact that the total number of monomers (Kuhn lengths) inside the capsid is fixed [42, 72],

$$N = \int d^3\mathbf{r} \Psi^2(\mathbf{r}). \quad (2.4)$$

We impose this constraint through a Lagrange multiplier, E , introduced below. Second,

there is a relation between the number of the end and branched points which should satisfy,

$$N_e = N_b + 2, \quad (2.5)$$

as there is only a single polymer in each capsid and no closed loops within the secondary structure of an RNA are allowed. The polymer is linear if $f_b = 0$, and the number of branched points increases with increasing value of f_b . For our calculations, we vary f_b and find f_e through Eq. (2.3) and Eq. (2.5). To this end, f_e is not a free parameter.

Extremizing the free energy with respect to the fields $\Psi(\mathbf{r})$ and $\Phi(\mathbf{r})$, subject to the constrain that the total number of monomers inside capsid is constant (Eq. (2.4)), we obtain three self-consistent non-linear coupled equations for the interior and exterior of the capsid,

$$\frac{a^2}{6} \nabla^2 \Psi(\mathbf{r}) = -E\Psi(\mathbf{r}) + \tau\beta\Phi_{in}(\mathbf{r})\Psi(\mathbf{r}) + \frac{1}{2} \frac{\partial W}{\partial \Psi} \quad (2.6a)$$

$$\begin{aligned} \nabla^2 \Phi_{in}(\mathbf{r}) = \frac{1}{\lambda_D^2} \sinh [\Phi_{in}(\mathbf{r})] - \frac{\tau}{2\lambda_D^2 \mu \beta \epsilon^2} \Psi^2(\mathbf{r}) \\ - \frac{1}{2\lambda_D^2 \mu \beta \epsilon^2} \rho(\mathbf{r}) \end{aligned} \quad (2.6b)$$

$$\nabla^2 \Phi_{out}(\mathbf{r}) = \frac{1}{\lambda_D^2} \sinh [\Phi_{out}(\mathbf{r})] \quad (2.6c)$$

where $\lambda_D = 1/\sqrt{8\pi\lambda_B\mu}$ is the (dimensionless) Debye screening length and E is Lagrange multiplier implementing the fixed monomer number inside capsid. The polymer concentration in the exterior of the capsid is considered to be zero, $\Psi = 0$. Equations (2.6) along with the constraints shown in Eqs. (2.4) and (2.5) 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 $\Psi(\mathbf{r})$ and $\Phi(\mathbf{r})$.

The boundary conditions for the two coupled differential Eqs. (2.6b) and (2.6c) can be obtained by minimizing the free energy with respect to $\Phi(\mathbf{r})$ field on the surface of the capsid and are,

$$\begin{aligned}\hat{n} \cdot \nabla \Phi_{in}(\mathbf{r})|_{r=R} &= \hat{n} \cdot \nabla \Phi_{out}(\mathbf{r})|_{r=R} \\ \Phi_{in}(\mathbf{r})|_{r=R} &= \Phi_{out}(\mathbf{r})|_{r=R} \\ \Phi_{out}(\mathbf{r})|_{r=\infty} &= 0.\end{aligned}\tag{2.7}$$

We employ Dirichlet boundary condition $\Psi(\mathbf{r})|_{r=R} = 0$ for the monomer density field at the capsid wall. Because of the symmetric monomer distribution, we set $\partial_r \Psi(\mathbf{r})|_{r=0} = 0$. We emphasize that the derivations of all equations given in this section can be found in the appendix of Ref. [54]. A more detailed derivation of the partition function and free energy for branched polymers can be found in Ref. [144].

2.3.1 N-terminal tails

Figure 2.1 shows a $T = 3$ structure with 180 N-terminal tails extending into the interior of capsid, distributed with icosahedral symmetry. Because of the repulsion between the positive charges residing on the N-terminal tails, and the fact that RNA wraps around them, we assume that the N-terminal tails take an extended configuration. To this end, we model the N-terminal tails of BMV capsids as solid cylinders, see Fig. 2.3(b). We note that the charged tails are placed inside the capsid, and we will use the same boundary conditions for them as those given in Eq. (2.7) at the surface.

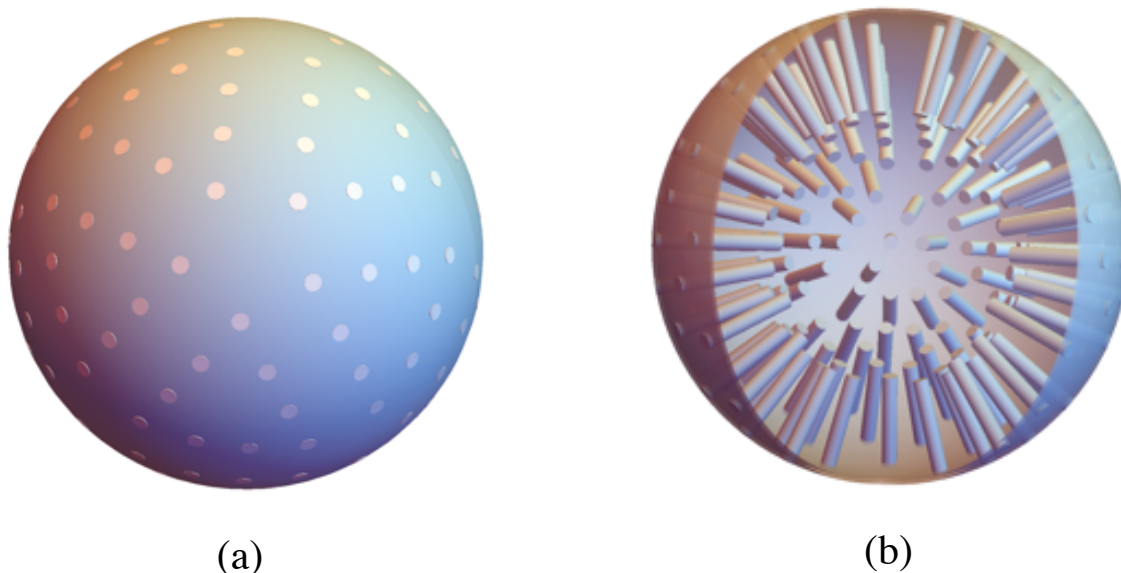


Figure 2.3: (a) The white circles indicate the locations of N-terminals on a $T = 3$ capsid. (b) 3D view of inside of a $T = 3$ capsid with 180 protruded regions representing N-terminals. There are eight positive charges on each cylinder (N-terminal tail) in a wild-type BMV capsid. The positive charges are not shown in the figure.

In the next section we will examine the impact of different charge distributions along N-terminal domains on the optimal genome length, which we will compare with the experimental results presented in Fig. 2.2. Since most of the positive charges are residing on the N-terminal tails, we consider that the charges of the coat proteins are only distributed in the cylindrical regions with no charges on the capsid wall.

For simplicity, we first consider a $T = 1$ capsid with only two positive charges on each of its 60 N-terminal tails and then focus on the $T = 3$ capsid of BMV.

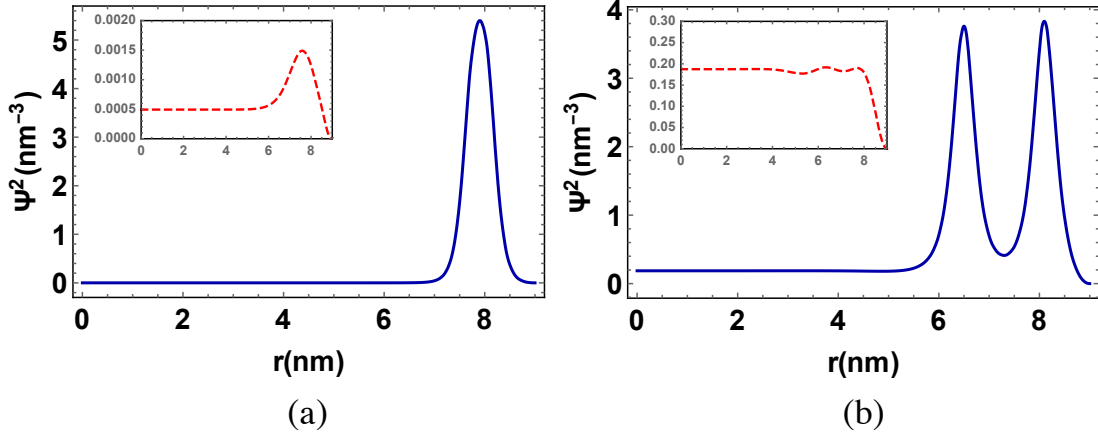


Figure 2.4: Genome density profile inside a $T = 1$ capsid as a function of the distance from the capsid center. The solid lines in the figure show the profiles along N-terminal tails, but the dashed graphs correspond to the direction without N-terminal tails (inset). (a) The plot illustrates the profile when the distance between the two charges is 0.2 nm with the total number of monomers $N = 658$. (b) The plot corresponds to the profile when the distance between the two charges is 1.4 nm with $N = 680$. See Fig. 2.5(a) for a schematic view of charge distributions. The length of the tail is 4 nm , and the size of each charged region is 0.2 nm . The polymers are branched with $f_b = 3.86$. The other parameters are salt concentration $\mu = 500 \text{ mM}$, the capsid radius $R = 9 \text{ nm}$, and the total charge on N-terminals $Qc = 120$.

2.4 Results

2.4.1 A capsid with 60 tails ($T=1$)

To obtain the optimal length of encapsidated genome in a $T = 1$ shell, we numerically solve the nonlinear coupled differential equations (2.6a), (2.6b), (2.6c), subject to the constraints given in Eqs. (2.4) and (2.5). We operate on the nonlinear coupled differential equations with finite element method and deal with the convergence issue employing the Newton method [6, 7, 101].

After finding the solutions for the fields $\Psi(\mathbf{r})$ and $\Phi(\mathbf{r})$ we insert them into Eq. (2.1) to obtain the free energy of polymer-capsid complex, F [52, 54, 51]. To obtain the encap-

solvation free energy, we need to calculate the free energy of a polymer free in solution and that of a positively charged shell and then subtract them both from the polymer-capsid complex free energy, F , given in Eq. (2.1). The capsid self-energy $[F(N=0)]$ due to the electrostatic interactions is calculated through Eqs. (2.6) and (2.7) in the limit as $N \rightarrow 0$ and should be explicitly subtracted from the polymer-capsid complex free energy, F . We emphasize that the focus here is on the solution conditions in which the capsid proteins can self-assemble in the absence of genome. We also note that previous works have shown that the free energy associated with a free chain (both linear and branched) is negligible under most experimental conditions [54, 124].

The results of our numerical calculations are given in Fig. 2.4 as a plot of the polymer concentration profile vs. r , the distance from the center of the shell for a branched polymer with the radius of capsid $R = 9 \text{ nm}$ at $\mu = 500 \text{ mM}$ salt concentrations. The total number of charges in the capsid is $Q_c = 120$ with two charges on each N-terminal tail. The length of N-terminal is 4 nm and the size of each charge is 0.2 nm (see Fig. 2.5(a)).

Figure 2.4(a) shows the genome profile if the distance between the two positive charges along the N-terminal tails is 0.2 nm while Fig. 2.4(b) corresponds to when the distance between the charges is 1.4 nm , see Fig. 2.5(a) for a schematic presentation of the distribution of charges in both cases. Note that the charged amino acids are yellow and neutral ones are blue in Fig. 2.5(a). The optimal number of monomers enclosed in the shell for Fig. 2.4(a) is $N = 658$ and for Fig. 2.4(b) is $N = 680$. The figure clearly shows that the polymer concentration is higher at the positions where the positive charges are located along the tails. When the distance between two charges is less than the Debye length

$\lambda_D = 0.438 \text{ nm}$, there is only one maximum in the profile. As the distance between the charges increases and goes beyond two Debye lengths, the genome density profile between the two charges goes almost to zero.

It is important to note that we have previously studied the impact of number of branched points, which is closely connected to the f_b value, on the length of the encapsidated genome and found that the length of genome increases with f_b [52]. Since our focus in this chapter is only on the effect of charge distribution along the N-terminals, we set $f_b = 3.86$ for all the calculations presented here in 2.4.1 and 2.4.2. In a previous chapter, we found that this value of f_b would create similar number of branch points as in the wild-type BMV genome [52]. The value of f_b does not play an important role in our findings of the effect of N-terminal charge distribution.

Figure 2.6 shows the encapsulation free energy as a function of N , the number of monomers, for a $T = 1$ structure. The dashed line in the figure corresponds to the case in which the distance between the charges is 0.2 nm and solid lines to when the distance between the charges is 1.4 nm , see Fig. 2.5(a) for a schematic of two charge distributions. As illustrated in Fig. 2.6, when the charges are closer to each other, the free energy of the system is lower; however, the minimum of the free energy moves towards longer chains as the distance between the charges increases.

Figure 2.5(b) shows the optimal length of encapsidated RNA as a function of the distance between two charges along the N-terminal domains. One charge is placed at the end of the N-terminal tail next to the capsid wall, but the location of the other varies from the wall all the way to the tip. The figure clearly shows that as the distance between charges

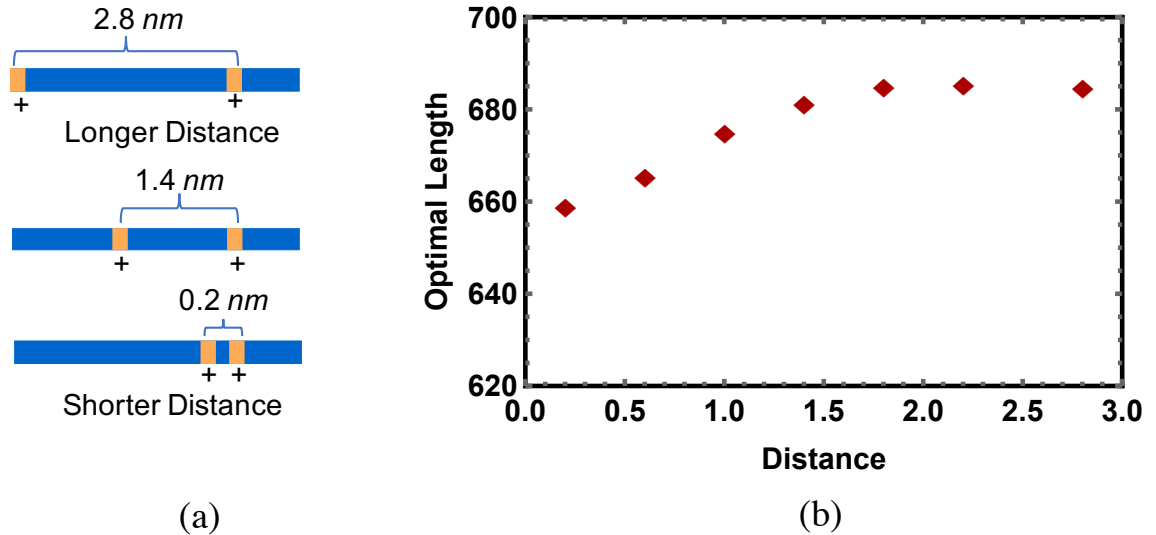


Figure 2.5: (a) Schematic of an N-terminal tail. The distance between two positive charges along the N-terminal domain increases from bottom to top. Each yellow rectangle is 0.2 nm and denotes one positive charged amino acid. The smallest distance between the two charges is 0.2 nm . From the shortest to the longest distance, we examine seven different cases. The largest distance between the two charges is 2.8 nm . The charge on the right side is next to the wall and its position is fixed. (b) Optimal length of RNA encapsulated as a function of the distance between two charges for a capsid with radius $R = 9 \text{ nm}$, the tail length 4 nm and salt concentration $\mu = 500 \text{ mM}$. RNA is modeled as an annealed branched polymer.

increases, the optimal length of the genome increases too. Thus, the location of charges along the N-terminal domains has an impact on the amount of the polymer packaged. It appears as the distance between the charges goes up, at some point the optimal length of packaged genome saturates and does not keep increasing. A careful examination of the first term in Eq. (2.1) shows that for this size of capsid and charge distribution, the optimal genome density is too small and the impact of entropy is not strong enough to have a significant role in the optimal length of genome. As the distance between the charges increases and becomes more than two Debye lengths ($\lambda_D = 0.438 \text{ nm}$ for $\mu = 500 \text{ mM}$), the

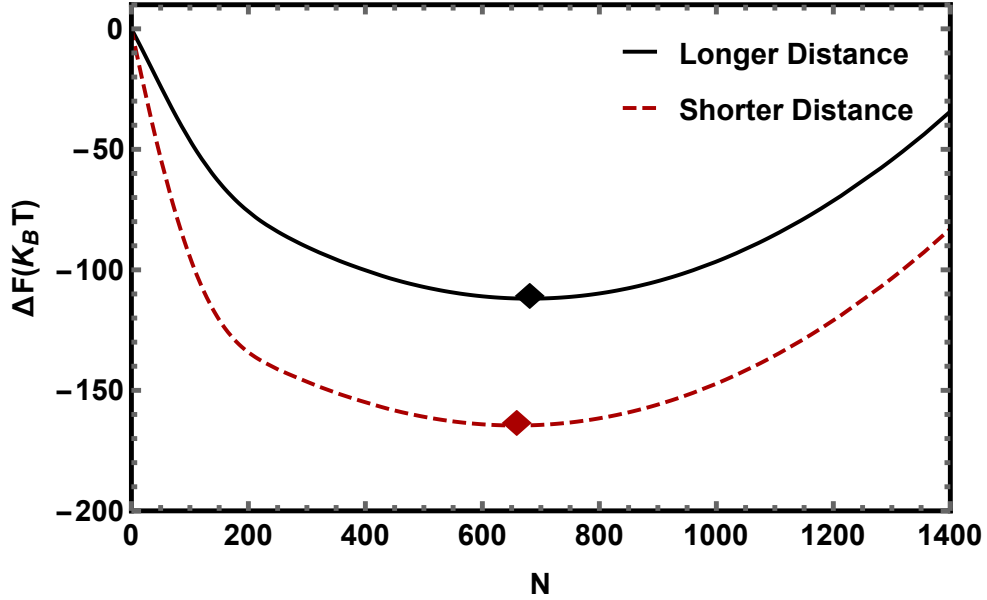


Figure 2.6: Encapsulation free energy as a function of monomer number, N . The dashed line corresponds to the case in which the distance between the two charges is short ($d = 0.2 \text{ nm}$) and the solid curve to when the distance between the charges is a little bit longer ($d = 1.4 \text{ nm}$), See Fig. 2.5(a) for a schematic of two charge distributions. The other parameters are the capsid radius $R = 9 \text{ nm}$, the tail length 4 nm , salt concentration $\mu = 500 \text{ mM}$ and the total positive charge on the capsid is $Qc = 120$. RNA is modeled as an annealed branched polymer and its fugacity is $f_b = 3.86$. The optimal number of packaged monomers for $d = 0.2 \text{ nm}$ is 658 while for $d = 1.4 \text{ nm}$ is 680.

electrostatic interaction becomes very weak between the two charges. Thus, the genome will be mostly adsorbed in the close proximity of each positive charge along the peptide. Note that even though entropy prefers a uniform genome density, the electrostatic interaction is much stronger and thus the optimal length of encapsidated genome first increases with the distance between the charges and then it remains more or less constant.

2.4.2 A capsid with 180 tails ($T=3$)

We now examine the impact of charge distribution along the N-terminal domain for a $T = 3$ capsid with 180 N-terminal tails. More specifically, we focus on the self-assembly

studies of Ni *et al.* [99] in which the impact on the length of packaged RNA of the location and distribution of positive charges along the N-terminal domains of BMV capsid proteins were studied [99]. Fig. 2.2(a) and Fig. 2.2(b) show the distribution of charges along N-terminal domains and the length of encapsulated RNA for different mutants, respectively. The schematic of the charge distribution along the N-terminals for various mutants and wild-type capsid proteins based on our model are illustrated in the left column of Fig. 2.7. The length of N-terminal is set equal to 5 *nm* for the wild-type and 6.5 *nm* for the mutants with eight extra amino acids. We assume all amino acids have the same size, which is set equal to 0.2 *nm*. The charged amino acids are yellow and neutral ones are blue as before.








Charge distribution	Virions	Effective Free Energy	Optimal Length	Percent Change (Theory)	Percent Change (Experiment)
	WT	-1935.67	3478.35	N/A	N/A
	2HA ₇	-1935.67	3478.35	0	+0.2
	2HA ₁₅	-1963.96	3479.95	+0.06	+2.4
	2H ₇	-3215.13	4547.57	+30.74	+19.6
	2H ₁₅	-3170.14	4536.67	+30.43	+21.2
	4S	-1935.67	3478.35	0	+1.7
	4R	-3720.75	4436.86	+27.56	+9.0

Figure 2.7: Table of seven charge distributions along N-terminals where each yellow rectangle represents a positively charged amino acid and blue triangles neutral ones. The table includes the optimal encapsulation free energy of the RNA confined into a spherical shell, the optimal length of encapsulated RNA, percent change (theory) of optimal length compared to the wild-type BMV and the percent change (experiment) from Fig. 2.2. The salt concentration is 500 *mM*. The radius of the capsid is 12 *nm*. For wild-type, 2HA₇, 2HA₁₅ and 4S, the total charge on capsid is $Qc = 1440$ but for 2H₇, 2H₁₅ and 4R is $Qc = 2160$. The tail length for wild-type, 4S, and 4R are 5 *nm* while for 2HA₇, 2HA₁₅, 2H₇, and 2H₁₅ are 6.5 *nm*. The Debye length λ_D for 500 *mM* is 0.438 *nm*.

