UNIVERSITY OF CALIFORNIA
RIVERSIDE

Physical Principles of the Assembly of Virus Particles and other Protein Nano Containers

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

Doctor of Philosophy

in

Physics

by

Sanaz Panahandeh

June 2022

Dissertation Committee:

Dr. Roya Zandi, Chairperson
Dr. John Barton
Dr. Thomas Kuhlman

The Dissertation of Sanaz Panahandeh is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I am very grateful to my advisor, Roya Zandi for her support, patience and encouragement over these many years. I greatly appreciate that she has been always available to advise and help me. My daughter arrived in the middle of my PhD studies. Raising a baby along with doing research has created many new challenges for me. I have been fortunate to have an advisor who not only guided me in research but also shared her own experience as a woman and mother in academia.

I would like to thank my committee members John Barton and Thomas Kuhlman for their useful comments.

I would like to thank Siyu Li for generously sharing her knowledge with me. I was very lucky to work with her on many projects.

I would also like to thank my other group-mates: Yinan Dong, Amin Safdari and Alireza Ramezani for their friendship and constructive discussions we had.

I would like to present my sincere thankfulness to my dear parents for their great role in my life and their numerous sacrifices for me and for my sisters. I lost my dad in the first year of my PhD studies. I regret that he did not live to see my PhD graduation. I love you and miss you so much baba!

I am also grateful to my sisters Sholeh and Narges, my aunt family Zahra, Hadi and Parisa for their tremendous support and understanding. I would like to thank my daughter, Hannah who joined us in the fourth year of my studies, for giving me unlimited happiness and pleasure. Finally, I would like to say thank you to my husband, Pouyan for his patience and tolerance over the last 6 years. Pouyan your love and support are the

strongest backups for me. Without you this work would never have been possible.

Chapter 2 is a version of S. Panahandeh, S. Li and R. Zandi, "The equilibrium structure of self-assembled protein nano-cages", Nanoscale, 10, 22802, (2018).

Chapter 3 is a version of S. Panahandeh, S. Li, B. Dragnea, and R. Zandi, "Virus assembly pathway inside a host cell", ACS Nano, 16, 1, 317–327, (2022).

Chapter 4 is a version of S. Panahandeh, S. Li, L. Marichal, R. Leite Rubim, G. Tresset, and R. Zandi, "How a virus circumvents energy barriers to form symmetric shells", ACS Nano, 14, 3, 3170–3180, (2020).

This dissertation is dedicated to my parents, Pouyan, and Hannah.

I love you forever!

ABSTRACT OF THE DISSERTATION

Physical Principles of the Assembly of Virus Particles and other Protein Nano Containers

by

Sanaz Panahandeh

Doctor of Philosophy, Graduate Program in Physics
University of California, Riverside, June 2022
Dr. Roya Zandi, Chairperson

Understanding how highly symmetric, robust, monodisperse protein nano-cages self-assemble can have major applications in various areas of bio-nanotechnology, such as drug delivery, biomedical imaging, and gene therapy. Among all the protein nano-cages, the viral shells have in particular received a lot of attention because of their abundance in nature with members infecting all kingdoms of life. The simplest viruses are made of a genome (RNA or DNA) and a protein shell called the capsid. The dissertation can be separated into two parts. The first part focuses on the equilibrium structure of nano cages. Using Monte Carlo simulation, we obtain global minimum energy structures in the absence and presence of genome. Our results suggest that the physical properties including the spontaneous curvature, flexibility, and bending rigidity of coat proteins are sufficient to predict the size, symmetry, and shape selectivity of the assembly products in the absence of genome. We find that in the presence of the genome, cargo-coat protein interactions also impact the structure and stability of the viral structures. We show that the equilibrium shells encapsidating small globular nucleic acid cargo can be assembled into non-icosahedral structures, which have

been observed in experiments involving small segments of RNA. However, increasing the size of the genome the equilibrium structure switches to icosahedral structures.

In the second part, we study the kinetic pathways of assembly of virus coat proteins around the genome. Our calculations show that the non-icosahedral structures assembling around small genomes are strained and less stable than icosahedral ones. Monitoring the assembly pathways of proteins shell, we find that the strained non-icosahedral structures can easily be split into fragments along stress lines and be re-assembled into the stable native icosahedral shells if the larger wild-type genome becomes available. We also present our results corresponding to two different assembly mechanisms: en masse assembly and the nucleation and growth pathway. Using Monte Carlo simulations, we specifically elucidate the role of elastic energy in the disorder-order transition in icosahedral viruses formed through both mechanisms. Self-assembly experiments on a model icosahedral plant virus show a disorderorder transition occurs under physiological conditions upon an increase in capsid protein concentrations. We use Monte Carlo simulations to explain this disorderorder transition and find that, as the shell grows, the structures of disordered intermediates in which the distribution of pentamers do not belong to the icosahedral subgroups become energetically so unfavorable that the caps can easily dissociate and reassemble, overcoming the energy barriers for the formation of perfect icosahedral shells. In addition, we monitor the growth of capsids under the condition that nucleation and growth is the dominant pathway and show that the key for the disorderorder transition in both en masse and nucleation and growth pathways lies in the strength of elastic energy compared to the other forces in the system including proteinprotein interactions and the chemical potential of free

subunits. Our findings explain, at least in part, why perfect virions with icosahedral order form under different conditions including physiological ones.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Viruses are the simplest biological organisms that mainly consist of genome that can be RNA or DNA and a capsid. Capsids are protein shells that protect the genome from the harsh environment out of the cells. The life cycle of a virus consists of six steps: Absorption during which virus attaches to the host cell membrane and enters the cell. Entry strategies are different in different viruses. For example most of the viruses without envelope enter cells through endocytosis. However, some other viruses such as bactriophages inject their genetic materials into the cell. After entering the host cell capsid disassembles into capsid proteins (CPs) during uncoating process and releases its genome. The released genome uses the machinery of the host cell to reproduce itself and CPs. Finally, the new CPs assemble around genomes and form new virions and exit the cell to infect other cells. The assembly of CPs into the viral shell has been the center of many attention over the past two decades and understanding of the physics underlying this process has been considerably advanced through experimental and theoretical studies [40, 103, 42, 11, 23].

Viruses can be categorized based on capsid shapes or their genome types. According to the capsid shapes viruses are classified into three families: cylindrical shape like tobacco mosaic virus (TMV), spherical shape like cowpea chlorotic mottle virus (CCMV) and conical shape like human immunodeciency viruses (HIV). Most of the spherical viruses adopt icosahedral symmetry (IO). In icosahedral structures proteins are clustered either in pentagonal or hexagonal units that pentamers are distributed evenly on the surface of capsid. One of the amazing functions that viruses are able of, is their ability to assemble spontaneously to highly symmetric IO structures not only *in vivo* but also in some cases *in vitro* . In 1967 the first spheical virus (CCMV) was reconstituted *in vitro* by Bancroft *et al.*. Note that While most of the spherical viruses have icosahedral symmetry, their sizes vary from 20nm to 100 nm in diameter. One way to categorize icosahedral structures with different number of subunits is the triangulation number that can be calculated following Casper and Klug model [10]. T number is defined as follows:

$$T = h^2 + hk + k^2 \tag{1.1}$$

, h and k are the number of steps that connects two adjacent pentamers along principal vectors of hexagonal lattice. The total number of protein subunits in an icosahedral capsid is 60T. There are always 12 pentamers in these capsids regardless to the size of capsid. Rest of the protein subunits form hexamers. Since h and k are positive integer values, T assumes certain numbers (T=1,3,4,7,...). Other than viruses, many other biological structures self-assemble from protein subunits into icosahedral and other highly symmetric structures such as clathrin vesicles [13, 17]. How these biological structures self-assemble to form such

precise structures from CPs has been the topic of many studies. However, still there are many questions that need to be answered.