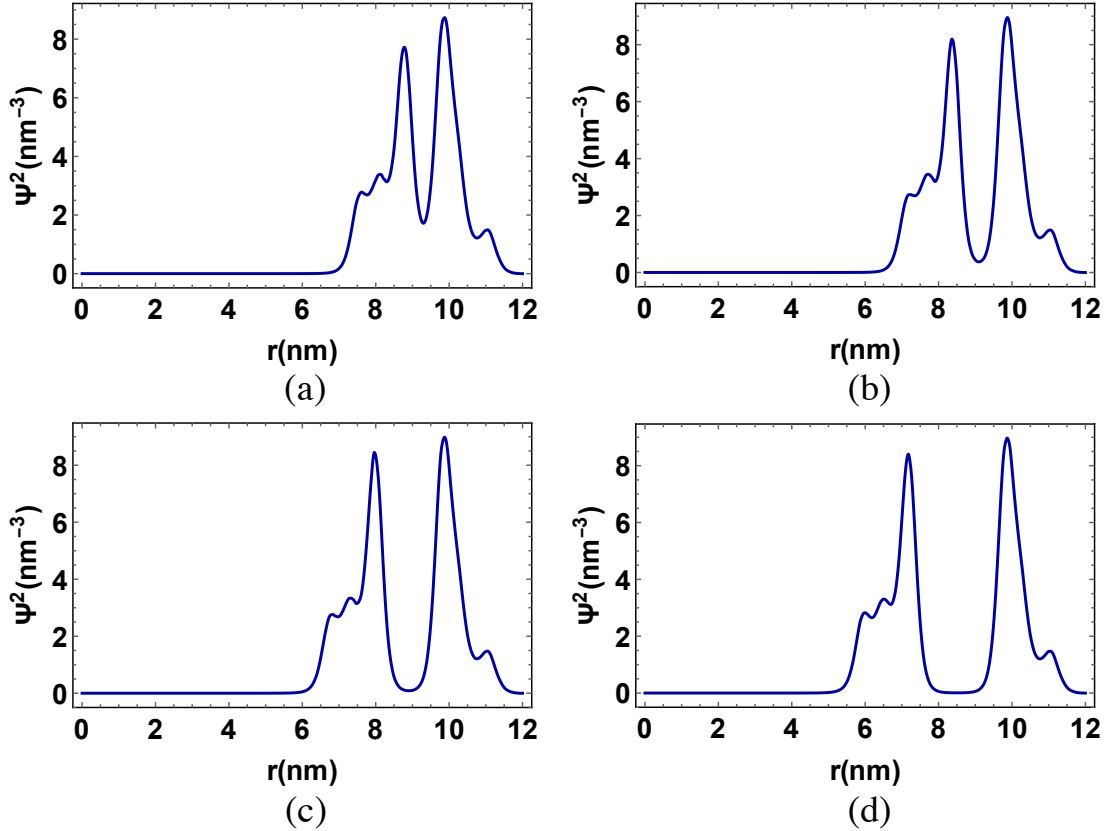


Figure 2.8: The genome density profile vs. r , the distance from the center of the capsid for four different charge distributions along the N-terminal domains: (a) wild-type (WT), (b) $2\text{HA}_{15}(\text{M1})$, (c) $2\text{HA}_{15}(\text{M2})$ and (d) 2HA_{15} . The first column in Figs. 2.7 and 2.9 show the schematics of N-terminal tails for each case. The peaks in the RNA profiles correspond to the position of positive charges along the N-terminal tails. As the distance between the charges located in the middle of N-terminal tails increases, the density of genome between the two peaks goes lower. However, the amount of RNA between two peaks due to the entropic contribution and the range of electrostatic interaction does not drop to zero in case of $2\text{HA}_{15}(\text{M1})$ (b). The genome density between the two peaks becomes smaller for $2\text{HA}_{15}(\text{M2})$ (c) and becomes almost zero for 2HA_{15} .

Following the same procedures as described above for a $T = 1$ structure, we first obtain the genome profile for a given number of nucleotides and then use it to calculate the free energy of the system. Figure 2.8 shows the genome profiles for the wild-type, 2HA₁₅ and two other mutant proteins. The schematic of charge distribution for each case is illustrated in Figs. 2.7 and 2.9. The total number of monomers in each plot in Fig. 2.8 is $N = 1390$, and the total number of charges in all capsids is $Qc = 1440$. There are eight positive charges on each N-terminal tail, whose length is 6.5 nm long for mutants and 5 nm for wild-type proteins. The genome is considered to be a branched polymer ($f_b = 3.86$).

Charge Distribution	Virions	Optimal Length (500mM)	Percent Change (500mM)	Optimal Length (300mM)	Percent Change (300mM)	Optimal Length (100mM)	Percent Change (100mM)
	2HA ₇ (M)	3431.4	N/A	2501.00	N/A	1831.2	N/A
	2HA ₇	3478.35	+1.37	2528.29	+1.09	1840.28	+0.50
	2HA ₁₅ (M1)	3522.73	+2.67	2544.39	+1.73	1847.94	+0.91
	2HA ₁₅ (M2)	3501.42	+2.04	2548.31	+1.89	1852.57	+1.17
	2HA ₁₅ (M3)	3492.47	+1.78	2541.79	+1.63	1853.44	+1.21
	2HA ₁₅	3479.95	+1.41	2529.88	+1.15	1850.17	+1.04

Figure 2.9: Table of six different charge distributions along N-terminals. As before each yellow rectangle represents an amino acid with a positive charge and blue rectangles represent neutral amino acids. The table includes the optimal length of encapsulated RNA for three different salt concentrations, 500 mM, 300 mM and 100 mM. The distance between the fourth positive charge (the fourth yellow rectangle) and the fifth positive charge from top to bottom is 0.2 nm, 0.6 nm, 1.0 nm, 1.4 nm, 1.8 nm and 2.2 nm. The percent change (theory) of the optimal length of encapsulated RNA for each mutant relative to the RNA encapsidated by mutant 2HA₇(M) is also presented in the table. The capsid radius is 12 nm and the tail length is 6.5 nm with total charges on the capsid $Qc = 1440$. Debye length is $\lambda_D = 0.979$ nm for $\mu = 100$ mM, $\lambda_D = 0.565$ nm for $\mu = 300$ mM and $\lambda_D = 0.438$ nm for 500 mM.

Figure 2.10 shows the free energy of a branched polymer packaged by the wild-type and mutant proteins of Fig. 2.7. The symbols in the figure correspond to the optimal

genome length for each case. The figure reveals that the encapsulation free energy of the wild-type, 2HA₇, 4S, and 2HA₁₅ are almost the same. Note that all these mutants have the same number of charges on their capsids. The values of the minimum free energy, the corresponding optimal genome length, and the percent change (theory and experiment) of encapsulated genome compared to the wild-type case are presented in Fig. 2.7.

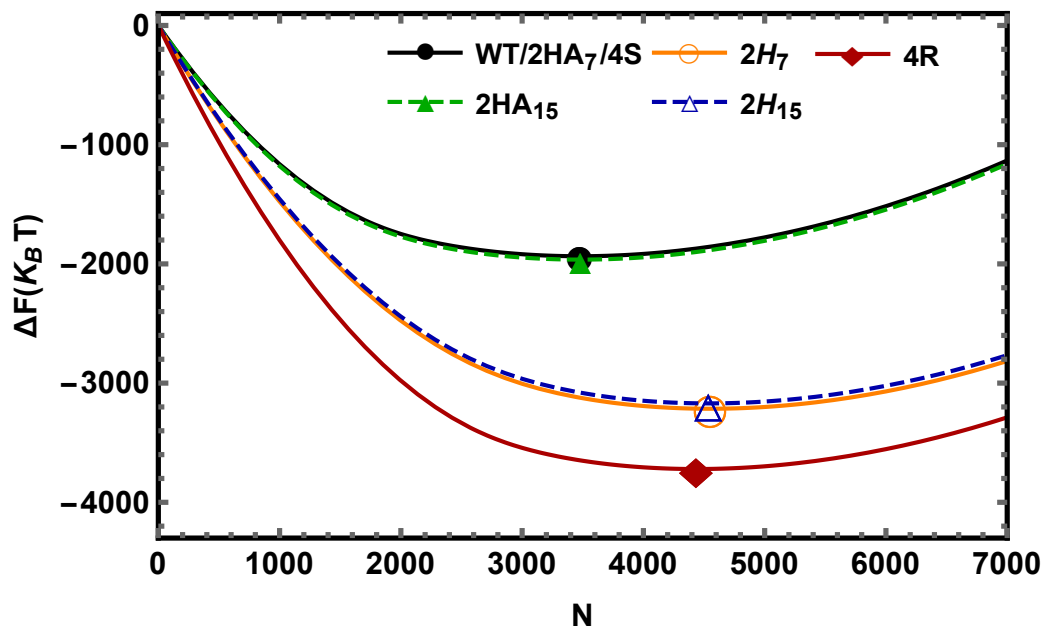


Figure 2.10: The encapsidation free energy as a function of monomer numbers for the mutants presented in Fig. 2.7. 2HA₇ and 4S have the same free energy as wild-type. The additional length inserted in 2HA₇ does not have a huge impact on the optimal encapsidated genome because it does not modify the distance between the charges along the N-terminal tails, see Fig. 2.7. The capsid radius is 12 nm and the tail length is 6.5 nm with total charges on the capsid $Q_c = 1440$. The salt concentration is 500 mM.

Consistent with the experimental data presented in Fig. 2.2 and the last column of Fig. 2.7, our theoretical calculations show that as the number of positive charges on the N-terminal tails increases, the optimal length of the genome increases too. The mutants 4R, 2H₇ and 2H₁₅ have four extra positive charges compared to wild-type proteins and they all

encapsidate longer genomes. Both mutants 2H₇ and 2H₁₅ have longer tails compared to 4R, and our results show that they encapsidate longer genomes, consistent with the experimental findings. Thus the length of N-terminal tails influences the amount of packaged RNA.

While there are many similarities between the experiments presented in Fig. 2.2 and our theoretical results shown in Fig. 2.7, there are also some differences. The comparison of the experiment and theory reveals that more genome is encapsidated by 2HA₁₅ proteins compared to wild-type or 2HA₇ proteins, which is not observed in our calculations. Note that to perform the numerical calculations, we consider that all amino acids have the same effective size (0.2 nm), and the Debye length in our system is $\lambda_D = 0.438$ nm. Since the parameter landscape is quite vast and there are several unknowns, instead of changing the size of each amino acid, we modify the distance between the fourth and fifth charged amino acids in the N-terminal tail of the mutant 2HA₁₅. More specifically, we systematically increase the distance between the fourth and fifth positive charges from 0.2 nm to 2.8 nm where the 8 amino acids were inserted for the case of the mutant 2HA₁₅ and then calculate the optimal length of encapsidated genome for three different salt concentrations of $\mu = 100, 300$ and 500 mM. As illustrated in Fig. 2.9, the optimal length of encapsidated genome depends on both the distance between the fourth and fifth positively charged amino-acids and the salt concentration. The figure reveals that as the distance increases from 0.2nm to 2.2nm, the optimal length of the encapsidated genome first increases and then later decreases.

To gain more insights into the experimental results, we also examined the impact on the optimal polymer length of a uniform charge distribution along the N-terminals versus

a tight one as presented in Fig. 2.11. As shown in the figure, for a given tail length and number of positive charges, when the charges are distributed more uniformly along the N-terminals, the optimal length of encapsidated genome becomes longer.



Charge Distribution	Virions	Effective Free energy	Optimal Length	Percent Change
	Tight	-2719.89	3387.44	N/A
	Loose	-2086.17	3471.32	+2.48%

Figure 2.11: Schematic of two N-terminal tails with different charge distributions. As before each yellow rectangle represents an amino acid with a positive charge and effective size of $d = 0.2 \text{ nm}$ and blue rectangles represent neutral amino acids but with the same size. The table includes the effective encapsulation free energy of the RNA confined into a spherical shell, the optimal length of encapsulated RNA, and the percent change (theory) of optimal length of packaged RNA with respect to the first charge distribution. The salt concentration is $\mu = 500 \text{ mM}$, and tail length is 4.5 nm for both cases. The total charge on the capsid is $Q_c = 1440$. When the charges are distributed more evenly, the optimal length of the encapsidated genome increases. For the first line of the table, the distance between yellow rectangles is either zero or 0.2 nm , while for the second one is either 0.2 nm or 0.4 nm .

2.4.3 Relationships between RNA topology and nucleocapsid structure in a model icosahedral virus

Here, we focus our work on the cowpea chlorotic mottle virus (CCMV), a model virus of the Bromoviridae family. Two RNA segments were used for this study: RNA2 of CCMV (C2) and RNA2 of the bovine rotavirus strain RF (RF2) [138]. These RNAs were chosen for their similar length (2767 and 2687 nucleotides, respectively) and RF2 also because it is not packaged as such by rotavirus capsid, in contrast to C2 by the CCMV CP. Thus, the secondary structure of RF2 has no reason for being optimized by evolutionary processes for packaging purpose. Moreover, because each of them codes for a different

protein, their sequence identities are very low, with very few regions of sequence similarity. See [94] for related experimental results.

The results of calculations are given in Fig. 2.12 as a plot of the density profile of branch points and nucleotide density (inset graph) as a function of r , the distance from the center of the capsid, see the section 2.3 Methods for the details. The blue dashed line is the nucleotide density profile which was calculated by the same procedures as above in 2.4.1 and 2.4.2. The black dashed line is the density profile of branch points for C2 and the red solid line is for RF2. The number of branch points was obtained through RNASubopt, a program in the Vienna RNA package [87]. Generating an ensemble of secondary structures for sequences of C2 and RF2, we calculated the thermally averaged number of branch points from the secondary structures of each RNA. We found that while the number of monomers for C2 and RF2 were close, i.e., 2767 and 2687, respectively, the difference between the number of branch points was more significant: 60 for C2 and 48 for RF2. The number of branch points for B1 was calculated previously [52] and was estimated to be 65.

Figure 2.12 clearly shows that the branch point density for C2 was larger than that for RF2 while their nucleotide densities (inset) were almost the same. Using the profiles for RF2 and C2, we calculated the encapsulation free energy and found that it was lower for C2 ($-3367 k_B T$ with k_B the Boltzmann constant and T the temperature) than for RF2 ($-3358 k_B T$). Since C2 and RF2 had almost the same nucleotide density profile inside the capsid, the difference of their encapsulated free energy was mainly due to the difference in their number of branch points. This slight difference in free energy explains why experimentally the nucleocapsids packaging C2 were statistically more uniform and better ordered across

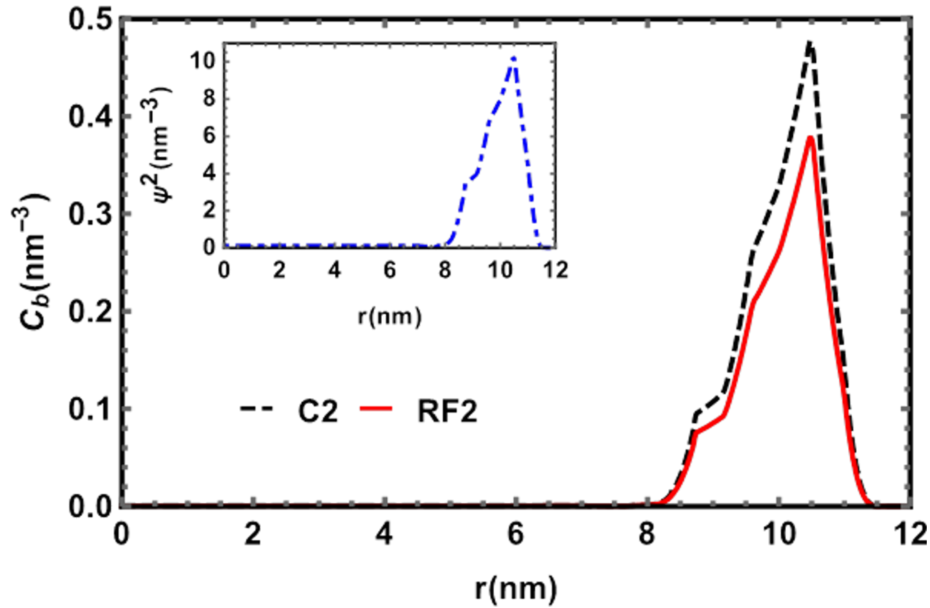


Figure 2.12: Branch density profile versus r , the distance from the center of the capsid. The dashed black line and red solid line correspond to the branch density profiles for C2 and RF2, respectively. The dot-dashed blue line (inset graph) is the nucleotide density profile for C2 and RF2, which is almost the same for both. The capsid radius is $R = 12$ nm, N-terminal tails length 4 nm, and the salt concentration 100 mM. The total number of positive charges on the capsid interior is 1800.

the sample than those packaging RF2.

2.5 Discussion

Despite the fact that many experiments have shown that the number of nucleotides packaged by capsid proteins increases with the number of charges on N-terminal tails, how the amount of encapsidated RNA depends on the distribution of the charges along and the length of the N-terminal domain of capsid proteins is not well-understood. Our results presented in Fig. 2.5(b) for T=1 capsid and Fig. 2.7 and Fig. 2.9 for T=3 viruses show that the electrostatic interaction alone is not sufficient to explain the dependence

of the amount of packaged RNA on the amino sequence of N-terminal tails in the BMV experiments [99]. For example, the amount of packaged RNA is different for the mutant 2HA₁₅ and 2HA₇ as illustrated in Fig. 2.2 whereas both have the same number of charges, very similar charge distribution and the same peptide length. This reveals the importance of specific interactions that depend on the exact type of amino acids, RNA secondary or tertiary structures, and packaging sequences/signals, which involves the highly specific, nonelectrostatic interactions between sections of RNA and capsid proteins [111, 131, 105]. Our mean-field theory does not include this effect and thus cannot explain the experimental observation due to specific interactions; nevertheless, our theory can describe how the length of N-terminal tails and distribution of charges along the peptide control the amount of RNA packaged by BMV capsid proteins, consistent with the experimental data.

The simple case of two charges on the N-terminal tails of the $T = 1$ capsid (Fig. 2.5) shows clearly that when the distance between two positive charges increases, the optimal length of RNA encapsulated into the capsid also increases. A careful examination of Eq. (2.1) shows that the length of the encapsulated polyelectrolyte increases with the distance in order for the chain to be uniformly distributed between the two charges, lowering the entropy contribution (the first term in Eq. (2.1)) as much as possible. Fig. 2.5 shows that the optimal length of genome saturates and remains more or less constant beyond certain distance between the two charges. This is mainly due to the fact that the optimal length of the genome for $T = 1$ is such that the density of genome is low. When the distance between the charges is more than two Debye lengths ($\lambda_D = 0.438 \text{ nm}$ for $\mu = 500 \text{ mM}$), the electrostatic interaction becomes very weak between the distant charges. It appears

that the chain then prefers to reside only in the immediate vicinity of each positive charge along the peptide. More specifically, as the distance between the charges increases, the electrostatic does not promote encapsidation of longer genomes. Thus, the optimal length of genome first increases and then it remains constant even if the distance between the charges increases further.

Figure 2.7 shows that the case for $T = 3$ structure relevant to BMV experiments is more complex. As seen in the figure, the mutant 4R whose charge is increased by substitution instead of insertion (keeping the length constant), has less encapsidated RNAs than do $2H_7$ and $2H_{15}$, while all three mutants have the same number of charges on their tails. Our calculations reveal that since $2H_7$ and $2H_{15}$ have longer N-terminal tails, longer genome is necessary for the chain to *uniformly* wrap around the tail keeping the entropic contribution in Eq. (2.1) low. However, the difference between the length of the genome encapsidated by wild-type proteins and the mutant $2HA_{15}$ proteins whose N-terminal length is increased by insertion of eight neutral amino acids is not as pronounced in our theory as in the experiments. This could be explained at least in part by the distance between the charged amino acids in the peptide. To understand the impact of the distance between the charges, we systematically examined the effect of the distance between the charges in the middle of the N-terminal tail as illustrated in Fig. 2.9. The results presented in the figure is quite intriguing as the optimal length of the genome first increases and then decreases for the three different salt concentrations presented in the figure.

The large distance between the two charges along the N-terminal tail provides more space for the genome to reside. The careful examination of Eq. (2.1) shows that

due to the entropic consideration, the genome will be distributed more or less uniformly along the N-terminal leading to the packaging of longer genomes. However, as the distance between the charges increases and goes beyond two Debye length ($\lambda_D = 0.438 \text{ nm}$ for $\mu = 500 \text{ mM}$, $\lambda_D = 0.565 \text{ nm}$ for $\mu = 300 \text{ mM}$ and $\lambda_D = 0.979 \text{ nm}$ for $\mu = 100 \text{ mM}$), the optimal length of RNA becomes shorter. This effect can be well-understood by investigating the genome profiles presented in Fig. 2.8. When the distance between the fourth and fifth charges is very large, there will be two distinct peaks in the genome profile with almost no nucleotides between the charges indicating that the negatively charged RNA prefers to be localized mainly around the positive charges. Figure 2.9 indicates that as the distance between the charges increases, at some point the optimal length of encapsidated RNA decreases resulting into the lower polymer density, which also reduces the entropy cost of formation of two completely separate peaks. Since the Debye length is longer for lower salt concentrations, the optimal length of genome starts decreasing at $d = 1.8 \text{ nm}$ for $\mu = 100 \text{ mM}$ (2HA₁₅(M3)), $d = 1.4 \text{ nm}$ for $\mu = 300 \text{ mM}$ (2HA₁₅(M2)) and $d = 1.0 \text{ nm}$ for $\mu = 500 \text{ mM}$ (2HA₁₅(M1)). While the behavior is the same for all three salt concentrations, the effect is less pronounced as the salt concentration decreases. Figure 2.11 supports that for a given length and number of positive charges, the more uniformly charges are dispersed along the N-terminals, the longer the optimal length of encapsidated genome becomes.

We emphasize that the goal of this chapter has been to qualitatively explain the experimental results and to explore the impact of entropy and electrostatic interaction that depend on the distance between the charges and not the details of protein structures. A better quantitative comparison between the experiments and theory can be obtained, if

the theory includes many other effects such as counter-ion condensation, the presence of divalent ions, structure of proteins and packaging signals discussed above. This study also compared the structure of CCMV nucleocapsids packaging different RNA segments with similar lengths but decreasing levels of compactness; C2 belongs to CCMV genome and RF2 is an unrelated segment derived from rotavirus genome. Because RF2 had a lower number of branch points than C2, theoretical calculations revealed that the free energy for packaging RF2 was slightly higher than for packaging C2.

In summary, in this chapter we explore whether the variation in RNA packaging by BMV mutants observed in the experiments of Ni *et al.* [99] and presented in Fig. 2.2 [99] can be understood by the mean-field theory incorporating electrostatics, excluded volume interaction and RNA conformational entropy. In particular, we have calculated, as a function of the number and location of charges in the peptide tails, the free energy of an RNA confined in a spherical shell interacting with the N-terminal tails and ions. We find that the combined effect of the electrostatic interaction and the genome entropy considerations can shed light on many experimental data relevant to BMV assembly. While our mean-field theory cannot explain all the experimental data, we have been able to show that the location and the distance between charges along the N-terminal tails significantly influence the amount of packaged RNA. Understanding the factors contributing to the virus assembly and RNA packaging will pave the path for interfering with the different stages of the virus life cycle.

Chapter 3

Exact Solution for elastic networks on curved surfaces

3.1 Abstract

The problem of characterizing the structure of an elastic network constrained to lie on a frozen curved surface appears in many areas of science and has been addressed by many different approaches, most notably, extending linear elasticity or through effective defect interaction models. In this Letter, we show that the problem can be solved by considering nonlinear elasticity in an exact form without resorting to any approximation in terms of geometric quantities. In this way, we are able to consider different effects that have been unwieldy or not viable to include in the past, such as a finite line tension, explicit dependence on the Poisson ratio, or the determination of the particle positions for the entire lattice. Several geometries with rotational symmetry are solved explicitly. Comparison with

linear elasticity reveals an agreement that extends beyond its strict range of applicability. Implications for the problem of the characterization of virus assembly are also discussed.

3.2 Introduction

Deciphering the design principles of life is one of the lingering mysteries facing researchers in many areas of science. Among many biological systems, viruses, in particular, have received much more attention as they are ubiquitous pathogens in our environment with members infecting all kingdoms of life. Most viruses, from the simplest to the most complicated, and from the least to the most evolved, are constituted of a protein shell (or “capsid”) that encloses the viral genetic material (RNA or DNA) [68, 58]. Understanding the process of virus assembly is a fundamental challenge of ever-increasing interest, not only because it is a central stage of the viral life cycle, but also because it is the target of antiviral therapeutic strategies. The Coronavirus Disease 2019 (COVID-19) pandemic, connected to SARS-CoV-2 revealed more than ever the importance of identifying new ways to combat viruses. In this context, our current understanding of virus assembly is quite limited. The difficulties arise from the interplay between curvature and crystalline order and their role in determining the positions of lattice defects on elastic surfaces with non-zero Gaussian curvature [141, 104, 103, 95].

Fig. 3.1 illustrates the structure of two viruses with different geometries: (a) Herpes Simplex Virus (HSV) with icosahedral symmetry [152] and (b) Human Immunodeficiency Virus (HIV) [57] with a conical structure. In the case of HSV, the position of 12 pentagonal

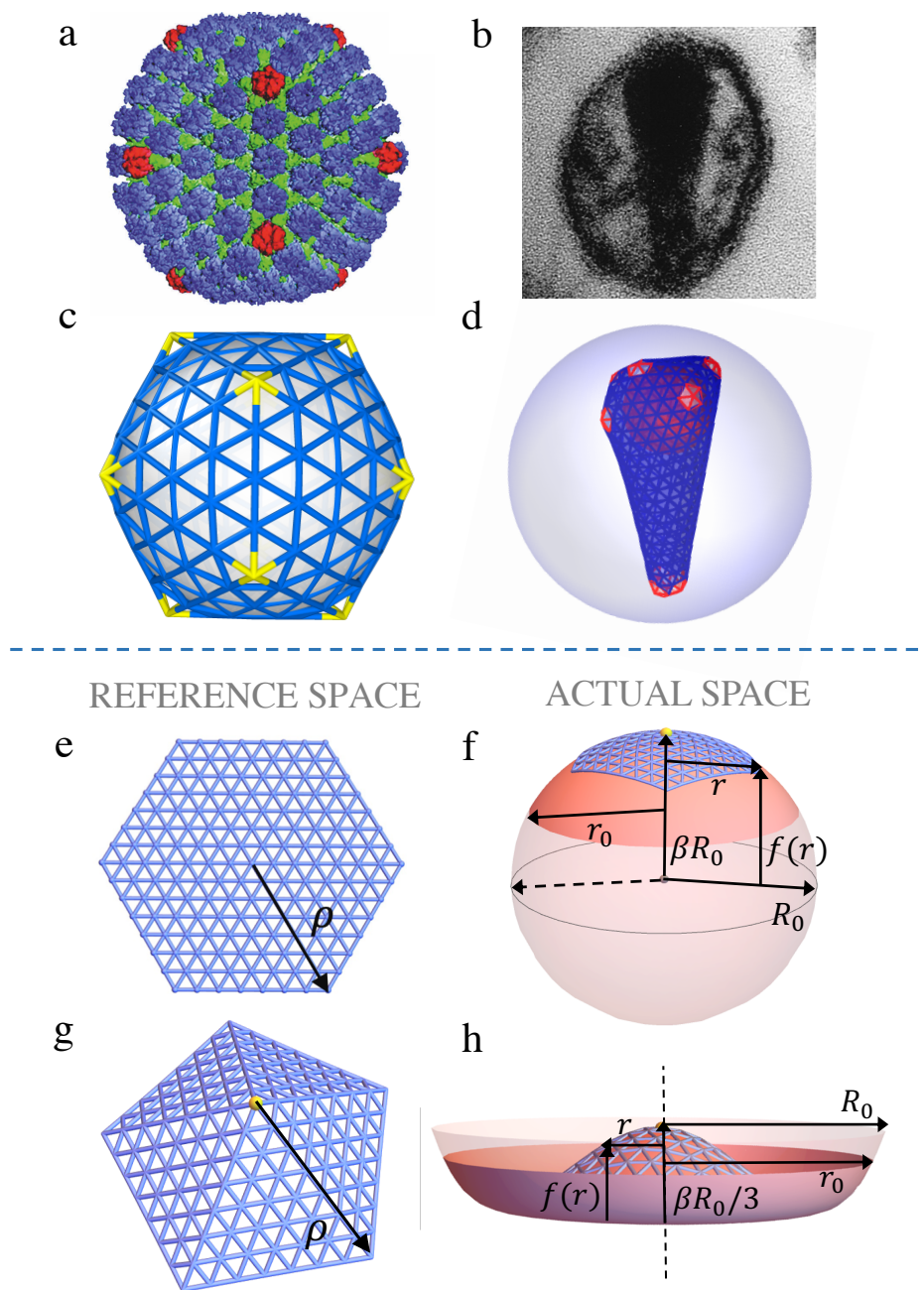


Figure 3.1: (a) 3D cryo-EM reconstruction of HSV [152]; (b) EM image of HIV surrounded with the lipid envelope [57]. The result of the computer simulations of (c) an HSV [83] and (d) an HIV with genome and membrane [100]. (e) The reference space consists of a flat surface without a disclination or (g) with a disclination at the center. The actual space is any manifold endowed with its natural metric. As an example, we consider (f) a spheroid($\beta = 1$) and (h) a sombrero($\beta = 3$).

defects is precise (Fig. 3.1c) to preserve the symmetry of the shell. In previous work, we have shown that as a triangular lattice such as a HSV shell (Fig. 3.1c) [83] grows over a spherical scaffold, defects appear one by one at the vertices of an icosahedron, explaining how error-free structures with icosahedral order assemble. For HIV shells, while there are often 5 defects at the smaller and 7 at the larger caps, defect positions can vary from one HIV structure to another. The computer simulations of Ref. [100] have shown how the presence of genome and membrane contributes to the formation of the conspicuous conical HIV structures; nevertheless, currently, there is no clear understanding of what determines the position of the defects as the surfaces with non-zero Gaussian curvature such as the conical shell of HIV grow.

Capsid formation dynamics is just one example of the more general class of problems consisting of crystal growth on curved geometries. Other examples are faceted insect eyes, liquid crystals, curved array of microlenses in optical engineering systems, other protein cages in addition to viral capsids such as platonic hydrocarbons, heat shock proteins, ferritins, carboxysomes, silicages, multicomponent ligand assemblies, clathrin vesicles and many other cellular organelles, as reviewed in Ref. [147], for example. These problems have been mostly addressed by extending linear elasticity [121, 96, 3, 67, 27, 75] or through effective defect interaction models [107, 17, 16, 86, 62, 1]. Either approach brings significant limitations: linear elasticity and its extensions impose certain approximations on geometric quantities such as Gaussian curvature (see discussion in Sec. 3.3.7, specially Eq. 3.54) and fail to satisfy global topological and geometrical constraints, most notably the Gauss-Bonnet theorem [84]. Effective defect interaction models satisfy global properties exactly,

but with uncontrolled approximations and obvious deficiencies: they predict a universal independence of all measurable quantities on the Poisson ratio ν_p and it is unknown, thus far, how to include other relevant free energy terms such as line tension. Furthermore, it requires computing the inverse square Laplacian, a formidable complex task beyond simple cases.

In this chapter, we formulate general elasticity in curved surfaces based on geometric invariants and solve the equations exactly. We build upon our previous results for spherical caps [84, 85] combined with covariant elasticity [48, 97]. Thus, the approach not only surmounts all the limitations of the previous theories, but provides an exact solution to the problem. Furthermore the theory allows to explore the impact of line tension and the Poisson ratio on the assembly of curved surfaces. The method is general and may be generalized to surfaces without specific symmetries. In this way, the approach provides the first step towards obtaining a complete theory for the self assembly of non-spherical virus capsids.

3.3 Methods

3.3.1 Summary of the methods

The free energy of a partially formed elastic shell can be written as

$$\begin{aligned}
 F &= F^{elastic} + F^{bending} + F^{abs} + F^{line} \\
 &= \int d^2 \mathbf{x} \sqrt{g} \left[\mathcal{F}^{elastic} + \mathcal{F}^{bending} \right] + F^{abs} + F^{line} .
 \end{aligned}
 \tag{3.1}$$

where the first and second terms correspond to the stretching and bending energies of the patch respectively, the third term represents the attractive monomer-monomer interaction promoting the crystal growth and the last term is associated with the cost of the rim energy due to the fact that the subunits at the boundary have fewer neighbors than the ones at the interior of the surface. The elastic term $\mathcal{F}^{elastic}$ (see Eq. 3.9) contains a quadratic term [77] in the strain tensor $u_{\alpha\beta}$

$$u_{\alpha\beta} = \frac{1}{2} [g_{\alpha\beta} - \bar{g}_{\alpha\beta}] , \quad (3.2)$$

where $g_{\alpha\beta}(\mathbf{x})$ is the *actual metric* (the metric of the curved surface) and $\bar{g}_{\alpha\beta}(\bar{\mathbf{x}})$ is the *reference metric* describing a perfect lattice, *i.e.* the one consisting of equilateral triangles (Fig. 3.1e-h). Note that the Gaussian curvature of the reference metric is zero, except possibly, on a finite number of points that define disclination cores [84]. If a_L is the lattice constant of the perfect triangular lattice in reference space, the number of particles $N \gg 1$ making the crystal with a given area \hat{A} is

$$\hat{A} = \int d^2\mathbf{x} \sqrt{\bar{g}(\mathbf{x})} = \int d^2\bar{\mathbf{x}} \sqrt{\bar{g}(\bar{\mathbf{x}})} = \frac{\sqrt{3}}{2} N a_L^2 . \quad (3.3)$$

The second term in Eq. 3.1 can be expressed in terms of the two radii of curvature $(R_i)_{i=1,2}$,

$$\mathcal{F}^{bending} = \kappa \left[\left(\frac{1}{R_1} - H_0 \right)^2 + \left(\frac{1}{R_2} - H_0 \right)^2 \right] , \quad (3.4)$$

with κ the bending rigidity and H_0 the mean spontaneous curvature of the constituents or subunits. We emphasize that the free energy density, Eq. 3.1, has no trivial solution. The only surfaces allowing zero strains have either zero Gaussian curvature: a plane or cylinder

($q = 0$) or a delta function at the origin like a cone ($q = 1$). The absolute minimum of the bending rigidity implies a surface with $R_1 = R_2 = \frac{1}{H_0}$, a sphere. There is no surface that simultaneously minimizes both the elastic and bending energies. The third term in Eq. 3.1, $F^{abs} = -\Pi\hat{A} < 0$ with Π the attractive interaction per unit area due to favorable hydrophobic contacts between subunits, is the driving force for crystal growth [96].

In this chapter we consider a frozen geometry, so the actual metric $g_{\alpha\beta}(\boldsymbol{x})$ is that of the corresponding surface. The defect distribution is also given, so the reference metric $\bar{g}_{\alpha\beta}(\bar{\boldsymbol{x}})$ is fixed as well. The problem consists of mapping the actual and reference space $\boldsymbol{x} = \mathcal{U}(\bar{\boldsymbol{x}})$. In other words, \mathcal{U} is the function that determines how the perfect lattice in reference space maps to the deformed one in actual space, as illustrated in Fig. 3.1e-h. For simplicity, in this chapter we consider only surfaces of revolution, defined by $x = r \cos(\theta), y = r \sin(\theta), z = f(r)$ with actual metric

$$ds^2 = (1 + f'(r)^2) dr^2 + r^2 d\theta^2 , \quad (3.5)$$

see Fig. 3.1. Further, we consider an isotropic reference metric

$$ds^2 = \rho'(r)^2 dr^2 + \alpha^2 \rho(r)^2 d\theta^2 = d^2\rho + \alpha^2 \rho^2 d\theta^2 \quad (3.6)$$

with $\alpha = 1 - \frac{q}{6}$ and $q = 0, \pm 1$ corresponding to no disclination or with a disclination at the origin. The two metrics are basically *incompatible* [97], that is, for a fixed $f(r)$ there is no choice of $\rho(r)$ that will make the strain tensor Eq. 3.2 vanish, as their Gaussian curvatures generally differ. In the isotropic case, the function \mathcal{U} describing the mapping from the

actual to reference space (or reference to actual) can be expressed as $\rho(r)$ or $r(\rho)$. To make the presentation of the chapter simple, we consider a situation in which the surface is given through $f(r)$ (see Fig. 3.1). The elastic energy given in Eq. 3.1 depends on $\rho(r)$ and thus becomes Eq. 3.15. And then the problem consists of finding the optimal $\rho(r)$ that minimizes the free energy Eq. 3.1. Following Ref. [84, 48], this leads to

$$\nabla_{\alpha}\sigma^{\alpha\beta} + (\bar{\Gamma}_{\gamma\nu}^{\beta} - \Gamma_{\gamma\nu}^{\beta})\sigma^{\gamma\nu} = 0 , \quad (3.7)$$

where $\sigma^{\alpha\beta}$ is the stress tensor and $\Gamma_{\gamma\nu}^{\beta}$ are the Christoffel symbols for the reference and actual metrics (see Sec. 3.3.3). This is a one dimensional non-linear differential equation with just one unknown $\rho(r)$ (see Eq. 3.20). We solve Eq. 3.7 subject to the following boundary condition,

$$n_{\rho}\sigma^{\rho\lambda}\bar{g}_{\lambda\nu} = -\frac{\tau}{r_A}n_{\nu} . \quad (3.8)$$

where τ is the line tension, $n^{\mu} = g^{\mu\nu}n_{\nu}$ is the normal to the boundary within the surface and r_A is the curvature of the boundary. For a surface with rotational symmetry and a circular boundary $r = r_0$, this equation simply becomes Eq. 3.33. For a tensionless boundary, obviously $\tau = 0$. In Sec. 3.3.4, we provide a detailed derivation of Eq. 3.8 from the line energy $F^{line} = \tau \oint_{\partial D} ds$ with ds an infinitesimal length for the boundary ∂D in actual space. We also show how Eqs. 3.1-3.8 reduce to standard linear elasticity and provide an explicit analytical solutions within linear elasticity for both a spheroid, $f(r) = \beta\sqrt{R_0^2 - r^2}$ (Fig. 3.1f) and a sombrero surface, $f(r) = \beta R_0/3 \left(1 - (r/R_0)^2 + (r/R_0)^4\right)^{3/2}$ (Fig. 3.1h), where β is a unitless number, in Sec. 3.3.7.

To obtain the free energy of the system, we first calculate $\rho(r)$ through Eq. 3.7, which minimizes Eq. 3.1. The plots in Fig. 3.2 below each show the optimal $\rho(r)$ or $r(\rho)$ for both spheroid and sombrero surfaces with no disclination or with one disclination at the center. It is important to note that with the exact $r(\rho)$, we can reconstruct the lattice in actual space; the positions of the lattice in reference space $(\rho_i, \theta_i)_{i=1\dots N}$ are known and consists of N equilateral triangles with lattice constant a_L (if $N \gg 1$), see Eq. 3.3. Then, from $r(\rho)$, the positions in actual space $(r(\rho_i), \theta_i)_{i=1\dots N}$ are obtained, as illustrated in Fig. 3.2. Thus, we find a solution to the problem of finding the best possible triangulation consisting of equilateral triangles that cover a given surface. Note that this solution is independent of the underlying potential among the constituent particles, and therefore, hereon we refer to this triangulation as the *universal mapping lattice*.

Plugging the solutions of $\rho(r)$ into Eq. 3.1, we obtain the free energy of the system. First, we consider the case of free boundary conditions as illustrated in Fig. 3.3. For comparison, we show the predictions from linear elasticity, which become exact in the limit of small curvature ($\frac{\hat{A}}{R_0^2} \rightarrow 0$), both for the defect free case $q_i = 0$ and a single disclination $q_i = 1$.

Very generally, we find that the applicability of elasticity theory extends to relatively large curvatures ($\frac{\hat{A}}{R_0^2} \approx 1$). For the spheroid, linear elasticity remains qualitatively correct for the entire range explored, but this is not the case for the sombrero surface, see Fig. 3.3, where linear elasticity breaks down and cannot be extended beyond a certain limit.

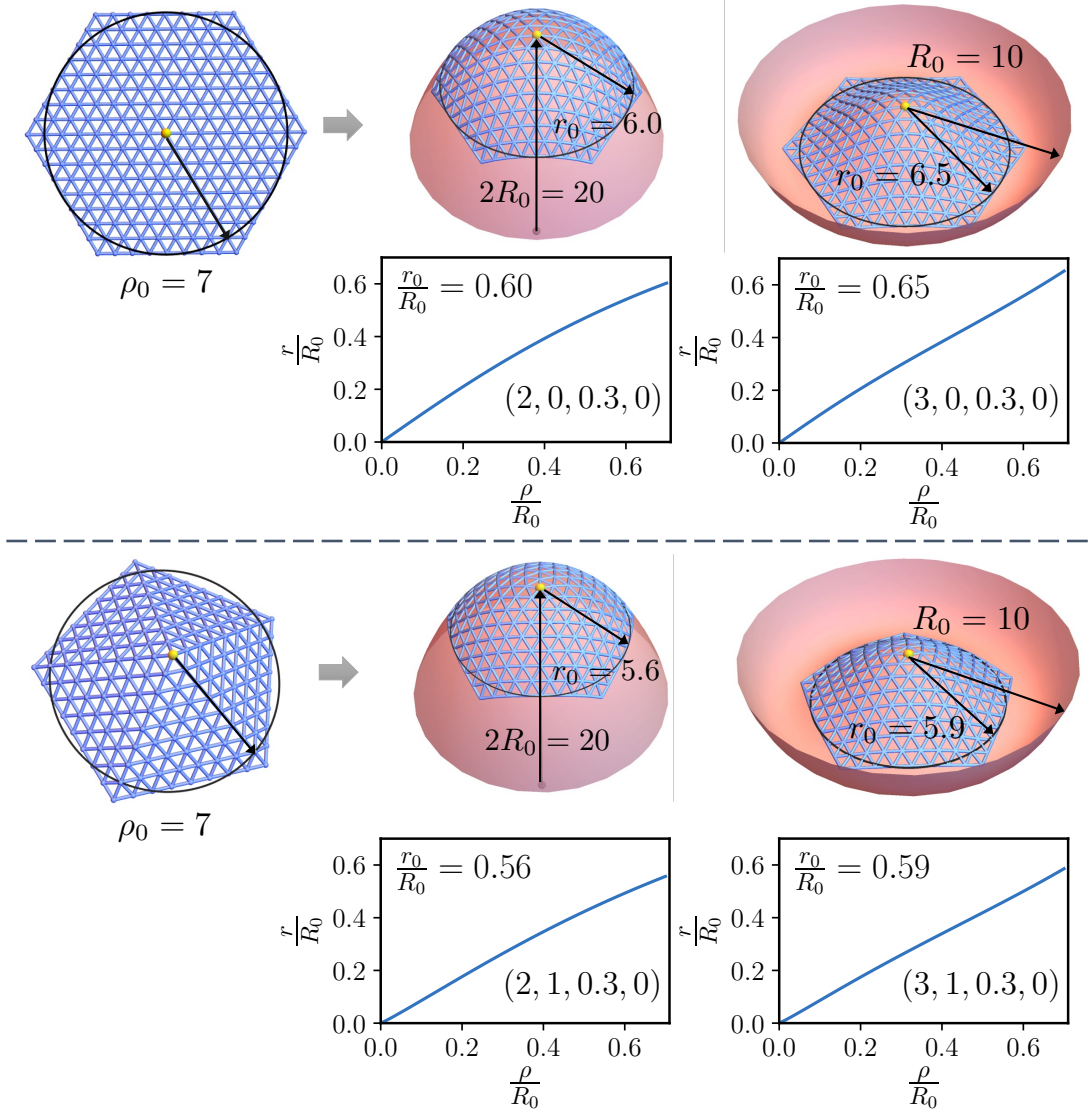


Figure 3.2: Lattice reconstruction for both the spheroid and sombrero. The lengths R_0 , ρ_0 and r_0 are given in units of the lattice constant a_L . The plots under spheroid and sombrero surfaces indicate the corresponding mapping function from the reference to actual space. $\frac{\rho_0}{R_0} = 0.7$ for each plot. The number in the parenthesis in each figure denote $(\beta, q_i, \nu_p, \hat{\tau})$.