Understanding how highly symmetric, robust, monodisperse protein nano-cages self-assemble can have major applications in various areas of bio-nanotechnology, such as drug delivery, biomedical imaging and gene therapy. In chapter 2, We develop a model to investigate the assembly of protein subunits into the structures with different sizes and symmetries. Using Monte Carlo simulation, we obtain global minimum energy structures. Our results suggest that the physical properties including the spontaneous curvature, flexibility and bending rigidity of coat proteins are sufficient to predict the size, symmetry and shape selectivity of the assembly products. Further, on a thermodynamic basis, we discuss the polymorphism of nano-cages observed in assembly experiments.

Recent experiments done by Bond *et al.*, revealed formation of a capsid from brome mosaic virus (BMV) coat proteins around oligonucleotides with structures other than BMV native structure [5]. Investigating the source of emergence of these new structures which were not found in our previous studies in the absence of genome is topic of chapter 3. The spontaneous self-assembly of CPs around their genome is mainly driven by the attractive electrostatics interaction between the positive charges on capsid proteins and the negative charges on the genome. Despite its importance and many decades of intense research, how the virus selects and packages its native RNA inside the crowded environment of a host cell cytoplasm in the presence of an abundance of nonviral RNA and other anionic polymers has remained a mystery. In chapter 3, we perform a series of simulations to monitor the growth of viral shells and find the mechanism by which cargocoat protein interactions can

impact the structure and stability of the viral shells. We show that coat protein subunits can assemble around a globular nucleic acid core by forming nonicosahedral cages, which have been recently observed in assembly experiments involving small pieces of RNA. We find that the resulting cages are strained and can easily be split into fragments along stress lines. This suggests that such metastable nonicosahedral intermediates could be easily reassembled into the stable native icosahedral shells if the larger wild-type genome becomes available, despite the presence of a myriad of nonviral RNAs.

To study more of the role of genome on kinetic pathways of virus assembly, we develop a Monte Carlo simulation in grand canonical ensemble in chapter 4 and study the assembly of CPs following two different pathways: "en masse" and "nucleation and growth". In "en masse" pathway, CPs initially get absorbed by genome and form disordered structures. The experimental results show a disorder-order transition of capsid structures around genome when CPs concentration increases. By performing series of simulations we explain the source of disorder-order transition under physiological conditions. Besides, we study the effect of protein concentration, hydrophobicity, and elastic energy on the formation of T=3 viruses. Finally in chapter 5, we summarize our findings and will suggest several research ideas that can improve our understanding of virus assembly pathways.

Chapter 2

The equilibrium structure of self-assembled protein nano-cages

2.1 Introduction

Self-assembly of monodispersed protein cages is ubiquitous in nature. Because of their biocompatibility, stability and low toxicity, protein cages have important roles in many biological processes, medicine and bio-nanotechnology. Examples of protein cages include platonic hydrocarbons, heat shock proteins, ferritins, carboxysomes, silicages, multicomponent ligand assemblies, clathrin vesicles and virus shells, to name a few [52, 4, 2, 31, 66]. The protein shells are necessary for both protection and delivery of various cargos in biological systems. For instance, ferritin stores iron and exists in almost every living organism.

Among all biological entities, viruses in particular have optimized the feat of packaging of genetic materials and other anionic cargos into a protein shell called the capsid,

recognized as one of the most efficient nano-containers for trafficking genetic material in nature [81, 27]. Most protein cages self-assemble from a large number of one or a few different types of protein subunits into complex supramolecular structures with diameters ranging from 10 to 500 nm [43]. Quite remarkably under many circumstances, viruses spontaneously assemble *in vitro* from protein building blocks into highly symmetric shells [37, 14]. Most spherical viruses adopt structures with icosahedral symmetry [16, 100, 48] characterized by a structural index T number, which assume only certain integers (1, 3, 4, 7, ...) [10]. The number of protein subunits in icosahedral shells is often 60 times the T-number.

Other protein cages can adopt several other symmetric structures. For example, clathrin shells form icosahedral structures in addition to many other symmetric shells [13, 17] depending on the size of their cargo. Nevertheless protein cages with icosahedral symmetry are by far the most abundant in nature. Figure 2.1 illustrates the structure of Lumazine synthase with $T = 1$ symmetry[69], a Clathrin shell with tetrahedral symmetry [32], an encapsulin nanocompartment from M. xanthus with $T = 3$ structure [68] and the Hepatitis B virus (HBV) capsid with $T = 4$ symmetry.

Despite the abundance of protein shells in nature, the role of building blocks and the factors contributing to the stability, size and shape selectivity of nanostructures are not well-understood. To this end, there is a precedent need to take a bottom-up approach and to understand at the fundamental scale the impact of building blocks on the design and formation of functional nano-shells.

Extensive work has explored the effect of spontaneous radius of curvature (dihedral angle) of building blocks on the equilibrium structure of protein cages [41]. For instance,

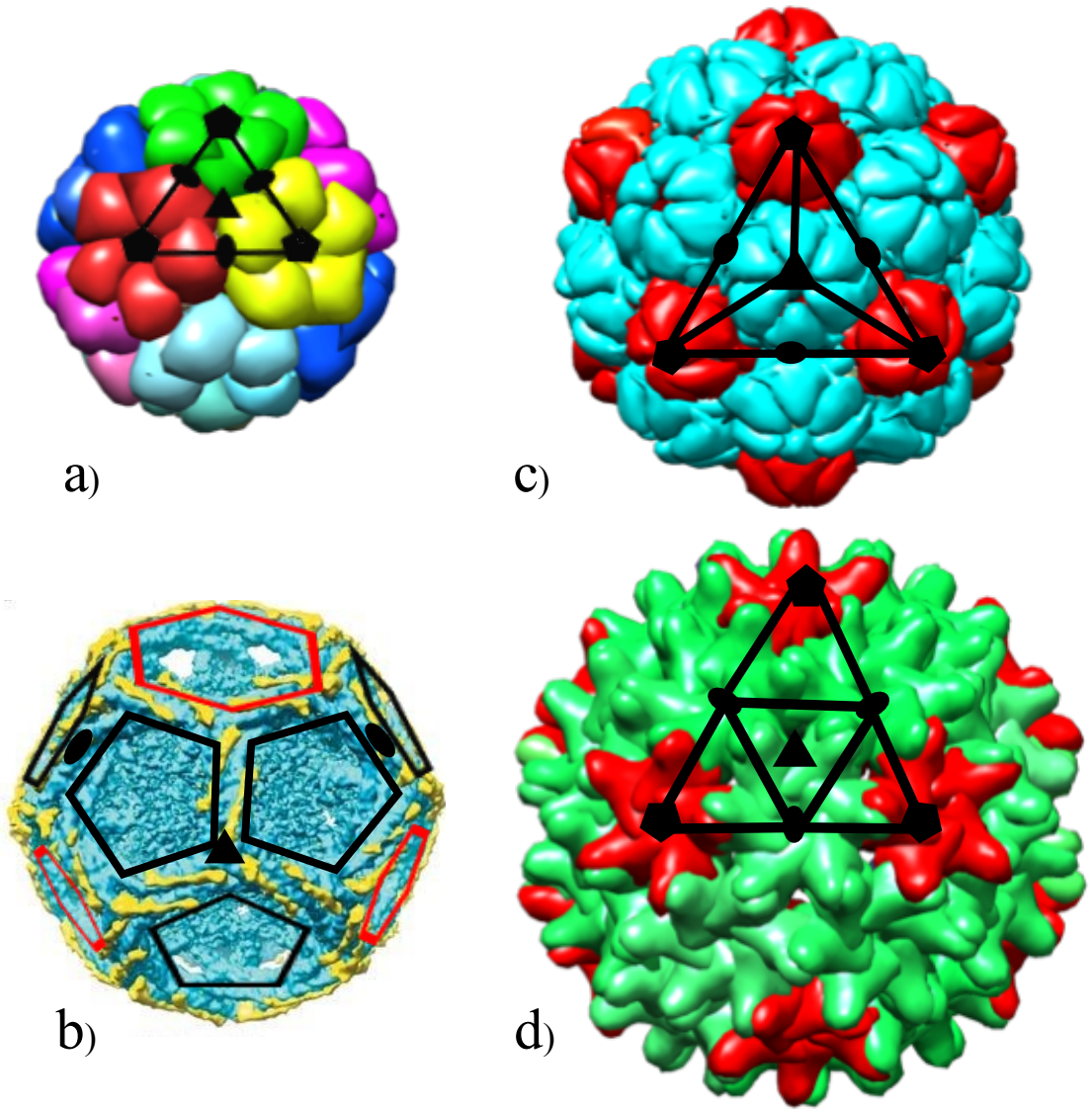

Figure 2.1: Structures of some protein cages. (a) Lumazine synthase is an enzyme with icosahedral symmetry ($T = 1$) constructed of 60 identical protein subunits. The colors are added to highlight each pentamer. (b) Mini-coat has tetrahedral symmetry [32]. The two-fold and three-fold symmetry axes are marked with small black ovals and triangles respectively. (c) Encapsulin from M. xanthus with $T = 3$ structure is made of 180 identical protein subunits. The position of two, three and five-fold symmetry axes are marked in the picture. (d) Hepatitis B virus, a $T = 4$ structure. The darker color (red) in (c) and (d) are pentamers. All the structures except (b) are reproduced using UCSF Chimera packages (http://www.rbvi.ucsf.edu/chimera).

using the Monte Carlo (MC) simulations, Chen *et al.* studied the self-assembly of attractive cone-shaped particles into different structures [12]. They obtained a sequence of clusters and found that the symmetry and stability of formed structures depend on the cone angle or the preferred angle between subunits. Similar sequence of structures was obtained with attractive spherical particles but under certain convexity constraints, equivalent of changing the preferred dihedral angle between subunits [11].

The simple case of N spherical colloids or circular disks interacting through Lennard-Jones potential constrained to move on the surface of a sphere also shows that the equilibrium structure of shells depends on the number of building blocks and the preferred angle between disks or Lennard-Jones particles [104]. As the preferred angle between disks or colloids changes, structures with different size and symmetries form.

More recently, Paquay *et al.* studied the equilibrium structures of interacting Morse particles residing on the surface of a sphere and found similar structures and magic numbers as observed in the case of LJ particles [78]. Nevertheless, the impact on the equilibrium structures of the mechanical properties of building blocks including flexibility and bending rigidity have not previously been studied. While the dynamical structures of protein shells under non-equilibrium conditions as a function of bending rigidity and stretching modulus of building blocks have been thoroughly investigated in Ref. [99], due to irreversible steps in the shell growth, the structures obtained in those simulations might be completely far from equilibrium.

In this chapter we investigate the equilibrium structure of nano-shells and the important factors contributing to their stability and symmetry. Using MC simulations

combined with the bond flipping method [51, 85] we study the structure of protein cages as a function of the spontaneous curvature as well as stretching and bending rigidity of building blocks, advancing our knowledge for producing high yield nano-cages with specific size and shape.

While spontaneous curvature is an important factor in defining the size of the shell, we find that the flexibility and bending rigidity of building blocks can completely modify the size and final symmetry of the shells. Quite interestingly the sequence of clusters or magic numbers and their associated shells obtained in our equilibrium studies, coincide not only with the structure of viruses displaying icosahedral symmetry but with other non-icosahedral protein cages observed in other systems such as clathrin shells.

Furthermore, we find that there are striking similarities between the minimum energy structure phase diagram and the one obtained through irreversible growth [99], as a function of the mechanical properties and spontaneous curvature of building blocks. Both diagrams display the same symmetry for almost the same regions of the parameter space. We emphasize that the location of pentamers in a shell defines its symmetry. If the position of a pentamer in a symmetric shell is slightly displaced, the symmetry will be broken. The fact that despite the irreversible steps, pentamers form in the "right" positions, preserving the symmetry during the irreversible shell growth, is quite unexpected. The similarities between irreversible and equilibrium structures in this chapter can be explained at least in part by the recent work of Li *et al.*. Using continuum elasticity theory, they have shown that as an elastic shell grows, there are regions on the spherical cap that strongly adsorb disclinations (pentamers), *i.e.*, there is a high affinity for the formation of disclinations at

certain locations as the cap grows. We will discuss this effect more in the Discussion and Summary section of this chapter.

We find some differences between two phase diagrams too. A few symmetric structures grown in irreversible simulations [99] do not constitute the minimum free energy structures. Furthermore, we obtain additional symmetric structures in the equilibrium simulations, which were not observed in the growth simulations under irreversible conditions.

We note here that in our MC simulations the assembled shells are obtained under the condition that the thermal fluctuations are negligible compared to the other energies in the system, and as such the equilibrium structures coincide with the minimum energy structures.

It is also worth mentioning that it is now widely accepted that the preferred curvature and mechanical properties of subunits depend on the solution conditions such as pH and salt concentration [28]. The interplay of protein geometry, repulsive electrostatic and attractive hydrophobic interactions define the equilibrium properties (bending and stretching moduli and spontaneous curvature) of subunits, nevertheless no systematic experimental data are known for these parameters. Understanding the role of stiffness and preferred curvature of building blocks could lead to generation of a range of new materials and novel structures.

2.2 Method

To study the equilibrium structures, we consider stretchable equilateral triangular subunits, representing building blocks of protein cages, as illustrated in Fig. 2.1a and

d. Since capsids are primarily built from pentamers and hexamers, triangular subunits (trimers) have been widely employed to study the assembly of viral shells and other protein nanocages [98, 99, 40, 60]. The elastic energy of a growing capsid, thus becomes, the sum of the stretching E_s and bending E_b energies [45, 56]. The stretching energy results from the deformation of the triangles from their equilateral shape and sums over all bonds. Assuming that each bond is built from a linear spring with the equilibrium length b_0 and the spring constant k_s, the stretching energy can be written as,

$$E_s = \sum_i \sum_{a=1}^{3} \frac{k_s}{2} \, (b_i^a - b_0)^2 \tag{2.1}$$

with i the triangular subunit index, b_0 the equilibrium length of the edges, and b_i^a the length of the a^{th} edge in the i^{th} subunit that can be stretched or compressed.

The bending energy is due to the deviation of local curvature from the preferred one (to be defined as a parameter in this study) and is calculated by summing over all neighboring pairs of subunits with $< i, j >$ indexing pairs of adjacent subunits and k_b being the bending stiffness as follows,

$$E_b \;\; = \;\; \sum_{<i,j>} k_b[1 - \cos(\theta_{<i,j>} - \theta_0)]. \tag{2.2}$$

The preferred angle between adjacent subunits and spontaneous radius of curvature are related through $\sin(\theta_0/2) = \; (12R_0^2/b_0^2 \; - 3)^{-1/2}$ with R_0 the spontaneous radius of curvature. The angle $\theta_{<ij>}$, where $\cos\theta_{<ij>} = \hat{n_i} \cdot \hat{n_j}$, is between the unit normal vectors $\hat{n_i}$ and $\hat{n_j}$ of the two adjacent subunits i and j sharing an edge.