3.3.2 Explicit formulas for the different quantities

In this section, we provide explicit expressions for the formulas in the main text.

The explicit form of $\mathcal{F}^{elastic}$ is

$$\mathcal{F}^{elastic} = \frac{1}{2} A^{\alpha\beta\gamma\delta} u_{\alpha\beta} u_{\gamma\delta}, \quad (3.9)$$

where $u_{\alpha\beta}$ (see Eq. 3.2 in the main text) is the strain tensor and

$$A^{\alpha\beta\gamma\delta} = \frac{Y}{1 - \nu_p^2} \left[\nu_p g^{\alpha\beta} g^{\gamma\delta} + (1 - \nu_p) g^{\alpha\gamma} g^{\beta\delta} \right], \quad (3.10)$$

with Y the Young Modulus, ν_p the Poisson ratio and $g^{\alpha\beta}$ the actual metric.

The Gaussian curvature is

$$K = \frac{\det(\partial_i \partial_j f)}{(1 + (\nabla f)^2)^2} = \frac{f'(r) f''(r)}{r(1 + f'(r)^2)^2}, \quad (3.11)$$

and the mean curvature (with the convention that $R_i = R > 0, i = 1, 2$ for the sphere) is

$$\begin{aligned} 2H &= -\nabla \cdot \left(\frac{\nabla f}{(1 + (\nabla f)^2)^{1/2}} \right) \\ &= - \left(\frac{f''(r)}{(1 + f'(r)^2)^{3/2}} + \frac{f'(r)}{r(1 + f'(r)^2)^{1/2}} \right) \end{aligned} \quad (3.12)$$

see Ref. [147] for the details. The two curvatures can be obtained from the equation

$$\begin{aligned} K &= \frac{1}{R_1 R_2} \\ 2H &= \frac{1}{R_1} + \frac{1}{R_2}, \end{aligned} \quad (3.13)$$

such that $\frac{1}{R_1} = H + \sqrt{H^2 - K}$ and $\frac{1}{R_2} = H - \sqrt{H^2 - K}$ with H and K given in Eqs. 3.11 and 3.12. The stress tensor is

$$\sigma^{\alpha\beta} = \frac{1}{\sqrt{g}} \frac{\delta F}{\delta u_{\alpha\beta}} = A^{\alpha\beta\gamma\delta} u_{\gamma\delta}. \quad (3.14)$$

3.3.3 Particularization to surface revolution

In this section, we provide various quantities for the surfaces of revolution, defined by $x = r \cos(\theta)$, $y = r \sin(\theta)$, $z = f(r)$ with actual metric defined in Eq. 3.5. The nonzero Christoffel symbols are given in Ref. [84] and are

symbol	Γ_{rr}^r	$\Gamma_{\theta\theta}^r$	$\Gamma_{\theta r}^\theta$
reference	$\frac{\rho''(r)}{\rho'(r)}$	$-\alpha^2 \frac{\rho(r)}{\rho'(r)}$	$\frac{\rho'(r)}{\rho(r)}$
actual	$\frac{f'(r)f''(r)}{1+f'(r)^2}$	$\frac{-r}{1+f'(r)^2}$	$\frac{1}{r}$

The elastic energy given in Eq. 3.1 depends on $\rho(r)$,

$$\begin{aligned} \frac{F^{elastic}}{Y\pi} &= \frac{1}{4(1-\nu_p^2)} \int dr r (1+f'(r)^2)^{1/2} \\ &\left[\left(1 - \frac{\rho'(r)^2}{1+f'(r)^2}\right)^2 + \left(1 - \frac{\alpha^2 \rho(r)^2}{r^2}\right)^2 + \right. \\ &\left. + 2\nu_p \left(1 - \frac{\rho'(r)^2}{1+f'(r)^2}\right) \left(1 - \frac{\alpha^2 \rho(r)^2}{r^2}\right) \right] \end{aligned} \quad (3.15)$$

The stress tensor given in Eq. 3.14 becomes,

$$\begin{aligned} \sigma^{rr} &= \frac{Y}{2(1-\nu_p^2)(1+f'(r)^2)} \left[1 - \frac{\rho'(r)^2}{1+f'(r)^2} + \right. \\ &\left. + \nu_p \left(1 - \left(\frac{\alpha\rho(r)}{r}\right)^2\right) \right] \\ \sigma^{r\theta} &= \sigma^{\theta r} = 0 \\ \sigma^{\theta\theta} &= \frac{Y}{2r^2(1-\nu_p^2)} \left[\nu_p \left(1 - \frac{\rho'(r)^2}{1+f'(r)^2}\right) + \right. \\ &\left. + 1 - \left(\frac{\alpha\rho(r)}{r}\right)^2 \right] \end{aligned} \quad (3.16)$$

Using Eq. 3.15 and Eq. 3.16, the general form of Eq. 3.7 for $\beta = r$ becomes

$$\begin{aligned} \partial_r \sigma^{rr} + \bar{\Gamma}_{rr}^r \sigma^{rr} + \bar{\Gamma}_{rr}^r \sigma^{rr} + \bar{\Gamma}_{\theta\theta}^r \sigma^{\theta\theta} + \\ + \Gamma_{rr}^r \sigma^{rr} + \Gamma_{\theta r}^\theta \sigma^{rr} = 0, \end{aligned} \quad (3.17)$$

which can be written as

$$\frac{d\sigma^{rr}}{dr} + \left(\bar{\Gamma}_{rr}^r + \Gamma_{rr}^r + \Gamma_{\theta r}^\theta \right) \sigma^{rr} + \bar{\Gamma}_{\theta\theta}^r \sigma^{\theta\theta} = 0 \quad (3.18)$$

The explicit form of the derivative of the stress tensor is,

$$\begin{aligned}
\frac{d\sigma^{rr}}{dr} &= \frac{-Y}{(1-v_p^2)} \frac{f'(r)f''(r)}{(1+f'(r)^2)^2} \left[1 - \frac{\rho'(r)^2}{1+f'(r)^2} + \right. \\
&+ \left. \nu_p \left(1 - \left(\frac{\alpha\rho(r)}{r} \right)^2 \right) \right] + \\
&+ \frac{Y}{(1-v_p^2)(1+f'(r)^2)} \left[\frac{\rho'(r)^2 f'(r)f''(r)}{(1+f'(r)^2)^2} - \right. \\
&- \left. \frac{\rho'(r)\rho''(r)}{1+f'(r)^2} + \nu_p \frac{\alpha^2\rho(r)^2}{r^3} - \nu_p \frac{\alpha^2\rho(r)\rho'(r)}{r^2} \right]
\end{aligned} \tag{3.19}$$

The equation determining $\rho(r)$, Eq. 3.18, becomes

$$\begin{aligned}
&\frac{-f'(r)f''(r)}{(1+f'(r)^2)} \left[1 - \frac{\rho'(r)^2}{1+f'(r)^2} + \right. \\
&+ \left. \nu_p \left(1 - \left(\frac{\alpha\rho(r)}{r} \right)^2 \right) \right] + \frac{1}{1+f'(r)^2} \times \\
&\times \left[\frac{\rho'(r)^2 f'(r)f''(r)}{(1+f'(r)^2)^2} - \frac{\rho'(r)\rho''(r)}{1+f'(r)^2} + \nu_p \frac{\alpha^2\rho(r)^2}{r^3} - \right. \\
&- \left. \nu_p \frac{\alpha^2\rho(r)\rho'(r)}{r^2} \right] + \left(\frac{\rho''(r)}{\rho'(r)} + \frac{f'(r)f''(r)}{1+f'(r)^2} + \frac{1}{r} \right) \times \\
&\times \frac{1}{2(1+f'(r)^2)} \left[1 - \frac{\rho'(r)^2}{1+f'(r)^2} + \right. \\
&+ \left. \nu_p \left(1 - \left(\frac{\alpha\rho(r)}{r} \right)^2 \right) \right] - \frac{\alpha^2\rho(r)}{2r^2\rho'(r)} \times \\
&\times \left[\nu_p \left(1 - \frac{\rho'(r)^2}{1+f'(r)^2} \right) + 1 - \left(\frac{\alpha\rho(r)}{r} \right)^2 \right] = 0 .
\end{aligned} \tag{3.20}$$

3.3.4 Boundary condition

The boundary conditions can be obtained through the variations of $F^{area} = F^{elastic} + F^{bending}$ in Eq. 3.1 and the reparameterizations of the actual metric,

$$\delta F^{area} = - \int d^2 \mathbf{x} \partial_\rho (\sqrt{g} \sigma^{\rho\mu} \bar{\xi}_\mu) = \oint \sqrt{g} dx^\mu \varepsilon_{\mu\nu} \sigma^{\nu\rho} \bar{\xi}_\rho . \quad (3.21)$$

If there is a line tension τ , the contribution to the free energy is then

$$F^{line} = \tau \oint ds = \tau \oint \sqrt{\langle g \rangle} dl = \tau \oint dx^\mu g_{\mu\nu} t^\nu , \quad (3.22)$$

where $\langle g \rangle = g_{\mu\nu} \frac{dx^\mu}{dl} \frac{dx^\nu}{dl}$ and

$$t^\mu = \frac{1}{\sqrt{\langle g \rangle}} \frac{dx^\mu}{dl} , \quad (3.23)$$

is the unit vector (in the actual metric) tangent to the boundary curve with $dx^\mu = \sqrt{\langle g \rangle} t^\mu dl$.

The variations to the free energy Eq. 3.22 as shown in Ref. [84] then become,

$$\begin{aligned} \delta F^{line} &= \frac{\tau}{2} \oint dx^\mu \delta g_{\mu\nu} t^\nu = -\tau \oint dx^\mu \nabla_\mu \xi_\nu t^\nu \\ &= \tau \oint dx^\mu \xi_\nu \nabla_\mu t^\nu \end{aligned} \quad (3.24)$$

Since $\delta (F^{area} + F^{line}) = 0$, the appropriate boundary condition is then,

$$\tau g_{\rho\nu} \nabla_\mu t^\rho = -\sqrt{g} \epsilon_{\mu\rho} \sigma^{\rho\lambda} \bar{g}_{\lambda\nu} . \quad (3.25)$$

Note the tangent vector is

$$t^\rho \nabla_\rho t^\mu = -\frac{1}{r_A} n^\mu, \quad (3.26)$$

where r_A is the curvature of the curve defining the boundary. The normal to the tangent vector can be written as

$$n_\mu = \sqrt{\bar{g}} \varepsilon_{\mu\rho} t^\rho. \quad (3.27)$$

The boundary condition, thus, becomes

$$n_\rho \sigma^{\rho\lambda} \bar{g}_{\lambda\nu} = -\frac{\tau}{r_A} n_\nu. \quad (3.28)$$

This condition is slightly different from the one given in Ref. [84]. This is because $\xi^\mu = \bar{\xi}^\mu$ so that $g_{\mu\nu} \xi^\nu = \xi_\mu \neq \bar{\xi}_\mu = \bar{g}_{\mu\nu} \xi^\nu$, which was overlooked in previous work.

As an example, if the problem has rotational symmetry and the boundary is a circle $r = r_0$, then $\theta(s) = s/r_0$. The tangent vectors then are

$$t^r = 0 \quad t^\theta = \frac{1}{r_0}, \quad (3.29)$$

and the normal vectors

$$n_r = \sqrt{1 + f'(r_0)^2} \quad n_\theta = 0. \quad (3.30)$$

Then Eq. 3.26 becomes

$$\begin{aligned} t^\rho \nabla_\rho t^r &= \Gamma_{\theta\theta}^r \frac{1}{r_0^2} = -\frac{1}{r_0(1 + f'(r_0)^2)} \\ &= -\frac{1}{r_0 \sqrt{1 + f'(r_0)^2}} n^r \end{aligned} \quad (3.31)$$

with

$$\frac{r_A}{r_0} = \sqrt{1 + f'(r_0)^2}. \quad (3.32)$$

Finally, the boundary condition becomes equal to,

$$\bar{g}_{rr}(r_0)\sigma^{rr}(r_0) = -\frac{1}{\sqrt{1 + f'(r_0)^2}} \frac{\tau}{r_0}. \quad (3.33)$$

For a sphere $f'(r) = -\frac{r}{\sqrt{R_0^2 - r^2}}$ and thus

$$r_A = \frac{r_0}{\sqrt{1 - \left(\frac{r_0}{R_0}\right)^2}}. \quad (3.34)$$

3.3.5 Free energy normalization

We will consider a dimensionless free energy normalized per particle, that is

$$f \equiv \frac{F}{Y\bar{A}} = \frac{2F}{\sqrt{3}NYa_L^2}, \quad (3.35)$$

hence the area in *reference space*, see Eq. 3. This area is given by

$$\hat{A} = \pi \left(1 - \frac{q_i}{6}\right) \rho_0^2. \quad (3.36)$$

Given two systems with the same number of particles, the one with the smallest free energy per particle, Eq. 3.35 is the stable minimum.

3.3.6 About units

The free energy, see Eq. 1 is

$$F = \int d^2 \mathbf{x} \sqrt{g} \left[\mathcal{F}^{elastic} + \mathcal{F}^{bending} \right] + F^{abs} + F^{line} . \quad (3.37)$$

Note that the stress tensor, Eq. A6 is given by

$$\sigma^{\alpha\beta} = \frac{1}{\sqrt{g}} \frac{\delta F}{\delta u_{\alpha\beta}} = A^{\alpha\beta\gamma\delta} u_{\gamma\delta} = Y \times (\text{Terms that Depend on } \nu_p) \quad (3.38)$$

Note also, that the ratio of the Young modulus and the line tension defines a coefficient l_A with units of length

$$\frac{\tau}{Y} \equiv l_A \quad (3.39)$$

Therefore, through the boundary conditions Eq. 8, the quantity

$$\frac{\sigma^{\alpha\beta}}{Y} = h(\nu_p, l_A) , \quad (3.40)$$

does not directly depend on the Young modulus. Also,

$$F^{abs} = -N\Delta F = -\frac{2\Delta F}{\sqrt{3}a_L^2} \frac{\sqrt{3}}{2} Na_L^2 = -\Pi \hat{A} \text{ with } \hat{A} = \int d^2 \mathbf{r} \sqrt{\bar{g}(\mathbf{x})} . \quad (3.41)$$

Therefore $\Pi = \frac{2\Delta F}{\sqrt{3}a_L^2}$, and \hat{A} is the area in reference space.

The line tension term is a function of the perimeter (P), given by

$$\frac{F^{line}}{YL^2} = \frac{\tau}{YL^2} \oint_{\partial D} ds \equiv \frac{\tau P}{YL^2} . \quad (3.42)$$

Finally, the free energy Eq. 3.37 is

$$\frac{F}{L^2 Y} = f(\nu_p, \frac{l_A}{L}) + \frac{\kappa}{YL^2} \int d^2 \mathbf{x} \sqrt{g} \left[\left(\frac{1}{R_1} - H_0 \right)^2 + \left(\frac{1}{R_2} - H_0 \right)^2 \right] + \frac{\Pi}{Y} \frac{\hat{A}}{L^2} + \frac{\tau P}{YL^2} , \quad (3.43)$$

where L is a characteristic dimension of the system. In general we will choose $L^2 = \hat{A}$, so

$$\frac{F}{Y \hat{A}} = f(\nu_p, \frac{l_A}{\sqrt{\hat{A}}}) + \frac{\kappa}{Y \hat{A}} \int d^2 \mathbf{x} \sqrt{g} \left[\left(\frac{1}{R_1} - H_0 \right)^2 + \left(\frac{1}{R_2} - H_0 \right)^2 \right] + \frac{\Pi}{Y} + \frac{\tau P}{Y \hat{A}} , \quad (3.44)$$

which defines the effective linear tension $\hat{\tau} = \frac{\tau a_L}{Y \hat{A}}$ and dimensionless area \hat{A}/a_L^2 , so that all lengths are expressed in terms of the lattice constant a_L .

3.3.7 Connection with linear elasticity theory

Here we show that the covariant formalism defined by Eq. 1, Eq. 2 reduce to the known formulas from elasticity theory when the displacements are small. Within elasticity theory, the reference metric (without disclinations)

$$\bar{g}_{\alpha\beta} = \delta_{\alpha\beta} \quad (3.45)$$

The surface is described in the Monge gauge,

$$z = h(x, y) . \quad (3.46)$$

The mapping $\mathbf{x} = \mathcal{U}(\bar{\mathbf{x}})$ is given by

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{u}(\bar{\mathbf{x}}) , \quad (3.47)$$

where \mathbf{u} is the displacement. Then, the actual metric becomes

$$\begin{aligned} g_{\alpha\beta}(\bar{\mathbf{x}}) &= \bar{\partial}_\alpha \bar{r}^i(\bar{\mathbf{x}}) \bar{\partial}_\beta \bar{r}^j(\bar{\mathbf{x}}) = \delta_{\alpha\beta} + \bar{\partial}_\alpha u_\beta + \bar{\partial}_\beta u_\alpha + \\ &\quad \bar{\partial}_\alpha u_\gamma \bar{\partial}_\beta u_\gamma + \bar{\partial}_\rho h \bar{\partial}_\gamma h (\delta_{\alpha\rho} \delta_{\gamma\beta} + \delta_{\alpha\beta} \delta_{\rho\gamma} + \bar{\partial}_\alpha u_\rho \delta_{\beta\lambda} + \bar{\partial}_\alpha u_\rho \bar{\partial}_\beta u_\gamma) \\ &\approx \delta_{\alpha\beta} + \bar{\partial}_\alpha u_\beta + \bar{\partial}_\beta u_\alpha + \bar{\partial}_\alpha h \bar{\partial}_\beta h \end{aligned} \quad (3.48)$$

If only linear terms in \mathbf{u} and the leading term in h are kept the strain tensor Eq. 2 becomes

$$u_{\alpha\beta} = \frac{1}{2} (\bar{\partial}_\alpha u_\beta + \bar{\partial}_\beta u_\alpha + \bar{\partial}_\alpha h \bar{\partial}_\beta h) . \quad (3.49)$$

The actual metric is

$$g_{\alpha\beta}(\bar{\mathbf{x}}) = \bar{g}_{\alpha\beta}(\bar{\mathbf{x}}) + 2u_{\alpha\beta}(\bar{\mathbf{x}}) \quad (3.50)$$

The leading elastic part of the free energy, consistent with the expansion Eq. 3.49 becomes

$$\begin{aligned}\mathcal{F}^{elastic} &= \frac{1}{2} \frac{Y}{1 - \nu_p^2} (\nu_p (u_{\alpha\alpha}^2 + (1 - \nu_p) u_{\alpha\beta} u_{\alpha\beta})) \\ &= \frac{1}{2} (2\mu (u_{\alpha\beta})^2 + \lambda (u_{\alpha\alpha})^2)\end{aligned}\tag{3.51}$$

expressed in terms of the Lamé coefficients λ, μ instead of the Young modulus $Y = \frac{4\mu(\mu+\lambda)}{2\mu+\lambda}$ and Poisson ratio $\nu_p = \frac{\lambda}{2\mu+\lambda}$. For fixed geometry, that is for a given f , this is exactly the same free energy and strains as used in linear elasticity theory, see for example Ref. [121].

The Airy function is the solution to the equation,

$$\frac{1}{Y_0} \bar{\Delta}^2 \chi(\bar{\mathbf{x}}) = s(\bar{\mathbf{x}}) - K(\bar{\mathbf{x}}) ,\tag{3.52}$$

see Ref. [121] for a full derivation. The laplacian $\bar{\Delta}$ refers to a flat metric. Adding an arbitrary disclination density is done by introducing singularities in the reference metric, as discussed for a central disclination in the main text. It is

$$s(\bar{\mathbf{x}}) = \frac{\pi}{3} \sum_{i=1}^N q_i \delta(\bar{\mathbf{x}} - \bar{\mathbf{x}}_i)\tag{3.53}$$

where $\bar{\mathbf{x}}_i$ are the positions of the N disclinations. The Gaussian curvature is obtained by expanding Eq. A3 to leading order, consistent with the expansion in the actual metric Eq. 3.48. Therefore

$$K(\bar{\mathbf{x}}) = -\frac{1}{2} \varepsilon_{\alpha\beta} \varepsilon_{\gamma\rho} \bar{\partial}_\beta \bar{\partial}_\rho (\bar{\partial}_\alpha h \bar{\partial}_\gamma h) .\tag{3.54}$$

We remark that Eq. 3.52 is written in terms of a flat metric, and the only contribution from the curved surface is through the approximated Gaussian curvature Eq. 3.54. Eq. 3.52 has the physical interpretation of the Gaussian curvature screening the disclination density.