Equations 2.1 and 2.2 reveal the presence of two important dimensionless parameters, the spontaneous radius of curvature R_0/b_0 and the Foppl von Karman (FvK) number

$$\gamma = k_s b_0^2/k_b, \tag{2.3}$$

which indicates the relative difficulty of deforming an equilateral triangular subunit compared to changing the dihedral angle between two adjacent subunits away from the preferred one. We note that both dimensionless parameters are normalized with respect to the size of the subunits b_0. Both γ and R_0/b_0 depend on the intrinsic property of proteins (shape and resistance to deformation), and also on the solution circumstances such as pH and ionic strength, and for closed cages also on differentials of osmolyte concentration[29, 107].

To obtain the lowest-energy configurations we employ a series of simulated annealing MC simulations [62]. We start from a triangulated spherical mesh with a random distribution of N_v vertices. Each MC step consists of N_v attempted bond movings, which involves removing and reattaching the edge connecting two vertices of two neighboring triangles such that the two vertices which were not connected before, they will be linked by an edge after the flip, as is shown in Fig. 2.2. Detachment and reconnection of the bonds are such that the total number of vertices N_v, subunits n_s and edges in the shell remain constant. Since our goal is to obtain the global minimum energy structure each edge swapping is followed by the shell relaxation during which vertices will move to the positions that minimize the total elastic energy. We employ the BFGS method to relax and minimize the energy of the shell [75]. The probability that the new relaxed structure with the new position of vertices be accepted is $\min(1, e^{(E_{old}-E_{new})/k_BT})$. E_{old} and E_{new} are the energies of the structures before and after the trial edge swapping, respectively. We generate a

Markov chain with Boltzmann probabilities by iterating the edge swapping until the energy converges. The edge swapping process is reversible to ensure detailed balance.

We repeat the above simulations with different initial configurations many times. To avoid local minimum free energy traps, we employ simulated annealing with both linear and non-linear schedules [12]. The results show no dependence on the cooling path as long as the cooling rate is very slow. We note again that since the thermal fluctuations are negligible compared to the other energies in the system, the structures we obtain correspond to the global minimum energy structures.

The above algorithm allows us to successfully change the position of pentamers and hexamers. In other words, during the simulations the location of disclinations is not fixed; they can move, and thus change the structure and symmetry of the shell. We perform MC simulations for all the structures ranging from $N_v = 12$ to 42 corresponding to the shells made of $n_s = 20$ to 80 number of subunits. In nature larger shells need some external help like scaffolding proteins or inner core to form symmetric shells [19], the focus of this study is, however, on the smaller shells that are able to spontaneously assemble without any core.

2.3 Results and Discussion

We carry out Monte Carlo simulations as described in the Method section of Chapter 2 for different numbers of subunits n_s. We start with a fixed preferred spontaneous radius of curvature $R_0/b_0 = 1.28$ but different values of γ. The results of the simulations are illustrated in Fig. 2.3 in the form of a plot of the minimized elastic energy per triangles ϵ_n (in

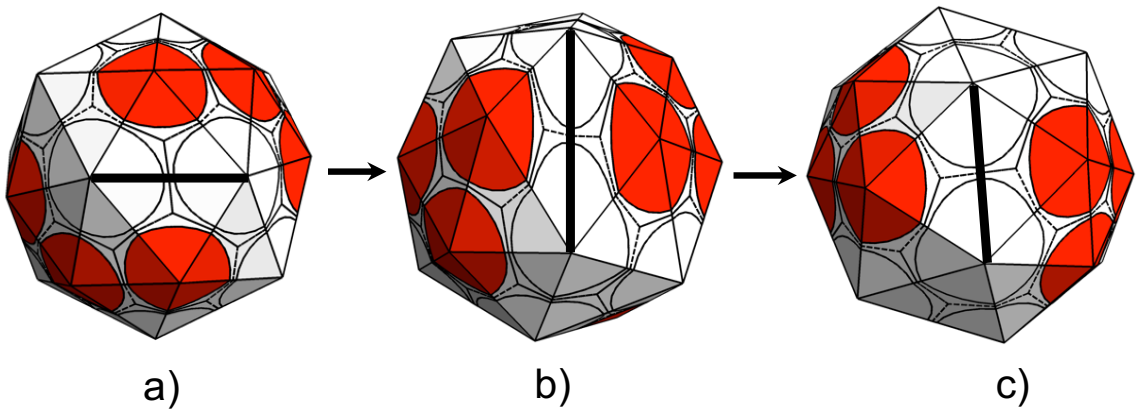

Figure 2.2: Bond moving method: (a) The thick black edge between two neighboring triangles is randomly chosen. (b) The black line is removed from its previous position and the two vertices that were not sharing a bond before the swap, are now connected. The darker (red) shades indicate the positions of pentamers and the white ones correspond to hexamers. By moving the bond from (a) to (b), the position of pentamers and hexamers are changed. (c) The system is energetically relaxed now after the swap.

units of k_BT) versus the number of subunits, n_s. The solid light line (green) corresponds to $\gamma = 0.5$, the dashed line to $\gamma = 1$, the dark solid line to $\gamma = 3$ and the dotted line to $\gamma = 8$. We emphasize that since γ is proportional to the ratio of stretching to bending modulus, for larger γs it is difficult to deform the subunits from their equilibrium equilateral shape but rather easy to bend them away from their preferred dihedral angle. For small γs, in contrast, the subunits can be easily deformed but it costs significantly more energy to modify the dihedral angle between the adjacent subunits from the preferred one.

Figure 2.3 illustrates that there are many local minima but no distinguished global minimum energy for $\gamma = 0.5$. The energy landscape indeed becomes flatter as $\gamma \to 0$. When γ becomes smaller, the subunits becomes more flexible, and thus the size of the shell changes with R_0/b_0. However, as γ becomes larger, the subunits become more rigid and avoid configurations leading to the deformation of equilateral triangles, and as such only

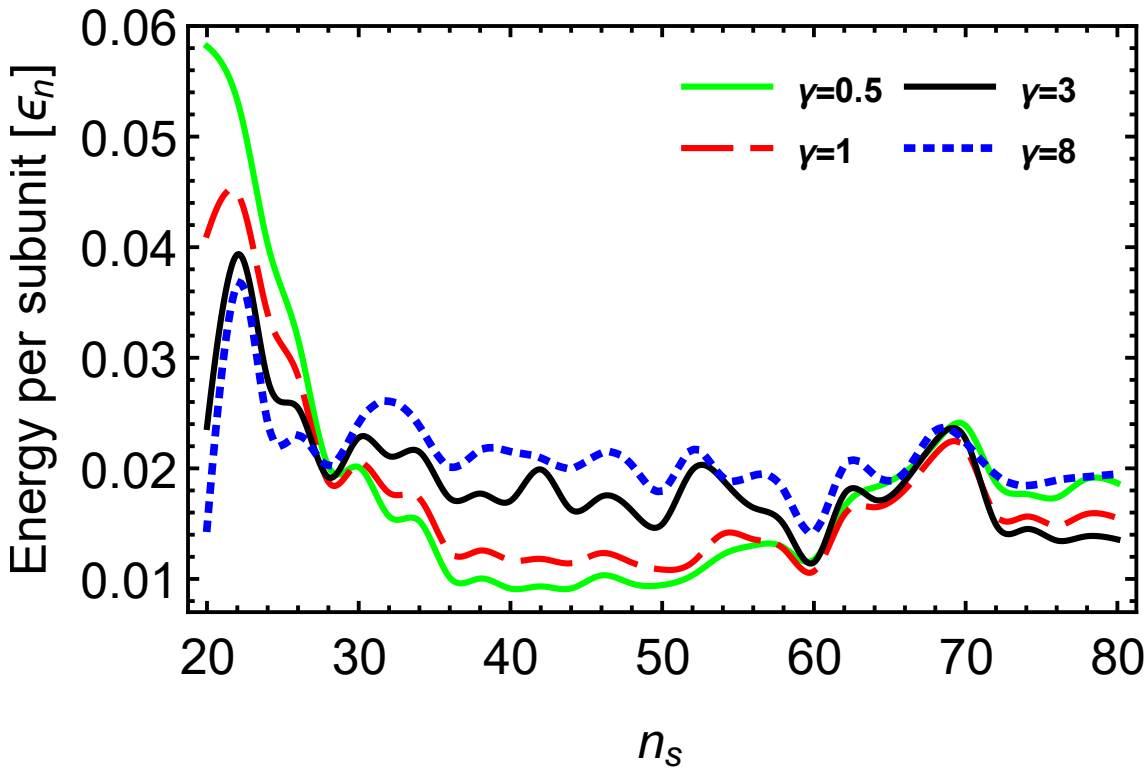

Figure 2.3: Energy per subunit versus number of subunits are displayed for $\gamma = 0.5,\ 1,\ 3$ and 8. The spontaneous radius of curvature is fixed at $R_0/b_0 = 1.28$. While the minimum energy for $\gamma = 0.5$ is at $n_s = 40$, for $\gamma = 1,\ 3$ and 8 the equilibrium structure is a $T = 3$ icosahedral structure.

the structures with icosahedral symmetry remain the global minimum energy structure. For $\gamma < 1$, other structures with different n_s compete or have lower energies than a $T = 3$ structure.

We next investigate the impact of the spontaneous radius of curvature on the global energy minima of Fig. 2.3 for various γs. Figure 2.4 illustrates the plot of energy per subunit versus R_0/b_0 for the global minimum energy structures (Fig. 2.3) at different γ-values. The curves in Fig. 2.4 can be divided into different segments, each representing different structure. The capital letter at the beginning of each segment reveals the symmetry

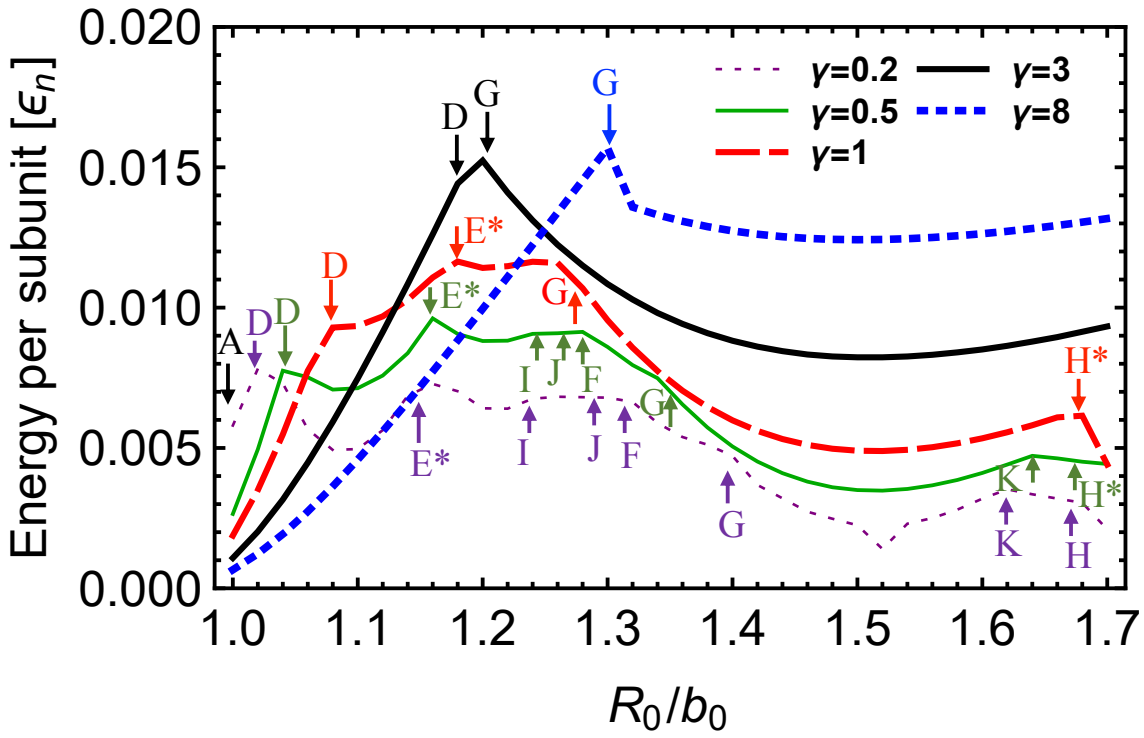

Figure 2.4: Plot of energy per subunit versus R_0/b_0 for $\gamma = 0.5,\ 1,\ 3$ and 8. Each curve can be divided into different segments. The capital letter at the beginning of each segment (from left to right) indicates the symmetry of the segment. The letter A pertains to the beginning of all curves. The corresponding structures are illustrated in Fig. 2.7a

.

and structure of that segment. Note that letter A pertains to the beginning of all curves. For instance, the letter A at the beginning of the thick dotted line shows that for $\gamma = 8$ the global minimum energy structure is a $T = 1$ icosahedral shell when $1 < R_0/b_0 < 1.3$. All the structures corresponding to the capital letters are illustrated in Fig. 2.7a.

The thick dotted line in Fig. 2.4 shows that even though the energy per subunit increases as R_0/b_0 increases, $T = 1$ remains the global minimum energy structure till $R_0/b_0 = 1.3$ when the icosahedral $T = 3$ becomes the global minimum energy structure. This effect is also apparent in Fig. 2.5, which is a plot of the number of subunits n_s versus

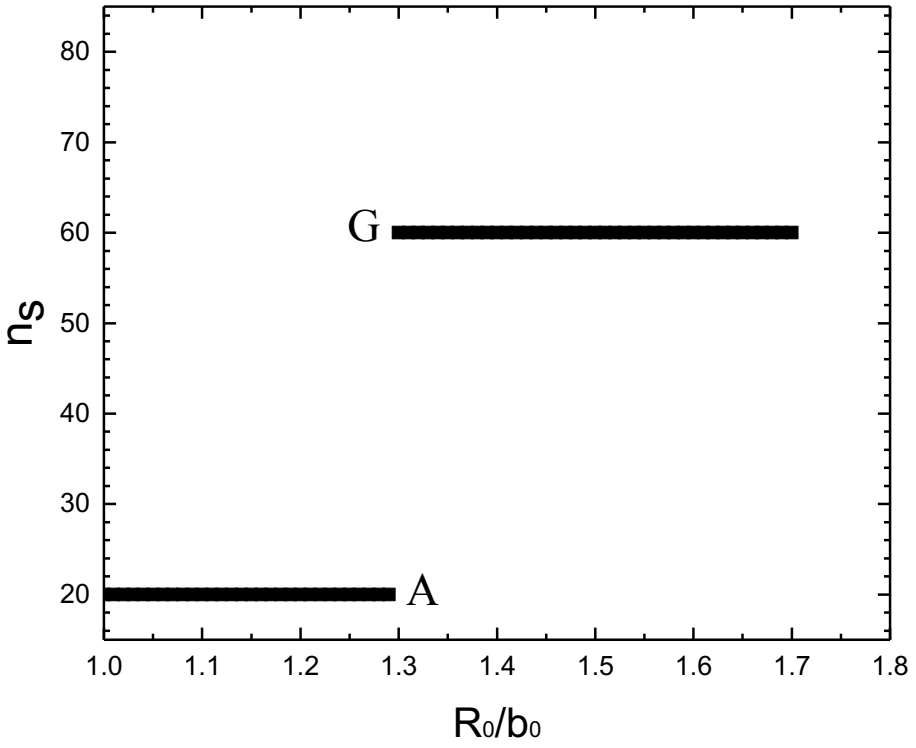

Figure 2.5: Plot of number of subunits in the equilibrium structures versus the spontaneous radius of curvature R_0/b_0 at $\gamma = 8$. The two flat lines in the plot correspond to $T = 1$ and $T = 3$ icosahedral structures.