The explicit form of the stress tensor is obtained from the Airy function as

$$\begin{aligned}
\sigma^{\rho\rho}(\rho) &= \frac{1}{\rho} \frac{d\chi(\rho)}{d\rho} & (3.55) \\
&= \frac{Y}{2\rho^2} \left(\frac{q_i}{6} \rho^2 \log\left(\frac{\rho}{\rho_0}\right) + \int_0^\rho dvv \int_v^{\rho_0} \frac{du}{u} \left(\frac{df}{du}\right)^2 - \frac{\rho^2}{\rho_0^2} \int_0^{\rho_0} dvv \int_v^{\rho_0} \frac{du}{u} \left(\frac{df}{du}\right)^2 \right) \\
\sigma^{\theta\theta}(\rho) &= \frac{1}{\rho^2} \frac{d^2\chi(\rho)}{d\rho^2} = \frac{1}{\rho^2} \frac{d(\rho\sigma^{\rho\rho}(\rho))}{d\rho} \\
&= \frac{Y}{2\rho^2} \left(\frac{q_i}{6} \log\left(e \frac{\rho}{\rho_0}\right) + \int_\rho^{\rho_0} \frac{du}{u} \left(\frac{df}{du}\right)^2 - \frac{1}{\rho^2} \int_0^\rho dvv \int_v^{\rho_0} \frac{du}{u} \left(\frac{df}{du}\right)^2 \right. \\
&\quad \left. - \frac{1}{\rho_0^2} \int_0^{\rho_0} dvv \int_v^{\rho_0} \frac{du}{u} \left(\frac{df}{du}\right)^2 \right).
\end{aligned}$$

The strain tensor is

$$\begin{aligned}
u_{\rho\rho} &= \frac{1}{Y} \left(\sigma^{\rho\rho} - \nu_p \rho^2 \sigma^{\theta\theta} \right) = \frac{1}{Y} \left(\frac{1}{\rho} \frac{d}{d\rho} - \nu_p \frac{d^2}{d^2\rho} \right) \chi \\
u_{\theta\theta} &= \frac{\rho^2}{Y} \left(\rho^2 \sigma^{\theta\theta} - \nu_p \sigma^{\rho\rho} \right) = \frac{\rho^2}{Y} \left(\frac{d^2}{d^2\rho} - \frac{\nu_p}{\rho} \frac{d}{d\rho} \right) \chi. & (3.56)
\end{aligned}$$

Note that these equations are also equivalent to Eq. 3.49

$$\begin{aligned}
u_{\rho\rho} &= \frac{du_\rho}{d\rho} + \frac{1}{2} \partial_\rho h \partial_\rho h & (3.57) \\
\frac{u_{\theta\theta}}{\rho^2} &= -\Gamma_{\theta\theta}^\rho u_\rho = \frac{u_\rho}{\rho}
\end{aligned}$$

with $r(\rho) = \rho + u_\rho(\rho)$. The free energy is

$$F = \pi \int_0^{\rho_0} d\rho \rho \left(\sigma^{\rho\rho} u_{\rho\rho} + \sigma^{\theta\theta} u_{\theta\theta} \right) . \quad (3.58)$$

Finally, the mapping $r(\rho)$ (or $\rho(r)$) can be obtained from solving either equation

$$\begin{aligned} 2u_{\rho\rho}(\rho) &= (1 + f'(r)^2) \left(\frac{dr}{d\rho} \right)^2 - 1 \\ 2u_{\theta\theta}(\rho) &= r^2(\rho) - \alpha^2 \rho^2 . \end{aligned} \quad (3.59)$$

The explicit solution of the previous equation, consistent at linear order is

$$r(\rho) = \rho - (1 - \alpha)\rho + \frac{u_{\theta\theta}(\rho)}{\rho} = \rho \left(1 - \frac{q_i}{6} + \frac{1}{Y} \left[\frac{d^2}{d\rho^2} - \frac{\nu_p}{\rho} \frac{d}{d\rho} \right] \chi \right) , \quad (3.60)$$

Note that had we used the first equation, the solution

$$\begin{aligned} \int_0^r dr \sqrt{1 + f'(r)^2} &= \int_0^\rho d\rho \sqrt{1 + 2u_{\rho\rho}} \\ r + \frac{1}{2} \int_0^r f'(r)^2 &= \rho + \int_0^\rho d\rho u_{\rho\rho} \\ r &= \rho + \int_0^\rho d\rho \frac{du_\rho}{d\rho} = \rho + u_\rho , \end{aligned} \quad (3.61)$$

where Eq. 3.57 has been used.

Addition of a line tension just adds the boundary condition

$$\sigma^{\rho\rho} = -\frac{\tau}{\rho_0} \quad (3.62)$$

to the stress tensor, and the additional free energy contribution

$$F^{line} = 2\pi\tau r(\rho_0) \equiv 2\pi\tau r_0 . \quad (3.63)$$

Analytical formulas for the Airy function, stress tensor, strain tensor, and free energy for two different surfaces, the spheroid and sombrero are given below.

Spheroid surface

It is given by (with R_0 the spheroid radius, not to be confused with $r_0 = \rho(r_0)$ the coordinate parameterizing the boundary) the equation

$$f(r) = \beta\sqrt{R_0^2 - r^2} . \quad (3.64)$$

The Airy function is

$$\chi = Y \frac{\beta^2}{16\rho_0^2} \left(-R_0^2 \log(R_0^2) \left(\rho^2 - 2\rho_0^2 \log(\rho) \right) \right) \quad (3.65)$$

$$+ \rho_0^2 \left(-\rho^2 - \left(R_0^2 - \rho^2 + R_0^2 \log\left(\frac{\rho^2}{R_0^2}\right) \right) \log(R_0^2 - \rho^2) \right) \\ + \rho^2 (R_0^2 - \rho_0^2) \log(R_0^2 - \rho_0^2) - R_0^2 \rho_0^2 Li_2 \left(1 - \frac{\rho^2}{R_0^2} \right) - \frac{\tau\rho^2}{2r_A} . \quad (3.66)$$

where Li_2 is the polylogarithm function $Li_2(x) = \sum_{n=1}^{\infty} \frac{x^n}{n^2}$.

The stress tensors is

$$\begin{aligned}
\sigma^{\rho\rho} &= Y \frac{1}{24\rho^2\rho_0^2} (3R_0^2\beta^2(\rho_0^2 - \rho^2) \log R_0^2 + 3\beta^2(\rho^2 - R_0^2)\rho_0^2 \log(R_0^2 - \rho^2) + \\
&+ \rho^2(2q_i\rho_0^2 \log\left(\frac{\rho}{\rho_0}\right) + 3\beta^2(R_0^2 - \rho_0^2) \log(R_0^2 - \rho_0^2))) - \frac{\tau}{r_A} \\
\sigma^{\theta\theta} &= Y \frac{1}{24\rho^4\rho^2} (-3R_0^2\beta^2(\rho^2 + \rho_0^2) \log R_0^2 + 3\beta^2(R_0^2 + \rho^2)\rho_0^2 \log(R_0^2 - \rho^2) \\
&+ \rho^2(2\rho_0^2(q_i + 3\beta^2 + q_i \log\left(\frac{\rho}{\rho_0}\right)) + 3\beta^2(R_0^2 - \rho_0^2) \log(R_0^2 - \rho_0^2))) - \frac{\tau}{r_A} \frac{1}{\rho^2} \quad (3.67)
\end{aligned}$$

The strain tensors is

$$\begin{aligned}
u_{\rho\rho} &= \frac{1}{24\rho^2\rho_0^2} (3R_0^2\beta^2((-1 + \nu_p)\rho^2) \log(R_0^2) - 3\beta^2(R_0^2(1 + \nu_p) + \\
&+ (-1 + \nu_p)\rho^2)\rho_0^2 \log(R_0^2 - \rho^2) + \rho^2(-2\rho_0^2(3\nu_p\beta^2 - q_i \log\left(\frac{\rho}{\rho_0}\right) + q_i\nu_p \log\left(\frac{e\rho}{\rho_0}\right)) + \\
&+ 3(-1 + \nu_p)\beta^2(\rho_0^2 - R_0^2) \log(R_0^2 - \rho_0^2))) + (\nu_p - 1) \frac{\tau}{Yr_A} \\
u_{\theta\theta} &= \frac{1}{24\rho_0^2} (3R_0^2\beta^2((-1 + \nu_p)\rho^2 - (1 + \nu_p)\rho_0^2) \log(R_0^2) + 3\beta^2(R_0^2(1 + \nu_p) - \\
&- (-1 + \nu_p)\rho^2)\rho_0^2 \log(R_0^2 - \rho^2) + \rho^2(2\rho_0^2(q_i + 3\beta^2 - q_i(-1 + \nu_p) \log\left(\frac{\rho}{\rho_0}\right) + \\
&+ 3(-1 + \nu_p)\beta^2(\rho_0^2 - R_0^2) \log(R_0^2 - \rho_0^2))) + \rho^2(\nu_p - 1) \frac{\tau}{Yr_A} . \quad (3.68)
\end{aligned}$$

If we expand $f(r)$ in powers of the spheroid radius R_0 ,

$$f(r) = \beta R_0 \left(1 - \frac{r^2}{2R_0^2}\right) . \quad (3.69)$$

The Airy function is

$$\chi = -\frac{Y\beta^2(\rho^4 - 2\rho^2\rho_0^2)}{64R_0^2} + \frac{Yq_i}{24}\rho^2 \left(\ln\left(\frac{\rho}{\rho_0}\right) - \frac{1}{2} \right) - \frac{\tau\rho^2}{2r_A}. \quad (3.70)$$

The stress tensors is

$$\sigma^{\rho\rho} = \frac{Y\beta^2}{16R_0^2}(\rho_0^2 - \rho^2) + \frac{Y}{2\rho^2} \left(\frac{q_i}{6}\rho^2 \log\left(\frac{\rho}{\rho_0}\right) \right) - \frac{\tau}{r_A} \quad (3.71)$$

$$\sigma^{\theta\theta} = \frac{Y\beta^2}{16R_0^2\rho^2}(\rho_0^2 - 3\rho^2) + \frac{Y}{2\rho^2} \left(\frac{q_i}{6} \log\left(e\frac{\rho}{\rho_0}\right) \right) - \frac{\tau}{r_A} \frac{1}{\rho^2}. \quad (3.72)$$

The strain tensors is

$$\begin{aligned} u_{\rho\rho} &= \frac{\beta^2}{16R_0^2} \left((3\nu_p - 1)\rho^2 - (\nu_p - 1)\rho_0^2 \right) + \frac{1}{12}q_i \log\left(\frac{\rho}{\rho_0}\right) - \frac{1}{12}q_i\nu_p \log\left(\frac{e\rho}{\rho_0}\right) \\ &\quad + (\nu_p - 1)\frac{\tau}{Yr_A} \\ u_{\theta\theta} &= \frac{\beta^2\rho^2}{16R_0^2} \left((\nu_p - 3)\rho^2 - (\nu_p - 1)\rho_0^2 \right) + \frac{\rho^2}{12} \left(-\nu_p q_i \log\left(\frac{\rho}{\rho_0}\right) + q_i \log\left(\frac{e\rho}{\rho_0}\right) \right) \\ &\quad + \rho^2(\nu_p - 1)\frac{\tau}{Yr_A}. \end{aligned} \quad (3.73)$$

The free energy is

$$F = \pi \left(Y \left(\frac{q_i^2\rho_0^2}{288} - \frac{q_i^2\beta^2\rho_0^4}{192R_0^2} + \frac{\beta^4\rho_0^6}{384R_0^4} \right) + \frac{(1 - \nu_p)\rho_0^2\tau^2}{Yr_A^2} \right). \quad (3.74)$$

Sombrero surface

It is given by the equation

$$f(r) = \frac{\beta R_0}{3} \left(1 - \left(\frac{r}{R_0} \right)^2 + \left(\frac{r}{R_0} \right)^4 \right)^{3/2} \quad (3.75)$$

where r is the radius of the sombrero surface. The Airy function is

$$\begin{aligned} \chi = & -\frac{Y\beta^2\rho^2}{57600R_0^{10}} \left(900R_0^8(\rho^2 - 2\rho_0^2) - 1000R_0^6(\rho^4 - 3\rho_0^4) + 675R_0^4(\rho^6 - 4\rho_0^6) - \right. \\ & \left. - 288R_0^2(\rho^8 - 5\rho_0^8) + 80(\rho^{10} - 6\rho_0^{10}) \right) - \frac{\tau\rho^2}{2r_A} \end{aligned} \quad (3.76)$$

The stress tensor is

$$\begin{aligned} \sigma^{\rho\rho} = & \frac{Y\beta^2}{480} \left(\frac{30(-\rho^2 + \rho_0^2)}{R_0^2} + \frac{50(\rho^4 - \rho_0^4)}{R_0^4} + \frac{45(-\rho^6 + \rho_0^6)}{R_0^6} + \frac{24(\rho^8 - \rho_0^8)}{R_0^8} + \right. \\ & \left. + \frac{8(-\rho^{10} + \rho_0^{10})}{R_0^{10}} \right) + \frac{Yq_i}{12} \log \left(\frac{\rho}{\rho_0} \right) - \frac{\tau}{r_A} \end{aligned} \quad (3.77)$$

$$\begin{aligned} \sigma^{\theta\theta} = & \frac{Y\beta^2}{480\rho^2} \left(\frac{30(-3\rho^2 + \rho_0^2)}{R_0^2} + \frac{50(5\rho^4 - \rho_0^4)}{R_0^4} - \frac{45(7\rho^6 - \rho_0^6)}{R_0^6} + \frac{24(9\rho^8 - \rho_0^8)}{R_0^8} + \right. \\ & \left. + \frac{8(-11\rho^{10} + \rho_0^{10})}{R_0^{10}} \right) + \frac{Yq_i}{12\rho^2} \log \left(\frac{e\rho}{\rho_0} \right) - \frac{\tau}{r_A\rho^2} \end{aligned} \quad (3.78)$$

The strain tensor is

$$\begin{aligned}
u_{\rho\rho} &= \frac{\beta^2}{480} \left(\frac{30((-1+3\nu_p)\rho^2 - (-1+\nu_p)\rho_0^2)}{R_0^2} - 50 \frac{(-1+5\nu_p)\rho^4 - (-1+\nu_p)\rho_0^4}{R_0^4} + \right. \\
&+ \frac{45(-1+7\nu_p)\rho^6 - (-1+\nu_p)\rho_0^6}{R_0^6} - \frac{24(-1+9\nu_p)\rho^8 - (-1+\nu_p)\rho_0^8}{R_0^8} + \\
&+ \left. \frac{8((-1+11\nu_p)\rho^{10} - (-1+\nu_p)\rho_0^{10})}{R_0^{10}} \right) + \frac{q_i}{12} \left(\log \left(\frac{\rho}{\rho_0} \right) - \nu_p \log \left(\frac{e\rho}{\rho_0} \right) \right) + (3.79) \\
&+ \frac{(-1+\nu_p)\tau}{Yr_A} \\
u_{\theta\theta} &= \frac{\rho^2\beta^2}{480} \left(\frac{30((-3+\nu_p)\rho^2 - (-1+\nu_p)\rho_0^2)}{R_0^2} - \frac{50((-5+\nu_p)\rho^4 - (-1+\nu_p)\rho_0^4)}{R_0^4} + \right. \\
&+ \frac{45((-7+\nu_p)\rho^6 - (-1+\nu_p)\rho_0^6)}{R_0^6} - \frac{24((-9+\nu_p)\rho^8 - (-1+\nu_p)\rho_0^8)}{R_0^8} + \\
&+ \left. \frac{8((-11+\nu_p)\rho^{10} - (-1+\nu_p)\rho_0^{10})}{R_0^{10}} \right) + \frac{q_i}{12} \left(\log \left(\frac{e\rho}{\rho_0} \right) - \nu_p \log \left(\frac{\rho}{\rho_0} \right) \right) + (3.80) \\
&+ \frac{(-1+\nu_p)\tau\rho^2}{Yr_A}
\end{aligned}$$

The free energy is

$$\begin{aligned}
F &= \pi \left(Y \left(\frac{q_i^2 \rho_0^2}{288} - \frac{q_i \beta^2 \rho_0^4 (900R_0^8 - 2000R_0^6 \rho_0^2 + 2025R_0^4 \rho_0^4 - 1152R_0^2 \rho_0^6 + 400R_0^8)}{172800R_0^{10}} + \right. \right. \\
&+ \frac{\beta^4 \rho_0^6}{8870400R_0^{20}} (23100R_0^{16} - 115500R_0^{14} \rho_0^2 + 278740R_0^{12} \rho_0^4 - 420420R_0^{10} \rho_0^6 + \\
&+ \left. \left. 438075R_0^8 \rho_0^8 - 326480R_0^6 \rho_0^{10} + 171248R_0^4 \rho_0^{12} - 59136R_0^2 \rho_0^{14} + 11200\rho_0^{16}) \right) - (3.81) \right. \\
&- \left. \frac{(-1+\nu_p)\rho_0^2 \tau^2}{Yr_A^2} \right)
\end{aligned}$$

3.3.8 Theory of defects, inverse Laplacian square

We finally note that Eq. 3.52 can be promoted to an equation

$$\Delta^2 \chi(\mathbf{x}) = s(\mathbf{x}) - K(\mathbf{x}) \quad (3.82)$$

	$\beta = 1$			$\beta = 2$			$\beta = 3$		
	$\frac{\hat{A}}{R_0^2}$	θ	θ_{AH}	$\frac{\hat{A}}{R_0^2}$	θ	θ_{AH}	$\frac{\hat{A}}{R_0^2}$	θ	θ_{AH}
Spheroid	2.13	0.75	0.73	0.68	0.43	0.36	0.32	0.29	0.27
Sombrero	5.71	1.18	1.17	3.83	1.00	0.89	0.42	0.34	0.35

Table 3.1: Transition points at which the surface with ($q_i = 1$) and without ($q_i = 0$) a disclination have the same energy. Here $\theta = \frac{r}{R_0}$ and θ_{AH} are the predictions reported in Ref. [1].

in terms of the actual metric. In this case, the values used for the Gaussian curvature and the disclination density are covariant and exact. This is the starting point of the effective theory of defects. For rotational symmetric cases, it is possible to solve the equation exactly, at least by numerical integration.

3.4 Results

Generally, we find that the applicability of elasticity theory extends to relatively large curvatures ($\frac{\hat{A}}{R_0^2} \approx 1$). For the spheroid, linear elasticity remains qualitatively correct for the entire range explored, but this is not the case for the sombrero surface, see Fig. 3.3, where linear elasticity breaks down and cannot be extended beyond a certain limit.

The points where the $q_i = 0$ and $q_i = 1$ curves cross each other define whether it is energetically favorable to have a disclination at the center or not. The results are quoted in Table 3.1. We compare with the most recent, and to our knowledge, most accurate predictions from Ref. [1]. Note that in some cases the differences are as high as 10%.

The case of a finite line tension is shown in Fig. 3.4. Basically, a large value of τ overshadows all the other energies of the system resulting into the collapse of all the free

energy plots into an almost universal curve defined by the line tension term. The overall free energy is in some agreement with linear elasticity theory, which, as shown in the inset, is also true for the elastic contribution, see Eq. 3.1. We provide similar plots for smaller values of the line tension (see Fig.3.6).

In the absence of line tension, linear elasticity predicts that the free energy is a function of the Young modulus Y alone, independent of the Poisson ratio ν_p in Sec. 3.3.7. This is implicitly assumed in models of interacting defects. Within the exact non-linear theory presented in this chapter, we show that the dependence of the Poisson ratio at vanishing line tension is, indeed, negligible (see Fig. 3.7). However, Fig. 3.5 reveals that there is a dependence of the Poisson ratio whenever the line tension is non-zero, see also Fig. 3.7. We have been able to extract explicitly the dependence of $\rho(r)$ or $r(\rho)$ on the Poisson ratio as the solution of the exact theory, see Figs. 3.2, 3.8 and 3.9.

As expected, for $\frac{\hat{A}}{R_0^2}$ small, the results are well described by linear elasticity, but as this value increases, they get progressively worse, and in the case of the sombrero, linear elasticity breaks down for sufficiently large values of $\frac{\hat{A}}{R_0^2}$.

3.5 Discussion

In summary, this chapter provides an exact solution to the problem of determining the structure of crystals on a curved surface by formulating the problem in terms of geometric invariants, which besides connecting it to the well developed field of differential

geometry of curves and surfaces, enables, what we believe, is a transparent interpretation of non-linear elasticity theory. In addition, we show that effects that have been difficult to consider in the past, such as a finite line tension or dependence on the Poisson ration ν_p are easily included.

Our exact solution provides a *universal triangulation*, as shown in Fig. 3.2, that is, a solution to the problem of providing the optimal tiling of an arbitrary surface with triangles as close to equilateral as possible. In previous studies, determining particle positions requires numerical minimization methods, either through discretizations of elasticity theory [121, 135, 85] or using explicit potentials, most typically inverse power laws, on a sphere [108, 16] and other geometries [9] as well. The problem with numerical minimizations is the cost and instabilities that appear for both a large number of particles N and/or complicated geometries. We note that our analytical approach does not suffer from any of these problems: it is independent of the number of particles N , see Eq. 3.3, and is stable for any differentiable surface.