R_0/b_0. There is a big jump in the number of subunits from $n_s = 20$ ($T = 1$) to $n_s = 60$ ($T = 3$) at $R_0/b_0 = 1.3$.

All the above effects can be seen more clearly in Fig. 2.7b in the form of a "structure" phase diagram of spontaneous radius of curvature R_0/b_0 and γ. Each shaded region in the diagram corresponds to a different shell whose structure and symmetry are illustrated in Fig. 2.7a. Figure 2.7b shows that the structures become more sensitive to the spontaneous radius of curvature as γ decreases. For instance, for $\gamma = 0.2$ as the spontaneous radius of

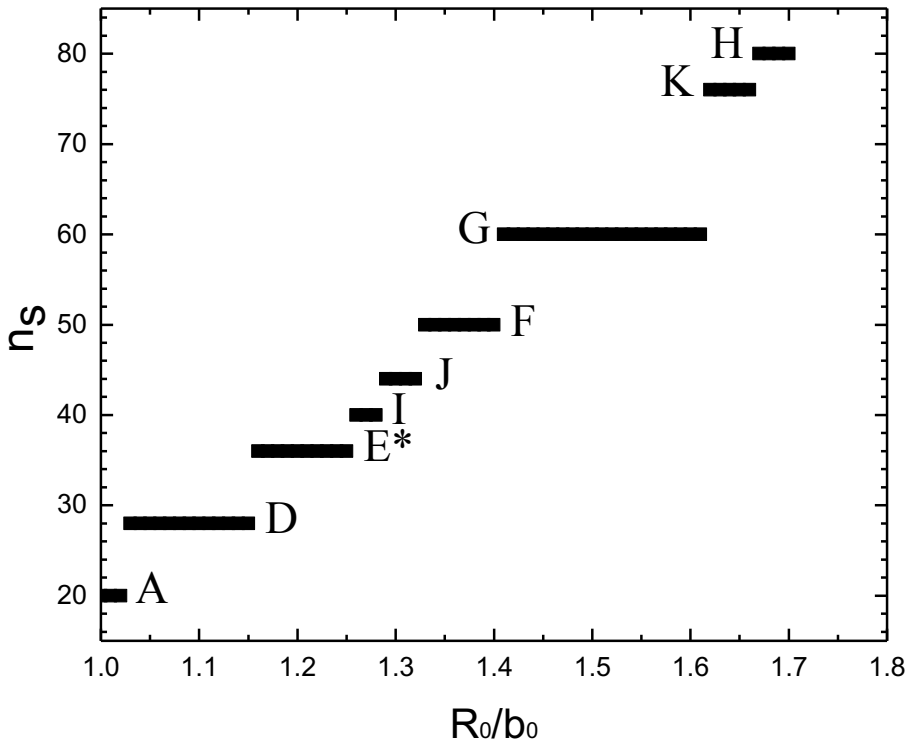

Figure 2.6: Number of subunits in the equilibrium structures versus the spontaneous radius of curvature R_0/b at $\gamma = 0.2$. The label next to each line shows the associated structure, illustrated in Fig. 2.7a .

curvature varies, we obtain nine different symmetric shells between $R_0/b_0 = 1$ and 1.7, see also Fig. 2.6. However, there are only two different structures at $\gamma = 8$ over a wide range of spontaneous curvature, $T = 1$ and $T = 3$. This is basically due to the fact that at larger γs the protein building blocks are stiffer and it is energetically more costly to deform them from their native shape. Since for icosahedral structures most proteins are sitting in equivalent positions, at high γ-values icosahedral structures are the minimum energy structures for the range of the spontaneous curvature studied, as illustrated in Fig. 2.7b.

The largest shell obtained in Fig. 2.7b contains $n_s = 80$ triangles corresponding to a $T = 4$ structure for smaller γ-values. Note that at intermediate γs, another equilibrium structure with the same number of subunits as a $T = 4$ shell ($n_s = 80$) but different symmetry exists, which we label it as H^* in Figs. 2.7a and 2.7b. While $T = 1$ and $T = 3$ occupy large regions in the equilibrium phase diagram, only a small region belongs to $T = 4$. This is consistent with several previous studies. First of all, a review of literature shows that there are fewer $T = 4$ structures in nature [9, 47]. Furthermore, in Refs. [11, 78] only the D_{5h} structure mentioned above was observed and no $T = 4$ icosahedral structures appeared in their simulations.

It is now interesting to compare the equilibrium "structure" phase diagram with the diagram obtained through irreversible assembly [99].

2.3.1 Equilibrium versus non-equilibrium

The non-equilibrium or irreversible structures are obtained based on the algorithm used in Ref. [99], and are illustrated in the form of a phase diagram of the dimensionless ratio of bending to stretching modulus γ and the spontaneous radius of curvature R_0/b_0 in

Fig. 2.8. As in the case of the equilibrium phase diagram, each color refers to a different symmetric structure. The details of obtaining non-equilibrium or irreversible structures are given in Chapter 3. Briefly, these structures grow following the local minimum energy pathway. Nevertheless, during the assembly process once a pentamer or hexamer forms, its position is permanently fixed. Due to these irreversible steps, the structures of assembled shells could be completely far from equilibrium. It appears that this "irreversible" algorithm has been successful in explaining many viral structures [44, 73, 56, 60].

We find it quite striking that the shell assembly along the local minimum energy path with the restrictive conditions of irreversible growth leads to the formation of shells almost identical to those obtained in equilibrium studies. These results are quite unexpected considering that the principles of detailed balance is violated in the irreversible growth and as such one would expect a big difference between the two phase diagrams.

Despite the similarities, there are some differences between the two phase diagrams, see Figs. 2.7b and 2.8. The shells that only appear in the irreversible phase diagram are illustrated in Fig. 2.9. Two small regions (structures B and C) in the irreversible phase diagram (Fig. 2.8) corresponding to $n_s = 24$ and 26 do not constitute the minimum free energy structures. In the equilibrium phase diagram, they are both replaced by the structure D, a clathrin shell, which has $n_s = 28$ and is called mini-coat. The other clathrin shells, hexagonal barrel (structure E) obtained in the irreversible growth has $n_s = 36$ with D_{6h} dihedral symmetry. The equilibrium structure of the shell with the same $n_s = 36$ has tennis ball symmetry, the structure E^* in Figs. 2.7b and 2.8.

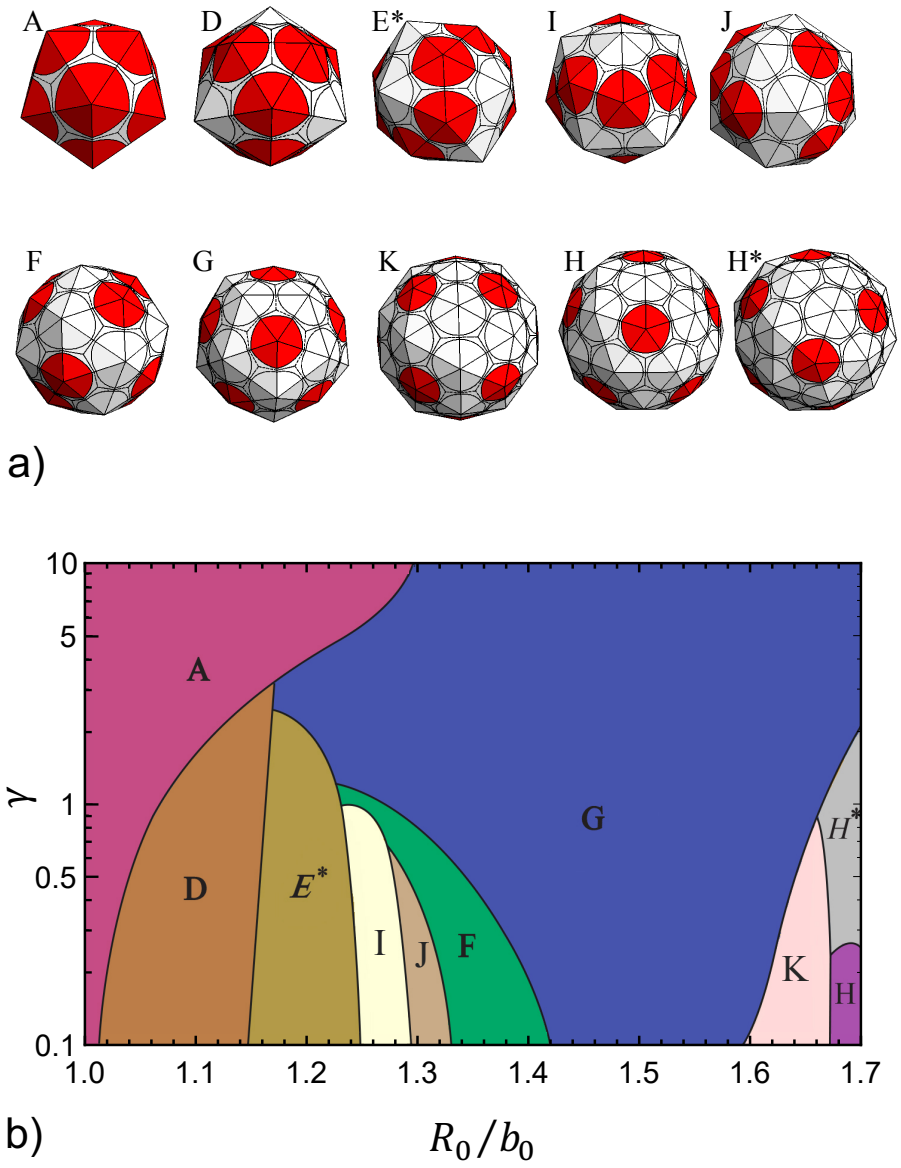

Figure 2.7: (a)The equilibrium structures obtained in the simulations corresponding to the labeled regions of the phase diagram illustrated in Figure b. The shells from left to right and top to bottom have $n_s = 20,\ 28,\ 36,\ 40,\ 44,\ 50,\ 60,\ 76,\ 80$ and 80 subunits and symmetries are icosahedral ($T = 1$), tetrahedral, D_2(tennis ball), D_2, D_2, D_3, icosahedral ($T = 3$), tetrahedral, icosahedral ($T = 4$) and D_{5h}, respectively. (b) Phase diagram of the equilibrium structures presenting various shells assembled for different values of γ and R_0/b. Each shaded region corresponds to a single equilibrium shell type . Region A corresponds to a shell with $n_s = 20$. The regions $D-K$ correspond to shells with $n_s = 28, 36, 50, 60, 80, 40, 44$ and 76 subunits. Both H and H^* structures have the same number of subunits $n_s = 80$.

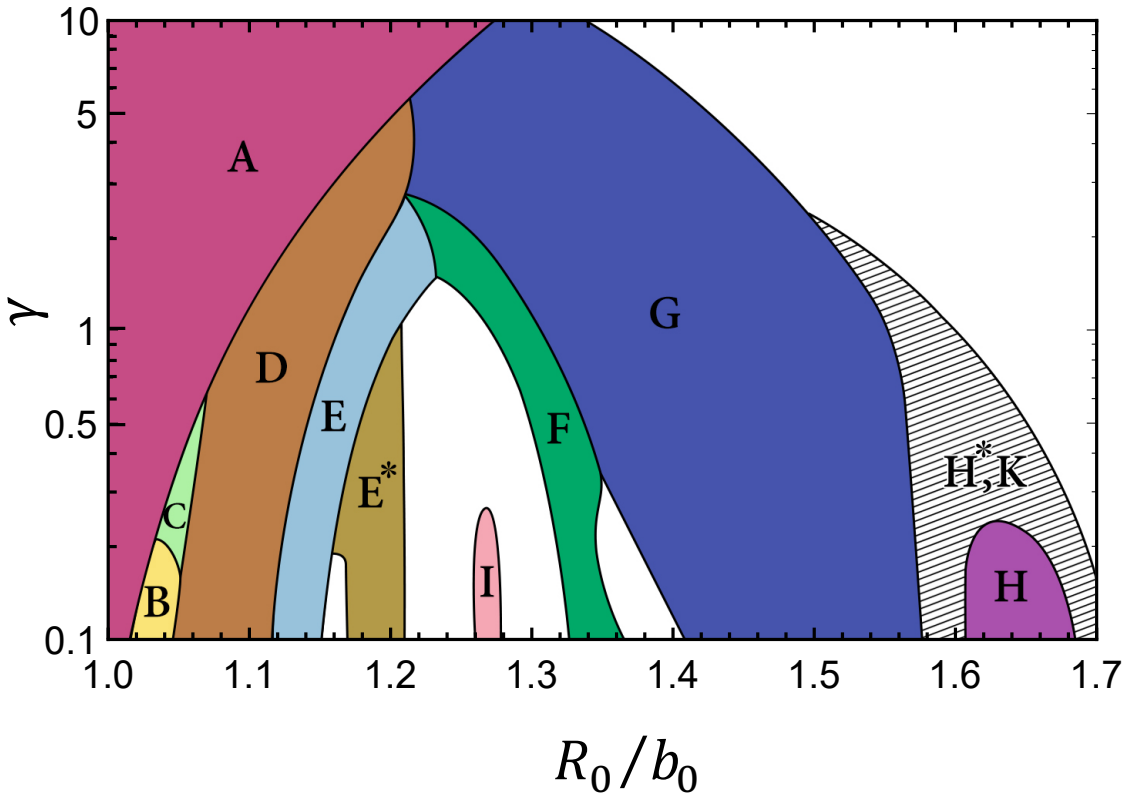

Figure 2.8: Phase diagram of the empty shells assembled for different values of γ and R_0/b_0 using the deterministic model. Each shaded area corresponds to a single type of structure that forms in that region. Shell structures with their corresponding symmetries are shown in Fig. 2.7 a and Fig. 2.9. In the hashed regions, in addition to symmetric shells, a number of similarly sized non-symmetric structures form. The white areas correspond to regions in which different types of shells without any specific symmetry are grown [99].

The white area in the irreversible phase diagram for $\gamma < 2$ and between $1.21 < R_0/b < 1.3$ ($n_s = 38 - 48$) corresponds to the region where many different types of shells without any specific symmetry are assembled. In contrast, there is no irregular structure in the equilibrium phase diagram, and we find the structures I ($n_s = 40$) and J ($n_s = 44$) with D_2 symmetry in that region.

Moreover, the regions corresponding to G and F structures cover a larger parameter space in the equilibrium phase diagram compared to the irreversible one, revealing the presence of energy barriers as the symmetric shells grow. The energy barrier is between a local minimum energy that the growing shell is trapped into it because of the irreversible assembly path and the global minimum energy.

Furthermore, the irregular shells formed between the G and F structures at lower γ-values in the irreversible phase diagram disappear and are replaced with the F one. The structure K with $n_s = 76$ and tetrahedral symmetry which forms between G and H (or H^*) in the equilibrium phase diagram, appears rarely in the irreversible one. In fact it only assembles at the boundary between the hashed and the white regions (Fig. 2.8), despite the fact that the K structure is smaller than H or H^*. Last but not least, the structures with icosahedral symmetry cover a wider region in the equilibrium phase diagram. The largest symmetric shell in Figs. 2.7b and 2.8 is $n_s = 80$. Note that in the absence of a cargo which could be a genome, an inner shell or scaffolding proteins, at low γ, when we increase the spontaneous radius of curvature only irregular shapes form. However, at large γ-values for ($R_0/b_0 \gg 1$), we obtain flat sheets or other structures with zero Gaussian curvature. In this study, we only focus on the assembly of small symmetric shells, as illustrated in Figs. 2.7b.