The approach developed in this chapter allows us to understand how the pentameric defects appear and interact with each other with clear implications in viral shells, specifically for the difficult cases of the assembly of nonspherical structures similar to those presented in Fig. 3.1. Determining the location of lattice defects in the growing shells with non-zero Gaussian curvature in the presence of boundaries with line tension for various values of Poisson ratio has been shown to be a very challenging task [61, 34, 47, 102]. The theory developed here paves the path for tackling the problem of crystalline growth pathways. While in this chapter we restricted our study to a fixed geometry with a given

number of particles and provided explicit solutions for problems with rotational symmetry, the approach is completely general for any geometry or number of disclinations, although its explicit description requires additional developments that will be provided in subsequent studies.

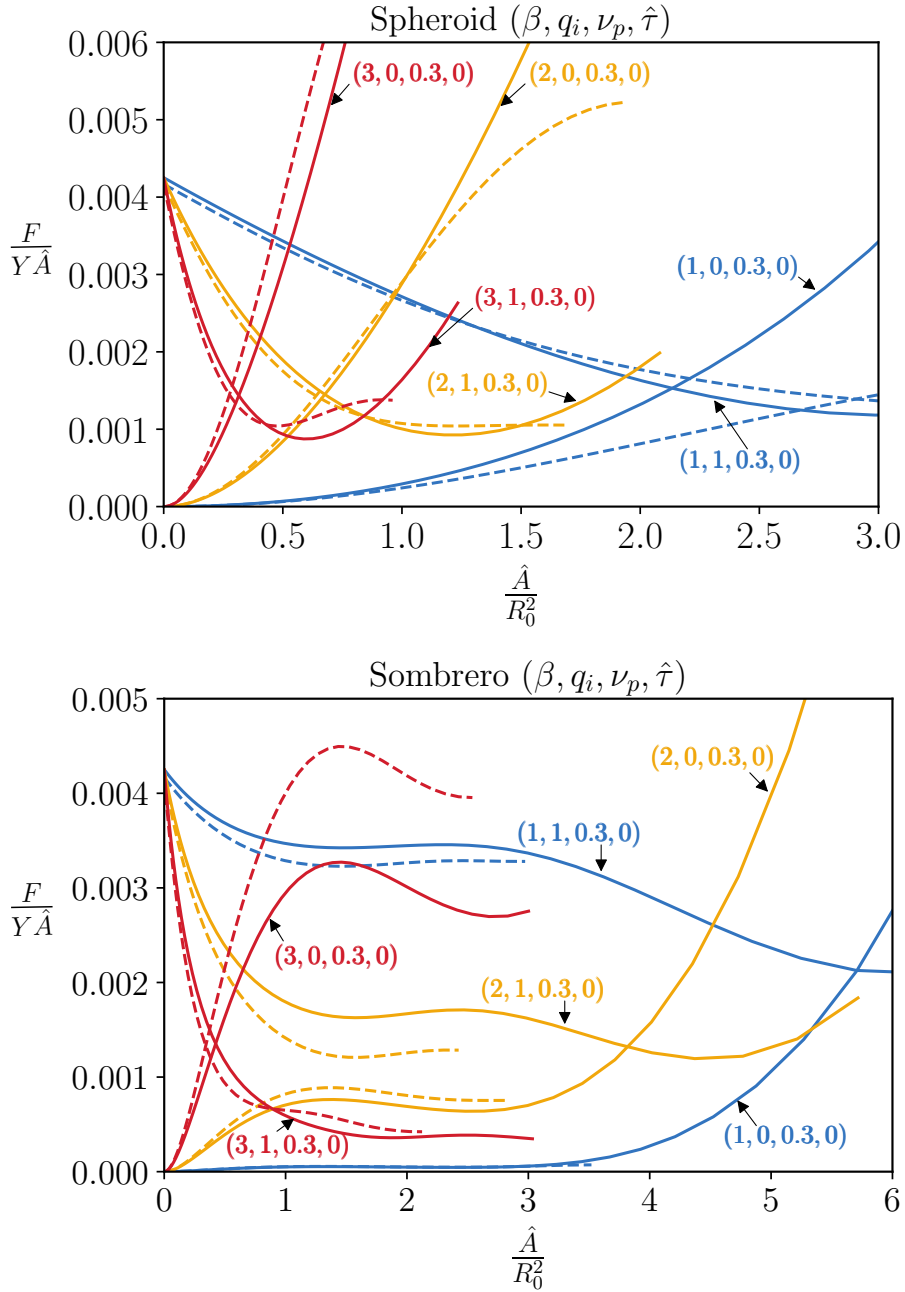


Figure 3.3: Free energy for a spheroid and sombrero without ($q_i = 0$) and with ($q_i = 1$) a disclination at the center at $\frac{\tau}{Y R_0} \equiv \hat{\tau} = 0$ (zero line tension) and at fixed Poisson ratio $\nu_p = 0.3$. The solid line corresponds to the exact results while the dashed line denotes the analytical results within linear elasticity. The three different colors represent different values of $\beta = 1, 2, 3$, indicating the magnitude of $f(r)$.

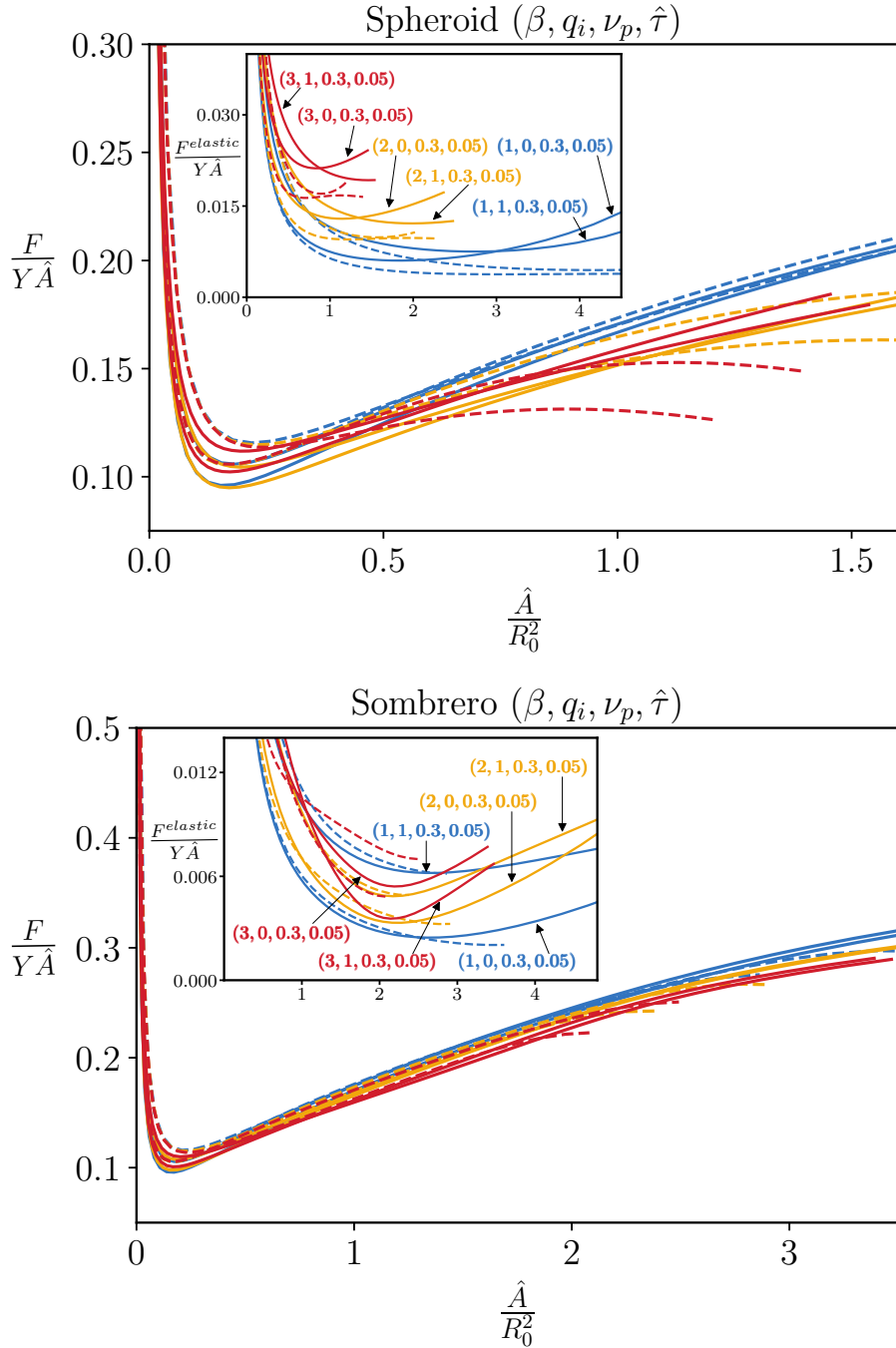


Figure 3.4: Free energy (Eq. 3.1) for both the spheroid and sombrero at finite line tension $\frac{\tau}{YR_0} = \hat{\tau} = 0.05$. The solid line corresponds to the exact results while the dashed line denotes the analytical results within linear elasticity.

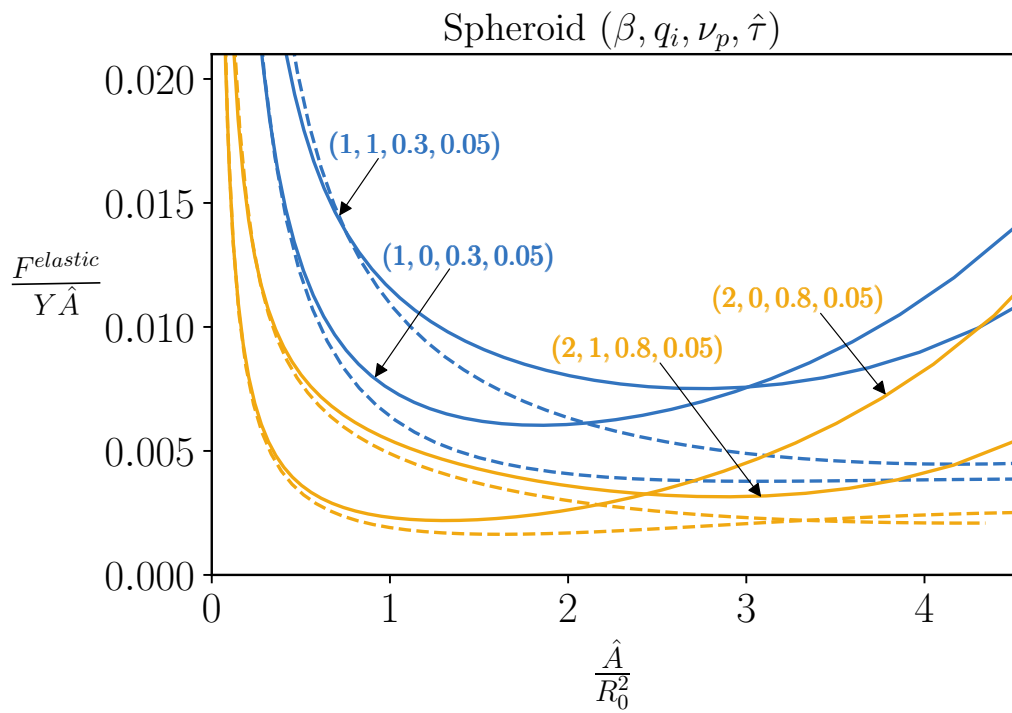


Figure 3.5: The elastic energy (see Eq. 3.1) for the spheroid at $\frac{\tau}{Y R_0} = \hat{\tau} = 0.05$ for two different values of the Poisson ratio $\nu_p = 0.3$ and 0.8 . The solid lines correspond to the exact results while the dashed lines denote analytical results within linear elasticity.

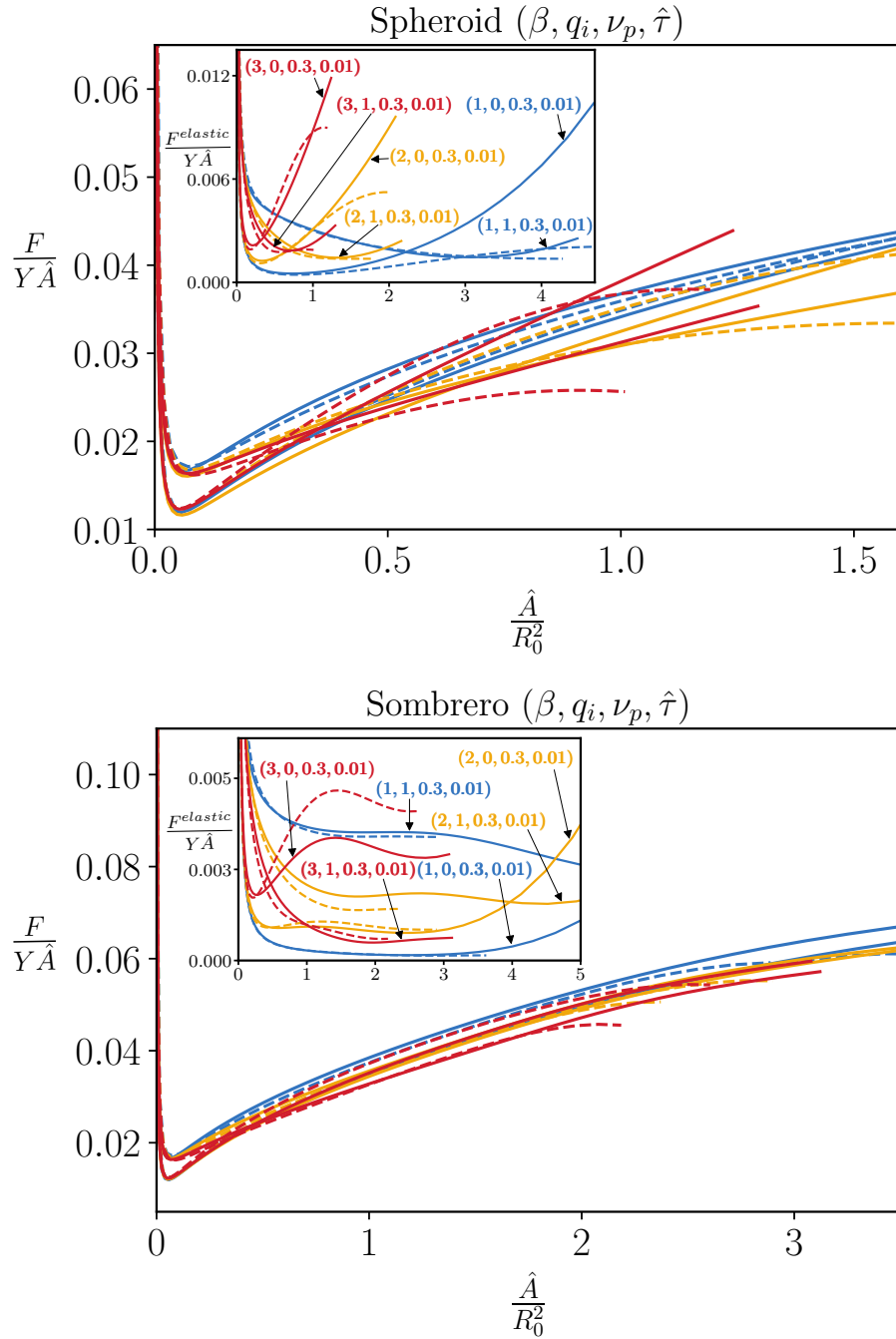


Figure 3.6: Free energy Eq. 1 for a spheroid and sombrero at $\hat{\tau} = 0.01$. The solid lines correspond to the exact results, while the dashed lines denote analytical results within linear elasticity.

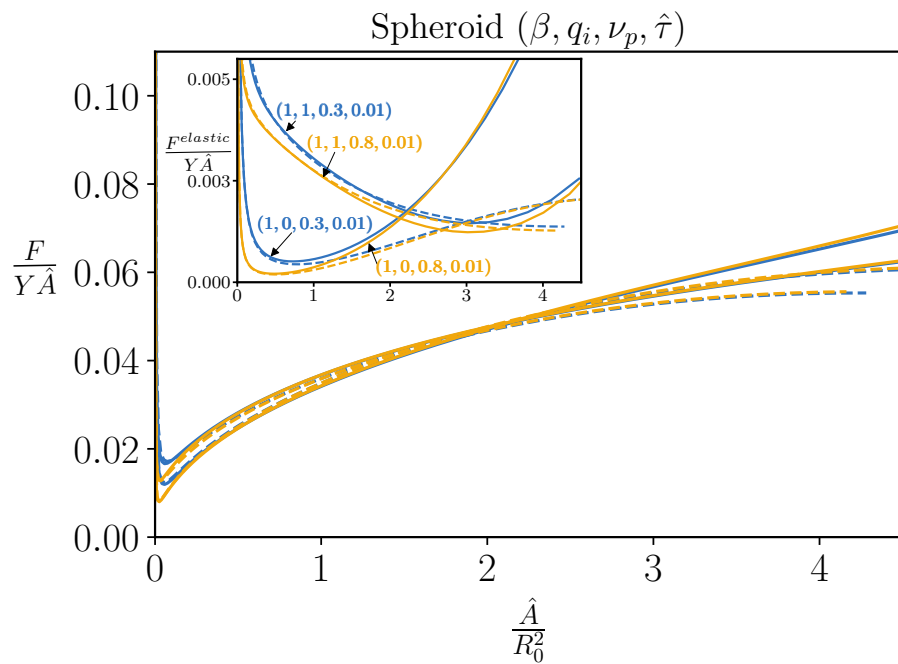
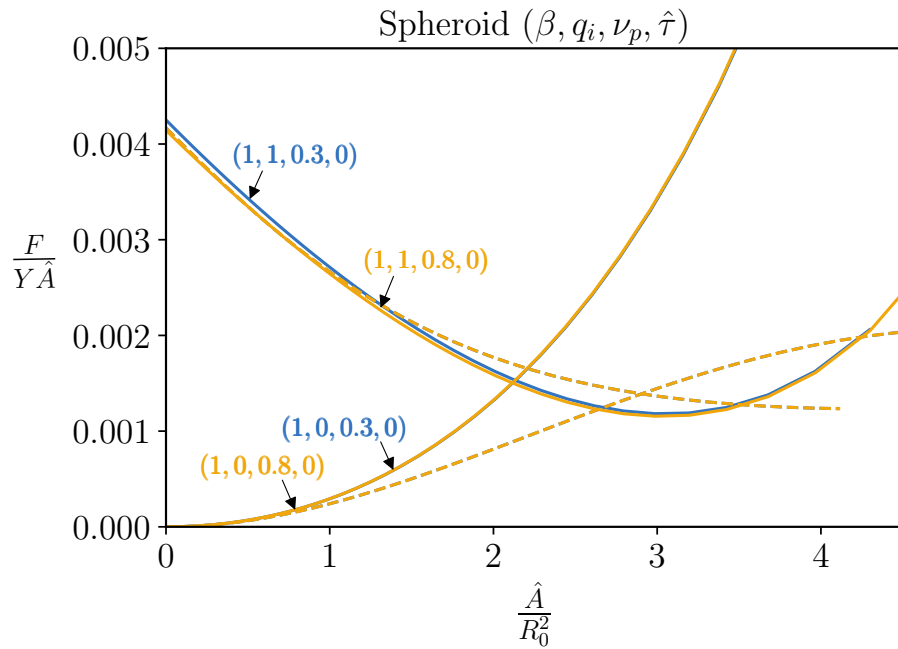


Figure 3.7: Free energy Eq. 1 for the spheroid at different Poisson ratios $\nu_p = 0.3$ and 0.8 . The solid lines correspond to the exact results while the dashed lines denote analytical results within linear elasticity.

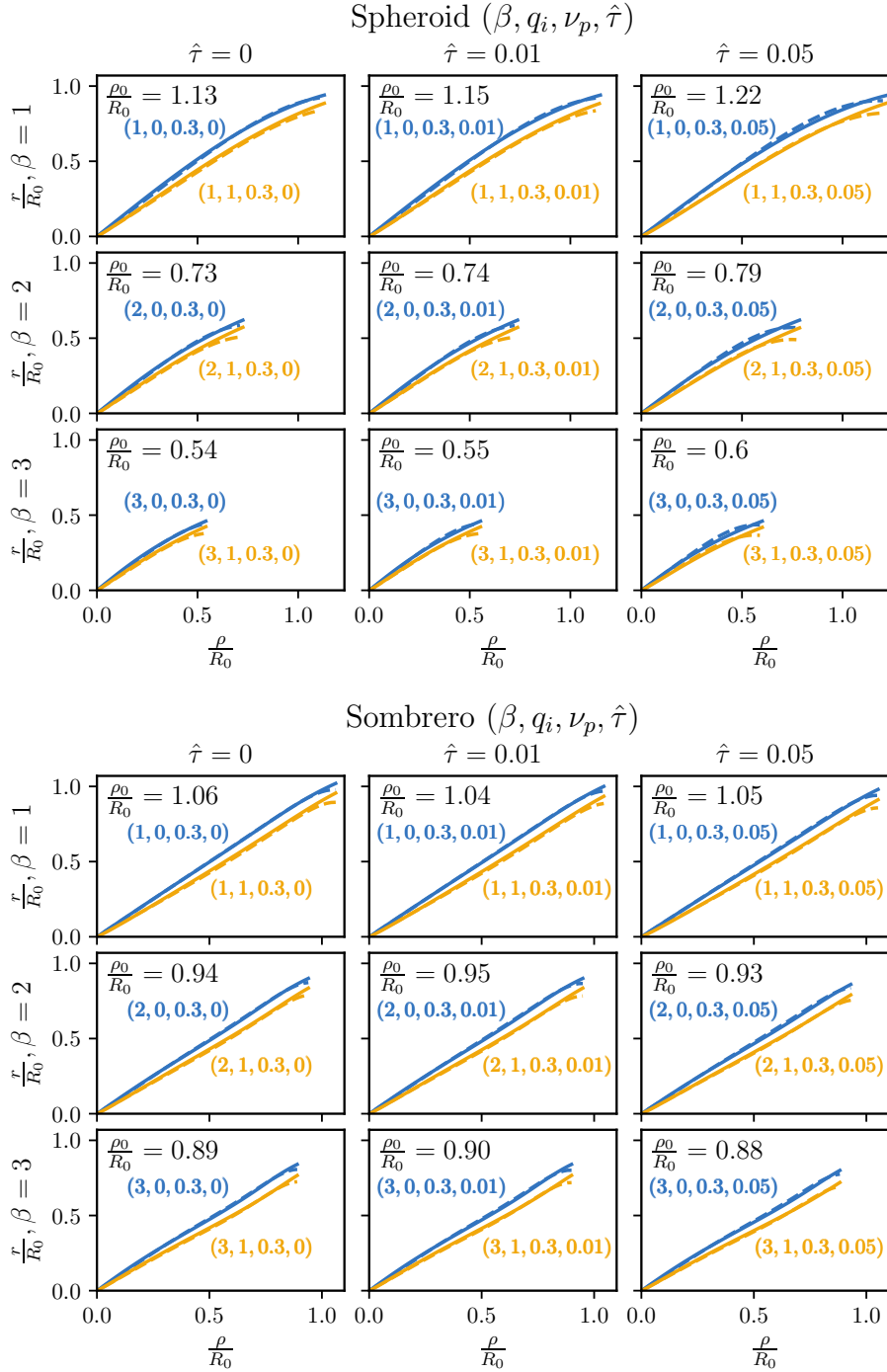


Figure 3.8: Solution Eq. B4 and Eq. 3.60 for the spheroid and sombrero relatively small ρ_0 compared with FIG. 3.9. The solid lines correspond to the exact result while the dashed lines denote analytical results within linear elasticity.

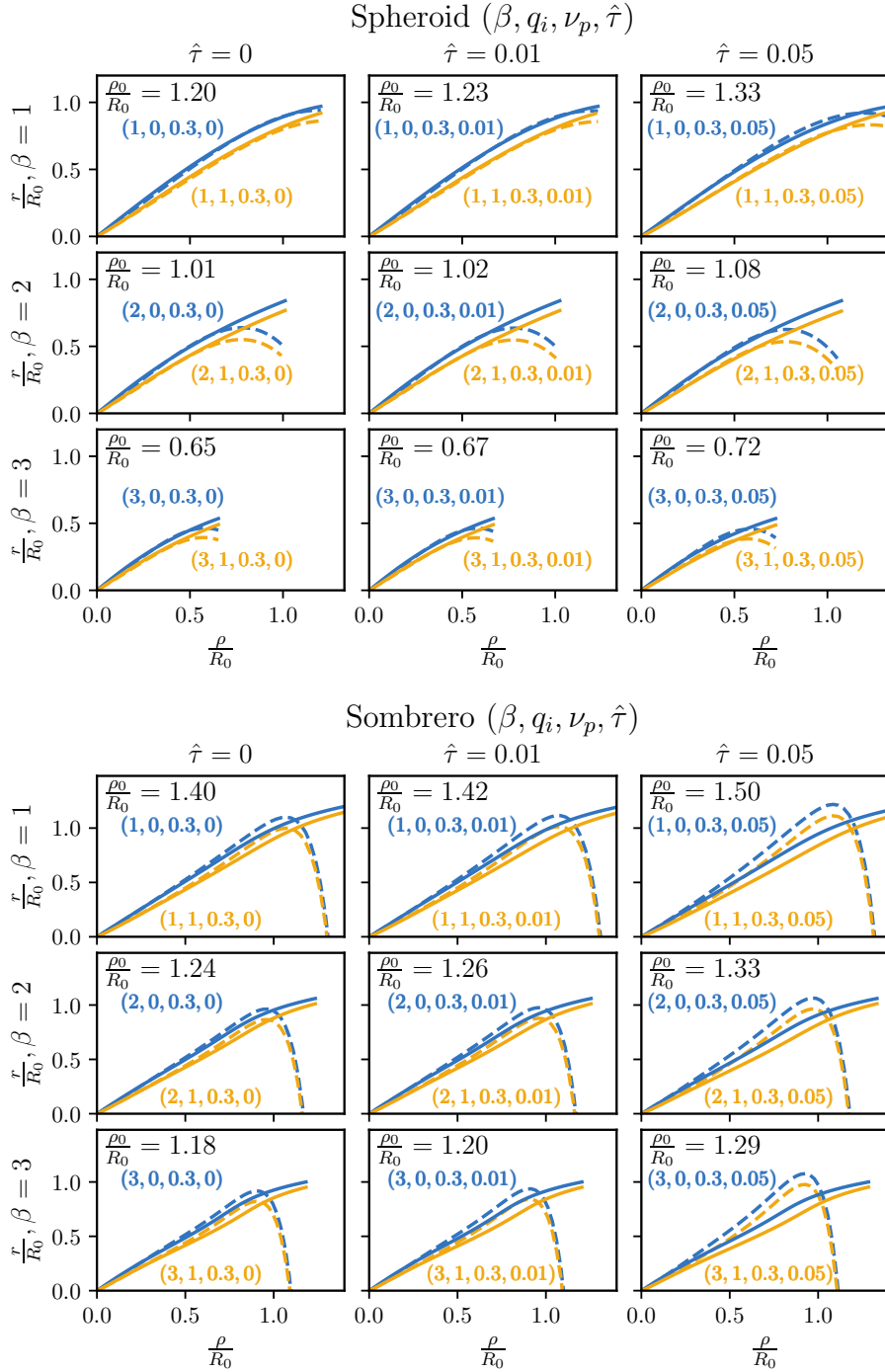


Figure 3.9: Solution Eq. B4 and Eq. 3.60 for the spheroid and sombrero with relatively large ρ_0 compared with FIG. 3.8. The solid lines correspond to the exact results while the dashed lines denote analytical results within linear elasticity.

Chapter 4

Conclusions

In this thesis, we investigate the physical principles underlying the formation of virus capsid assembly and genome packaging.

In the first chapter, we focus on understanding the process of RNA packaging by capsid proteins. Despite the fact that many experiments have shown that the number of nucleotides packaged by capsid proteins increases with the number of charges on N-terminal tails, how the amount of encapsidated RNA depends on the distribution of the charges along and the length of the N-terminal domain of capsid proteins is not well understood. By employing a mean-field theory incorporating electrostatics, excluded volume interaction, and RNA conformational entropy, we explore the variation in RNA packaging observed in BMV mutants observed in the experiments of Ni *et al.* [99]. Our calculations of the free energy of RNA confined in a spherical shell interacting with N-terminal tails and ions reveal that the combined effect of electrostatic interactions and genome entropy provides valuable insights into BMV assembly. We find that the location and distance between charges along

N-terminal tails significantly influence the amount of packaged RNA. By comparing the structure of CCMV nucleocapsids packaging various RNA segments with similar lengths but varying levels of compactness, our calculations further verify that the RNA segments with a greater number of branch points exhibit a lowered free energy.

However, it is essential to note that while our mean-field theory can explain many experimental observations, it cannot account for all the data. To achieve a better quantitative comparison with experimental results, we highlight the importance of considering additional factors such as counter-ion condensation, divalent ions, protein structure, and packaging signals.

The second chapter of the thesis focus on the elegant simplicity of virus structures and their connection to elasticity theory and defect theories based on topology and differential geometry. The chapter addresses the challenge of characterizing the structure of an elastic network confined to a curved surface, which arises in various scientific disciplines. Through a novel approach using geometric invariants, the chapter provides an exact solution for determining crystal structures on curved surfaces.

By providing an exact solution for determining crystal structures on curved surfaces, we formulate the problem in terms of geometric invariants, enabling a transparent interpretation of nonlinear elasticity theory. We demonstrate the inclusion of complex effects, such as finite line tension and dependence on the Poisson ratio, in our method. Our exact solution offers a universal triangulation, providing an optimal tiling of arbitrary surfaces with triangles as close to equilateral as possible. Importantly, our analytical approach overcomes the drawbacks of numerical minimization methods, providing stability and in-

dependence from the number of particles and surface complexity. The implications of our findings extend to understanding pentameric defects in viral shells and the assembly of non-spherical structures. By tackling the challenge of determining the location of lattice defects in growing shells with non-zero Gaussian curvature and boundaries with line tension, our theory paves the path for investigating crystalline growth pathways.

In conclusion, this thesis makes significant contributions to our understanding of virus capsid assembly, crystalline growth pathways, and genome packaging. The insights gained from exploring RNA packaging provide potential strategies for blocking viral infections or building capsids for gene therapy applications. Additionally, the exact solution for elastic networks on curved surfaces paves the path for tackling the problem of crystalline growth pathways and understanding how pentameric defects appear in nonspherical viral shells. By combining fundamental principles of physics and mathematical tools, this work sets the stage for further investigations and potential applications in various fields.

Bibliography

- [1] Siddhansh Agarwal and Sascha Hilgenfeldt. Simple, General Criterion for Onset of Disclination Disorder on Curved Surfaces. *Physical Review Letters*, 125(7):078003, 8 2020.
- [2] Akhtar Ali and Marilyn J. Roossinck. Rapid and efficient purification of cowpea chlorotic mottle virus by sucrose cushion ultracentrifugation. *J. Virol. Methods*, 141(1):84–86, 2007.
- [3] Amir Azadi and Gregory M Grason. Emergent Structure of Multidislocation Ground States in Curved Crystals. *Physical Review Letters*, 112(22):225502, 6 2014.
- [4] J. B. Bancroft. The self-assembly of spherical plant viruses. *Adv. Virus Res.*, 16:99, 1970.
- [5] J.B. Bancroft, G.J. Hills, and Rou Markham. A study of the self-assembly process in a small spherical virus formation of organized structures from protein subunits in vitro. *Virology*, 31(2):354–379, 1967.
- [6] W. Bangerth, R. Hartmann, and G. Kanschat. deal.II – a general purpose object oriented finite element library. *ACM Trans. Math. Softw.*, 33(4):24/1–24/27, 2007.
- [7] K.J. Bathe. *Finite Element Procedures*. Number pt. 2 in Finite Element Procedures. Prentice Hall, New Jersey, 1996.
- [8] V. A. Belyi and M. Muthukumar. Electrostatic origin of the genome packing in viruses. *PNAS*, 103:17174, 2006.
- [9] Enrique Bendito, Mark J Bowick, Agustin Medina, and Zhenwei Yao. Crystalline particle packings on constant mean curvature (Delaunay) surfaces. *Phys. Rev. E*, 88(1):12405, 7 2013.
- [10] Christian Beren, Yanxiang Cui, Antara Chakravarty, Xue Yang, A. L. N. Rao, Charles M. Knobler, Z. Hong Zhou, and William M. Gelbart. Genome organization and interaction with capsid protein in a multipartite RNA virus. *Proc. Natl. Acad. Sci. U.S.A.*, 117(20):10673–10680, 2020. Number: 20.
- [11] Christian Beren, Lisa L. Dreesens, Katherine N. Liu, Charles M. Knobler, and William M. Gelbart. The effect of RNA secondary structure on the self-assembly of viral capsids. *Biophys. J.*, 113(2):339–347, 2017.

- [12] Christian Beren, Lisa L. Dreesens, Katherine N. Liu, Charles M. Knobler, and William M. Gelbart. The Effect of RNA Secondary Structure on the Self-Assembly of Viral Capsids. *Biophysical Journal*, 113(2):339–347, 2017.
- [13] Alexander Borodavka, Surendra W. Singaram, Peter G. Stockley, William M. Gelbart, Avinoam Ben-Shaul, and Roman Tuma. Sizes of long rna molecules are determined by the branching patterns of their secondary structures. *Biophysical Journal*, 111(10):2077 – 2085, 2016.
- [14] I. Borukhov, D. Andelman, and H. Orland. Polyelectrolyte Solutions between Charged Surfaces. *Europhys. Lett.*, 32:499, 1995.
- [15] I. Borukhov, D. Andelman, and H. Orland. Random polyelectrolytes and polyampholytes in solution. *Euro. Phys. J. B*, 5:869, 1998.
- [16] M. Bowick, A. Cacciuto, D. R. Nelson, and A. Travasset. Crystalline Order on a Sphere and the Generalized Thomson Problem. *Physical Review Letters*, 89(18):185502, 10 2002.
- [17] Mark J. Bowick, David R. Nelson, and Alex Travasset. Interacting topological defects on frozen topographies. *Physical Review B*, 62(13):8738–8751, 10 2000.
- [18] I. Breßler, J. Kohlbrecher, and A. F. Thünemann. SASfit: A tool for small-angle scattering data analysis using a library of analytical expressions. *J. Appl. Cryst.*, 48(5):1587–1598, 2015.
- [19] R. F. Bruinsma, M. Comas-Garcia, R. F. Garmann, and A. Y. Grosberg. Equilibrium self-assembly of small RNA viruses. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, 93(3-1):032405, 2016. Number: 3-1.
- [20] Robijn F. Bruinsma and William S. Klug. Physics of viral shells. *Annu. Rev. Condens. Matter Phys.*, 6:245–268, 2015.
- [21] Robijn F. Bruinsma, Gijs J. L. Wuite, and Wouter H. Roos. Physics of viral dynamics. *Nat. Rev. Phys.*, pages 1–16, 2021.
- [22] R. D. Cadena-Nava, M. Comas-Garcia, R. F. Garmann, A. L. N. Rao, C. M. Knobler, and W. M. Gelbart. Self-Assembly of Viral Capsid Protein and RNA Molecules of Different Sizes: Requirement for a Specific High Protein/RNA Mass Ratio. *J. Virol.*, 86:3318, 2012.
- [23] 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*, 115:2386, 2011.
- [24] Ruben D. Cadena-Nava, Mauricio Comas-Garcia, Rees F. Garmann, A L N. Rao, Charles M. Knobler, and William M. Gelbart. Self-assembly of viral capsid protein and RNA molecules of different sizes: Requirement for a specific high protein/RNA mass ratio. *J. Virol.*, 86(6):3318–3326, 2012.

- [25] D. L. Caspar and A. Klug. Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.*, 27:1–24, 1962.
- [26] Donald LD Caspar and Aaron Klug. Physical principles in the construction of regular viruses. *Cold Spring Harbor Symp. Quant. Biol.*, 27:1–24, 1962.
- [27] Martin Castelnovo. Viral self-assembly pathway and mechanical stress relaxation. *Physical Review E*, 95(5):052405, 5 2017.
- [28] Pablo Ceres and Adam Zlotnick. Weak protein-protein interactions are sufficient to drive assembly of hepatitis B virus capsids. *Biochemistry*, 41(39):11525–11531, 2002. Number: 39.
- [29] Paul M Chaikin, Tom C Lubensky, and Thomas A Witten. *Principles of condensed matter physics*, volume 10. Cambridge university press Cambridge, 1995.
- [30] Jingzhi Chen, Yves Lansac, and Guillaume Tresset. Interactions between the molecular components of the cowpea chlorotic mottle virus investigated by molecular dynamics simulations. *J. Phys. Chem. B*, 122(41):9490–9498, 2018. Number: 41.
- [31] Ting Chen, Zhenli Zhang, and Sharon C. Glotzer. A precise packing sequence for self-assembled convex structures. *PNAS*, 104:717, 2007.
- [32] Yifan Cheng, Werner Boll, Tomas Kirchhausen, Stephen C. Harrison, and Thomas Walz. Cryo-electron tomography of clathrin-coated vesicles: Structural implications for coat assembly. *Journal of Molecular Biology*, 365(3):892–899, 2007.
- [33] M. Chevreuil, D. Law-Hine, J. Chen, S. Bressanelli, S. Combet, D. Constantin, J. Degrouard, J. Möller, M. Zeghal, and G. Tresset. Nonequilibrium self-assembly dynamics of icosahedral viral capsids packaging genome or polyelectrolyte. *Nat. Commun.*, 9:3071, 2018.
- [34] Maelenn Chevreuil, Didier Law-Hine, Jingzhi Chen, Stéphane Bressanelli, Sophie Combet, Doru Constantin, Jéril Degrouard, Johannes Möller, Mehdi Zeghal, and Guillaume Tresset. Nonequilibrium self-assembly dynamics of icosahedral viral capsids packaging genome or polyelectrolyte. *Nat. Commun.*, 9(1):3071, 2018.
- [35] Yoon Gi Choi and A. L. N. Rao. Packaging of brome mosaic virus RNA3 is mediated through a bipartite signal. *J. Virol.*, 77(18):9750–9757, 2003.
- [36] 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.*, 86:12271, 2012.
- [37] Mauricio Comas-Garcia. Packaging of genomic RNA in positive-sense single-stranded RNA viruses: A complex story. *Viruses*, 11(3):253, 2019. Number: 3.
- [38] Mauricio Comas-Garcia, Ruben D. Cadena-Nava, A. L. N. Rao, Charles M. Knobler, and William M. Gelbart. In vitro quantification of the relative packaging efficiencies of

- single-stranded RNA molecules by viral capsid protein. *J. Virol.*, 86(22):12271–12282, 2012. Number: 22.
- [39] Mauricio Comas-Garcia, Rees F. Garmann, Surendra W. Singaram, Avinoam Ben-Shaul, Charles M. Knobler, and William M. Gelbart. Characterization of viral capsid protein self-assembly around short single-stranded RNA. *J. Phys. Chem. B*, 118(27):7510–7519, 2014.
- [40] R.A. Crowther, J.T. Pinch, and B.M.F. Pearse. On the structure of coated vesicles. *Journal of Molecular Biology*, 103(4):785–798, 1976.
- [41] P.-G. de Gennes. *Scaling concepts in polymer physics*. Cornell University Press, Ithaca, New York, 1979.
- [42] P.-G. de Gennes. Polymers at an Interface. 2. Interaction between Two Plates Carrying Adsorbed Polymer Layers. *Macromolecules*, 15:492, 1982.
- [43] Johan A. den Boon, Arturo Diaz, and Paul Ahlquist. Cytoplasmic viral replication complexes. *Cell Host Microbe*, 8(1):77–85, 2010.
- [44] Sanjay Dharmavaram, Selene Baochen She, Guillermo Lázaro, Michael Francis Hagan, and Robijn Bruinsma. Gaussian curvature and the budding kinetics of enveloped viruses. *PLOS Computational Biology*, 15(8):1–22, 08 2019.
- [45] Yinan Dong, Siyu Li, and Roya Zandi. Effect of the charge distribution of virus coat proteins on the length of packaged rnas. *Phys. Rev. E*, 102:062423, 2020.
- [46] Dominique Durand, Corinne Vivès, Dominique Cannella, Javier Pérez, Eva Pebay-Peyroula, Patrice Vachette, and Franck Fieschi. NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi-flexible linkers. *J. Struct. Biol.*, 169(1):45–53, 2010.
- [47] Eric C Dykeman, Peter G Stockley, and Reidun Twarock. Solving a levinthal’s paradox for virus assembly identifies a unique antiviral strategy. *PNAS*, 111(14):5361–5366, 2014.
- [48] E. Efrati, E. Sharon, and R. Kupferman. Elastic theory of unconstrained non-Euclidean plates. *Journal of the Mechanics and Physics of Solids*, 57(4):762–775, 4 2009.
- [49] K. Elleuch, F. Lequeux, and P. Pfeuty. Crosslink Effects on Equilibrium Polymers. *J. Phys. I France*, 5:465, 1995.
- [50] Oren M. Elrad and Michael F. Hagan. Mechanisms of size control and polymorphism in viral capsid assembly. *Nano Letters*, 8(11):3850–3857, 2008. PMID: 18950240.
- [51] Gonca Erdemci-Tandogan, Henri Orland, and Roya Zandi. Rna base pairing determines the conformations of rna inside spherical viruses. *Phys. Rev. Lett.*, 119:188102, Oct 2017.

- [52] Gonca Erdemci-Tandogan, Jef Wagner, Paul van der Schoot, Rudolf Podgornik, and Roya Zandi. Rna topology remodels electrostatic stabilization of viruses. *Phys. Rev. E*, 89:032707, Mar 2014.
- [53] Gonca Erdemci-Tandogan, Jef Wagner, Paul van der Schoot, Rudolf Podgornik, and Roya Zandi. Effects of rna branching on the electrostatic stabilization of viruses. *Phys. Rev. E*, 94:022408, 2016.
- [54] Gonca Erdemci-Tandogan, Jef Wagner, Paul van der Schoot, Rudolf Podgornik, and Roya Zandi. Effects of rna branching on the electrostatic stabilization of viruses. *Phys. Rev. E*, 94:022408, 2016.
- [55] H. Fraenkel-Conrat and Robley C. Williams. Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components. *Proceedings of the National Academy of Sciences*, 41(10):690–698, 1955.
- [56] D. Franke, M. V. Petoukhov, P. V. Konarev, A. Panjkovich, A. Tuukkanen, H. D. T. Mertens, A. G. Kikhney, N. R. Hajizadeh, J. M. Franklin, C. M. Jeffries, and D. I. Svergun. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of Applied Crystallography*, 50(4):1212–1225, 2017.
- [57] Barbie K Ganser, Su Li, Victor Y Klishko, John T Finch, and Wesley I Sundquist. Assembly and analysis of conical models for the hiv-1 core. *Science*, 283(5398):80–83, 1999.
- [58] Rees F. Garmann, Mauricio Comas-Garcia, Charles M. Knobler, and William M. Gelbart. Physical principles in the self-assembly of a simple spherical virus. *Acc. Chem. Res.*, 49:48–55, 2016.
- [59] Rees F. Garmann, Mauricio Comas-Garcia, Charles M. Knobler, and William M. Gelbart. Physical principles in the self-assembly of a simple spherical virus. *Acc. Chem. Res.*, 49(1):48–55, 2016. Number: 1.
- [60] Rees F. Garmann, Mauricio Comas-Garcia, Melissa S. T. Koay, Jeroen J. L. M. Cornelissen, Charles M. Knobler, and William M. Gelbart. Role of electrostatics in the assembly pathway of a single-stranded RNA virus. *J. Virol.*, 88(18):10472–10479, 2014.
- [61] Rees F. Garmann, Aaron M. Goldfain, and Vinothan N. Manoharan. Measurements of the self-assembly kinetics of individual viral capsids around their rna genome. *PNAS*, 116(45):22485–22490, 2019.
- [62] Luca Giomi and Mark Bowick. Crystalline order on Riemannian manifolds with variable Gaussian curvature and boundary. *Physical Review B*, 76(5):054106, 8 2007.
- [63] Rivka Goobes, Gil Goobes, Charles T. Campbell, and Patrick S. Stayton. Thermodynamics of statherin adsorption onto hydroxyapatite. *Biochemistry*, 45(17):5576–5586, 2006.

- [64] A. Gopal, Egecioglu D.E., Yoffe A.M., Ben-Shaul A, Rao ALN, C. M. Knobler, W. M. Gelbart, and A. Ben-Shaul. Viral RNAs Are Unusually Compact. *PLoS ONE*, 9:e105875, 2014.
- [65] Ajaykumar Gopal, Defne E. Egecioglu, Aron M. Yoffe, Avinoam Ben-Shaul, Ayala L. N. Rao, Charles M. Knobler, and William M. Gelbart. Viral RNAs are unusually compact. *PLoS One*, 9(9):e105875, 2014.
- [66] Ajaykumar Gopal, Z. Hong Zhou, Charles M. Knobler, and William M. Gelbart. Visualizing large RNA molecules in solution. *RNA*, 18(2):284–299, 2012.
- [67] Gregory M Grason. Colloquium: Geometry and optimal packing of twisted columns and filaments. *Reviews of Modern Physics*, 87(2):401–419, 2015.
- [68] M. F. Hagan. Modeling Viral Capsid Assembly. *Adv. Chem. Phys.*, 155:(in press), 2013.
- [69] M. F. Hagan and R. Zandi. Recent advances in coarse-grained modeling of virus assembly. *Curr. Opin. Virol.*, 18:36–43, 2016.
- [70] Michael F. Hagan and Gregory M. Grason. Equilibrium mechanisms of self-limiting assembly. *Rev. Mod. Phys.*, 93:025008, Jun 2021.
- [71] Mathijs Janssen, Andreas Härtel, and René van Roij. Boosting capacitive blue-energy and desalination devices with waste heat. *Phys. Rev. Lett.*, 113:268501, Dec 2014.
- [72] H. Ji and D. Hone. Polymer adsorption on rough surfaces. 2. Good solvent conditions. *Macromolecules*, 21:2600, 1988.
- [73] Jennifer M. Johnson, Jinghua Tang, Yaw Nyame, Deborah Willits, Mark J. Young, and Adam Zlotnick. Regulating self-assembly of spherical oligomers. *Nano Lett.*, 5(4):765–770, 2005. Number: 4.
- [74] Willem K. Kegel and Paul van der Schoot. Competing hydrophobic and screened-coulomb interactions in hepatitis B virus capsid assembly. *Biophys. J.*, 86(6):3905–3913, 2004. Number: 6.
- [75] Andrej Košmrlj and David R. Nelson. Statistical Mechanics of Thin Spherical Shells. *Physical Review X*, 7(1):011002, 1 2017.
- [76] 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*, 119:1869–1880, 2015.
- [77] L Landau and E M Lifshitz. *Theory of Elasticity*. Butterworth-Heinemann; 3 edition, 1985.
- [78] 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*, 113(12):3813–3819, 2009. Number: 12.

- [79] S. I. Lee and T. T. Nguyen. Radial Distribution of RNA Genomes Packaged inside Spherical Viruses. *Phys. Rev. Lett.*, 100:198102, 2008.
- [80] Siyu Li, Gonca Erdemci-Tandogan, Jef Wagner, Paul Van Der Schoot, and Roya Zandi. Impact of a nonuniform charge distribution on virus assembly. *Physical Review E*, 96(2):1–7, 2017.
- [81] Siyu Li, Gonca Erdemci-Tandogan, Jef Wagner, Paul van der Schoot, and Roya Zandi. Impact of a nonuniform charge distribution on virus assembly. *Phys. Rev. E*, 96(2):022401, 2017.
- [82] Siyu Li, Henri Orland, and Roya Zandi. Self consistent field theory of virus assembly. *Journal of Physics: Condensed Matter*, 30(14):144002, mar 2018.
- [83] Siyu Li, Polly Roy, Alex Travesset, and Roya Zandi. Why large icosahedral viruses need scaffolding proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 115(43):10971–10976, 2018.
- [84] Siyu Li, Roya Zandi, and Alex Travesset. Elasticity in curved topographies: Exact theories and linear approximations. *Physical Review E*, 99(6):063005, 2019.
- [85] Siyu Li, Roya Zandi, Alex Travesset, and Gregory M. Grason. Ground States of Crystalline Caps: Generalized Jellium on Curved Space. *Physical Review Letters*, 123(14):145501, 9 2019.
- [86] Jack Lidmar, Leonid Mirny, and David R. Nelson. Virus shapes and buckling transitions in spherical shells. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 68(5):051910, 11 2003.
- [87] Ronny Lorenz, Stephan H. Bernhart, Christian Höner zu Siederdisen, Hakim Tafer, Christoph Flamm, Peter F. Stadler, and Ivo L. Hofacker. ViennaRNA package 2.0. *Algorithms Mol. Biol.*, 6(1):26, 2011.
- [88] V. L. Lorman and S. B. Rochal. Density-wave theory of the capsid structure of small icosahedral viruses. *Phys. Rev. Lett.*, 98(18):185502, 2007.
- [89] A Losdorfer Bozic, A Siber, and R Podgornik. Electrostatic theory of viral self-assembly. *J. Biol. Phys.*, 39:215, 2013.
- [90] T. C. Lubensky and J. Isaacson. Statistic of Lattice Animals and Dilute Branched Polymers. *Phys. Rev. A*, 20:2130, 1979.
- [91] Antoni Luque, David Reguera, Alexander Morozov, Joseph Rudnick, and Robijn Bruinsma. Physics of shell assembly: Line tension, hole implosion, and closure catastrophe. *The Journal of Chemical Physics*, 136(18):184507, 05 2012.
- [92] Stan J. Maassen, Jurriaan Huskens, and Jeroen J. L. M. Cornelissen. Elucidating the thermodynamic driving forces of polyanion-templated virus-like particle assembly. *J. Phys. Chem. B*, 123(46):9733–9741, 2019.

- [93] Laurent Marichal, Jéril Degrouard, Anouchka Gatin, Nolwenn Raffray, Jean-Christophe Aude, Yves Boulard, Sophie Combet, Fabrice Cousin, Stéphane Hourdez, Jean Mary, Jean-Philippe Renault, and Serge Pin. From protein corona to colloidal self-assembly: The importance of protein size in protein–nanoparticle interactions. *Langmuir*, 36(28):8218–8230, 2020.
- [94] Laurent Marichal, Laetitia Gargowitsch, Rafael Leite Rubim, Christina Sizun, Kalouna Kra, Stéphane Bressanelli, Yinan Dong, Sanaz Panahandeh, Roya Zandi, and Guillaume Tresset. Relationships between rna topology and nucleocapsid structure in a model icosahedral virus. *Biophysical Journal*, 120(18):3925–3936, 2021.
- [95] Farzaneh Mohajerani and Michael F Hagan. The role of the encapsulated cargo in microcompartment assembly. *PLoS Comput. Biol.*, 14(7):e1006351, 2018.
- [96] Alexander Yu. Morozov and Robijn F Bruinsma. Assembly of viral capsids, buckling, and the Asaro-Grinfeld-Tiller instability. *Physical Review E*, 81(4):041925, 4 2010.
- [97] Michael Moshe, Eran Sharon, and Raz Kupferman. Elastic interactions between two-dimensional geometric defects. *Physical Review E*, 92(6):062403, 12 2015.
- [98] David R Nelson. *Defects and geometry in condensed matter physics*. Cambridge University Press, 2002.
- [99] P. Ni, Z. Wang, X. Ma, N. C. 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.*, 419:284, 2012.
- [100] Jiyong Ning, Gonca Erdemci-Tandogan, Ernest L Yufenyuy, Jef Wagner, Benjamin A Himes, Gongpu Zhao, Christopher Aiken, Roya Zandi, and Peijun Zhang. In vitro protease cleavage and computer simulations reveal the hiv-1 capsid maturation pathway. *Nature Communications*, 7:13689, 2016.
- [101] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, New York, NY, second edition, 2006.
- [102] Sanaz Panahandeh, Siyu Li, Bogdan Dragnea, and Roya Zandi. Virus assembly pathways inside a host cell. *ACS Nano*, 16(1):317–327, 2022.
- [103] Sanaz Panahandeh, Siyu Li, Laurent Marichal, Rafael Leite Rubim, Guillaume Tresset, and Roya Zandi. How a virus circumvents energy barriers to form symmetric shells. *ACS nano*, 14(3):3170–3180, 2020.
- [104] Sanaz Panahandeh, Siyu Li, and Roya Zandi. The equilibrium structure of self-assembled protein nano-cages. *Nanoscale*, 10(48):22802–22809, 2018.
- [105] Nikesh Patel, Eric C. Dykeman, Robert H. A. Coutts, George P. Lomonosoff, David J. Rowlands, Simon E. V. Phillips, Neil Ranson, Reidun Twarock, Roman Tuma, and Peter G. Stockley. Revealing the density of encoded functions in a viral rna. *Proceedings of the National Academy of Sciences*, 112(7):2227–2232, 2015.

- [106] Nikesh Patel, Simon J. White, Rebecca F. Thompson, Richard Bingham, Eva U. Weiß, Daniel P. Maskell, Adam Zlotnick, Eric C. Dykeman, Roman Tuma, Reidun Twarock, Neil A. Ranson, and Peter G. Stockley. HBV RNA pre-genome encodes specific motifs that mediate interactions with the viral core protein that promote nucleocapsid assembly. *Nat. Microbiol.*, 2(8):1–10, 2017.
- [107] A Pérez-Garrido, M J W Dodgson, and M A Moore. Influence of dislocations in Thomson’s problem. *Physical Review B*, 56(7):3640–3643, 8 1997.
- [108] A. Pérez-Garrido, M. Ortuño, E. Cuevas, and J. Ruiz. Many-particle jumps algorithm and Thomson’s problem. *Journal of Physics A: Mathematical and General*, 29(9):1973–1978, 1996.
- [109] J. D. Perlmutter, C. Qiao, and M. F. Hagan. Viral genome structures are optimal for capsid assembly. *eLife*, 2, 2013.
- [110] Jason D. Perlmutter and Michael F. Hagan. Mechanisms of virus assembly. *Annu. Rev. Phys. Chem.*, 66:217–239, 2015.
- [111] Jason D Perlmutter and Michael F Hagan. The role of packaging sites in efficient and specific virus assembly. *Journal of molecular biology*, 427(15):2451–2467, 07 2015.
- [112] Jason D. Perlmutter, Matthew R. Perkett, and Michael F. Hagan. Pathways for virus assembly around nucleic acids. *J. Mol. Biol.*, 426(18):3148–3165, 2014.
- [113] J. Zachary Porterfield and Adam Zlotnick. A simple and general method for determining the protein and nucleic acid content of viruses by UV absorbance. *Virology*, 407(2):281–288, 2010.
- [114] Ali Punjani, John L. Rubinstein, David J. Fleet, and Marcus A. Brubaker. cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods*, 14(3):290–296, 2017.
- [115] F. Qu and T. J. Morris. Encapsidation of turnip crinkle virus is defined by a specific packaging signal and RNA size. *J. Virol.*, 71(2):1428–1435, 1997.
- [116] Vamseedhar Rayaprolu, Alan Moore, Joseph Che-Yen Wang, Boon Chong Goh, Juan R Perilla, Adam Zlotnick, and Suchetana Mukhopadhyay. Length of encapsidated cargo impacts stability and structure of in vitro assembled alphavirus core-like particles. *Journal of Physics: Condensed Matter*, 29(48):484003, 2017.
- [117] Y. P. Ren, S. M. Wong, and L. Y. Lim. In vitro-reassembled plant virus-like particles for loading of polyacids. *J. Gen. Virol.*, 87:2749, 2006.
- [118] W. H. Roos, R. Bruinsma, and G. J. L. Wuite. Physical virology. *Nat. Phys.*, 6(10):733–743, 2010. Number: 10.
- [119] Andrew Routh, Tatiana Domitrovic, and John E. Johnson. Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus. *Proc. Natl. Acad. Sci. U.S.A.*, 109(6):1907–1912, 2012.

- [120] Anette Schneemann. The structural and functional role of RNA in icosahedral virus assembly. *Annu. Rev. Microbiol.*, 60(1):51–67, 2006. Number: 1.
- [121] H S Seung and David R Nelson. Defects in flexible membranes with crystalline order. *Physical Review A*, 38(2):1005–1018, 7 1988.
- [122] A. Shafir, D. Andelman, and R. R. Netz. Adsorption and depletion of polyelectrolytes from charged surfaces. *J. Chem. Phys.*, 119:2355, 2003.
- [123] A. Siber, A. L. Bozic, and R. Podgornik. Energies and pressures in viruses: contribution of nonspecific electrostatic interactions. *Phys. Chem. Chem. Phys.*, 14:3746, 2012.
- [124] A. Siber and R. Podgornik. Nonspecific interactions in spontaneous assembly of empty versus functional single-stranded RNA viruses. *Phys. Rev. E*, 78:051915, 2008.
- [125] F. D. Sikkema, M. Comellas-Aragones, R. G. Fokkink, B. J. M. Verduin, J.J.L.M. Cornelissen, and R. J. M. Nolte. Monodisperse polymer-virus hybrid nanoparticles. *Org. Biomol. Chem.*, 5:54, 2007.
- [126] Surendra W. Singaram, Rees F. Garmann, Charles M. Knobler, William M. Gelbart, and Avinoam Ben-Shaul. Role of RNA branchedness in the competition for viral capsid proteins. *J. Phys. Chem. B*, 119(44):13991–14002, 2015.
- [127] Venkatesh Sivanandam, Deborah Mathews, Rees Garmann, Gonca Erdemci-Tandogan, Roya 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, may 2016.
- [128] Venkatesh Sivanandam, Deborah Mathews, Rees Garmann, Gonca Erdemci-Tandogan, Roya 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. *Sci. Rep.*, 6(1):26328, 2016.
- [129] J. A. Speir, S. Munshi, G. Wang, T. S. Baker, and J. E. Johnson. Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy. *Structure*, 3(1):63–78, 1995.
- [130] Jeffrey A Speir and John E Johnson. Nucleic acid packaging in viruses. *Curr. Opin. Struct. Biol.*, 22(1):65–71, 2012. Number: 1.
- [131] Peter G. Stockley, Reidun Twarock, Saskia E. Bakker, Amy M. Barker, Alexander Borodavka, Eric Dykeman, Robert J. Ford, Arwen R. Pearson, Simon E. V. Phillips, Neil A. Ranson, and Roman Tuma. Packaging signals in single-stranded RNA viruses: nature’s alternative to a purely electrostatic assembly mechanism. *J. Biol. Phys.*, 39:277, 2013.

- [132] Jingchuan Sun, Chris DuFort, Marie-Christine Daniel, Ayaluru Murali, Chao Chen, Kodetham Gopinath, Barry Stein, Mrinmoy De, Vincent M. Rotello, Andreas Holzenburg, C. Cheng Kao, and Bogdan Dragnea. Core-controlled polymorphism in virus-like particles. *Proceedings of the National Academy of Sciences*, 104(4):1354–1359, 2007.
- [133] Jingchuan Sun, Chris DuFort, Marie-Christine Daniel, Ayaluru Murali, Chao Chen, Kodetham Gopinath, Barry Stein, Mrinmoy De, Vincent M. Rotello, Andreas Holzenburg, C. Cheng Kao, and Bogdan Dragnea. Core-controlled polymorphism in virus-like particles. *Proc. Nat. Acad. Sci. USA*, 104(4):1354–1359, 2007.
- [134] Huand Tao, Zhangand Rui, and B. I. Shklovskii. Electrostatic theory of viral self-assembly. *Physica A*, 387:3059, 2008.
- [135] Alex Travesset. Universality in the screening cloud of dislocations surrounding a disclination. *Physical Review B*, 68(11):115421, 9 2003.
- [136] Guillaume Tresset, Jingzhi Chen, Maelenn Chevreuil, Naima Nhiri, Eric Jacquet, and Yves Lansac. Two-dimensional phase transition of viral capsid gives insights into subunit interactions. *Phys. Rev. Applied*, 7(1):014005, 2017. Number: 1.
- [137] Guillaume Tresset, Clémence Le Coeur, Jean-François Bryche, Mouna Tatou, Mehdi Zeghal, Annie Charpilienne, Didier Poncet, Doru Constantin, and Stéphane Bressanelli. Norovirus capsid proteins self-assemble through biphasic kinetics via long-lived stave-like intermediates. *J. Am. Chem. Soc.*, 135(41):15373–15381, 2013. Number: 41.
- [138] Cécile Troupin, Axelle Dehée, Aurélie Schnuriger, Patrice Vende, Didier Poncet, and Antoine Garbarg-Chenon. Rearranged genomic RNA segments offer a new approach to the reverse genetics of rotaviruses. *J. Virol.*, 84(13):6711–6719, 2010.
- [139] Luca Tubiana, Anže Lošdorfer Božič, Cristian Micheletti, and Rudolf Podgornik. Synonymous mutations reduce genome compactness in icosahedral ssRNA viruses. *Biophys. J.*, 108(1):194–202, 2015.
- [140] Reidun Twarock, Richard J Bingham, Eric C Dykeman, and Peter G Stockley. A modelling paradigm for RNA virus assembly. *Curr. Opin. Virol.*, 31:74–81, 2018.
- [141] Reidun Twarock and Antoni Luque. Structural puzzles in virology solved with an overarching icosahedral design principle. *Nat. Commun.*, 10(1):1–9, 2019.
- [142] Paul van der Schoot and Roya Zandi. Impact of the topology of viral rnas on their encapsulation by virus coat proteins. *J. Biol. Phys.*, 39(2):289–299, 2013.
- [143] Hans Vink. A new convenient method for the synthesis of poly(styrenesulfonic acid). *Makromol. Chem.*, 182:279–281, 1981.
- [144] J. Wagner, G. Erdemci-Tandogan, and Roya Zandi. Adsorption of annealed branched polymers on curved surfaces. *J. Phys.:Condens. Matter*, 27:495101, 2015.

- [145] Chunyan Wang, Jiagang Tu, Jun Liu, and Ian J Molineux. Structural dynamics of bacteriophage P22 infection initiation revealed by cryo-electron tomography. *Nature Microbiology*, 4(6):1049–1056, March 2019.
- [146] Aron M. Yoffe, Peter Prinsen, Ajaykumar Gopal, Charles M. Knobler, William M. Gelbart, and Avinoam Ben-Shaul. Predicting the sizes of large RNA molecules. *Proc. Natl. Acad. Sci. U.S.A.*, 105(42):16153–16158, 2008.
- [147] Roya Zandi, Bogdan Dragnea, Alex Travesset, and Rudolf Podgornik. On virus growth and form. *Phys. Rep.*, 847:1–102, mar 2020.
- [148] Roya Zandi, Bogdan Dragnea, Alex Travesset, and Rudolf Podgornik. On virus growth and form. *Phys. Rep.*, 847:1–102, 2020.
- [149] Roya Zandi, David Reguera, Robijn F. Bruinsma, William M. Gelbart, and Joseph Rudnick. Origin of icosahedral symmetry in viruses. *Proceedings of the National Academy of Sciences*, 101(44):15556–15560, 2004.
- [150] Cheng Zeng, Mercedes Hernando-Pérez, Bogdan Dragnea, Xiang Ma, Paul van der Schoot, and Roya Zandi. Contact mechanics of a small icosahedral virus. *Phys. Rev. Lett.*, 119(3):038102, jul 2017.
- [151] Yong-Zhen Zhang, Yan-Mei Chen, Wen Wang, Xin-Chen Qin, and Edward C. Holmes. Expanding the RNA virosphere by unbiased metagenomics. *Annu. Rev. Virol.*, 6(1):119–139, 2019. Number: 1.
- [152] Z Hong Zhou, Matthew Dougherty, Joanita Jakana, Jing He, Frazer J Rixon, and Wah Chiu. Seeing the herpesvirus capsid at 8.5 Å. *Science*, 288(5467):877–880, 2000.
- [153] A. Zlotnick, R. Aldrich, J. M. Johnson, P. Ceres, and M. J. Young. Mechanism of Capsid Assembly for an Icosahedral Plant Virus. *Virology*, 277:450, 2000.
- [154] Adam Zlotnick. Distinguishing reversible from irreversible virus capsid assembly. *Journal of Molecular Biology*, 366(1):14–18, 2007.