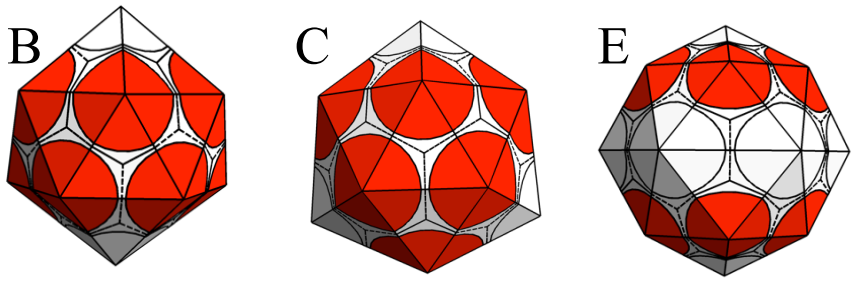

Figure 2.9: The symmetric shells growing under irreversible assembly conditions, which do not appear in the equilibrium phase diagram. The shells from left to right have $n_s = 24,\ 26$ and 36 subunits. The symmetries are D_6, D_3 and D_{6h} respectively. In the irreversible phase diagram (Fig. 2.8) they are labeled as $B,\ C$ and E. The E structure has the same number of subunits as the E^* structure but with different symmetry.

2.4 Summary

Despite the wide range of amino acid sequences and folding structures of coat proteins, many protein cages spontaneously self-assemble to form icosahedral or other symmetric structures. This reveals a "universal" behavior among most protein cages in nature. In this research, using the MC simulation and edge swapping method, we investigated the equilibrium structure of protein cages built from identical subunits. We, in particular, choose triangular subunits as they are a generic choice that applies to the majority of protein nano-cages. Since triangles form a hexagonal lattice in flat space and also assemble to create 12 disclinations for making closed polyhedrons, their behavior is similar to the building blocks of protein cages. While in this study and many others equilateral triangular subunits have been able to capture the universal behavior of viruses and other proteins cages [63, 98, 44, 72, 73, 60], we do not expect that the equilateral triangular mesh explains the behavior of systems in which anisotropy of subunits plays an important role in the final structure of protein cages.

Using triangular subunits, we studied the impact of the mechanical properties of building blocks on the symmetry and structure of small protein cages and constructed a phase diagram as a function of the spontaneous curvature of subunits and FvK number (the ratio of stretching to bending modulus), as shown in Fig. 2.7b. The phase diagram is significantly occupied with icosahedral shells, $T = 1$ and 3, which are common among viruses and many other protein cages. As illustrated in the figure, at low γ, where subunits can deform easily from the equilateral triangle, various structures form as a function of spontaneous curvature. However, by increasing γ, subunits become more rigid and structures with lower symmetries disappear. For instance for low $\gamma = 0.2$, the equilibrium structures are sensitive to the spontaneous radius of curvature and there is a smooth transition from one shell to the next one as illustrated in Fig. 2.6. Increasing to $\gamma = 8$, the equilibrium structures become less sensitive to R_0/b_0 and only icosahedral structures with $T = 1$ and 3 survive, indicating the robustness of these two structures.

Quite unexpectedly, we found that the equilibrium phase diagram, Fig. 2.7b, was very similar to the phase diagram obtained under irreversible conditions, see Fig. 2.8. As explained in Chapter 2, the irreversible simulations of Ref. [99] were performed following the local minimum energy path but under the condition that once a pentamer or hexamer formed, it could no longer dissociate or move. Since the principles of detailed balance were violated in the irreversible growth simulations, we did not expect to observe such a high degree of similarity between the two phase diagrams, see Figs. 2.7b and 2.8.

These results could be explained to some extent with the recent work of Li *et al.* who employed the continuum elasticity theory and studied the assembly pathway of

icosahedral shells. They found that as an elastic shell grows, there is a deep potential well attracting pentamers exactly at the locations that will become the vertices of an icosahedron when the shell is complete [60].

Due to the small size and discreteness of the system, the continuum elasticity theory cannot explain the symmetry of the shells observed in Figs. 2.7a and 2.9. However, the extensive similarities between equilibrium and irreversible phase diagrams in Figs. 2.7b and 2.8 indicate that for the symmetric shells other than icosahedral ones, there are also high-affinity regions for the formation of disclinations at specific locations during the growth process, which leads to the assembly of different types of symmetric shells, depending on the mechanical properties of protein subunits.

It is important to note that we often found one single global minimum energy structure in the phase diagrams presented in Figs. 2.7a and 2.9 for a given γ and R_0/b_0. However, in many biological systems, sometimes a few different types of protein cages co-exist in the same solution [64, 53]. For instance self-assembly studies of dimeric Hepatitis B Virus capsid protein mutant Cp1492 shows that empty $T = 3$ and $T = 4$ structures form in a ratio of about 95:5 at medium to high salt concentration and close to neutral pH [70, 101]. While the focus of our work is to find the optimal structure of protein cages as a function of mechanical properties of its building blocks, the polymorphism observed in several self-assembly studies can be explained through a careful examination of plots of energy per subunit versus number of subunits in Fig. 2.3. If the difference between the free energy per subunits in two different structures is small compared to the thermal energy k_BT, one expects to observe both structures, with relative populations given by the corresponding

Boltzmann factor $\exp(\Delta\epsilon/k_BT)$, with $\Delta\epsilon$ the difference between the free energy per subunit in the two shells.

Lastly we emphasize that even though in this work we did not explicitly study the impact of salt and pH on the structure of protein shells, the solution environment such as salt and pH can modify the number of charges on the protein subunits, which in turn can change the stiffness and the spontaneous dihedral angle of building blocks. While the results obtained in this chapter can explain why various protein shells with different symmetry appear in nature, at this point there is not enough experimental data to allow us to connect our variables R_0/b_0 and γ (the ratio of stretching to bending modulus) to the experimental conditions such as pH and salt.

To examine several concepts presented in this article, it would be interesting to carry out a set of systematic experiments as a function of pH and salt concentration with various mutated proteins, which in consequence have different mechanical properties. One then can construct an experimental phase diagram similar to the one shown in Fig. 2.7b. A quantitative comparison between experiments and theory will result in a better understanding of the protein-protein interaction and the parameters that contribute to the formation of various protein cages with extensive potential application in various area of material science, gene delivery and medicine.

Chapter 3

Virus Assembly Pathways inside a Host Cell

3.1 Introduction

Positive-strand RNA viruses represent the largest genetic class of viruses, including many human, animal, and plant pathogens [50]. Their virions contain mRNA-sense single stranded (ss) RNA which is protected by a protein cage called the capsid [30]. The capsids of more than half of the RNA virus genera have icosahedral symmetry [83, 100, 48, 96]. Since the spontaneous assembly of molecular constituents into a functional virion is one of the key steps in the virus life cycle, its physical and chemical bases have been intensely studied on a number of model systems. A wealth of information has emerged from *in vitro* assembly studies and detailed models that explain several salient features of the virus self-assembly have been proposed [103, 42, 11, 23].

For all ssRNA icosahedral viruses, genome encapsidation happens at assembly, in cytosolic compartments where the opportunity of encapsidating other, non-cognate cargo also exists [83]. *In vitro* self-assembly studies suggest that the negatively-charged RNA recruits coat proteins (CPs) through multiple electrostatic interactions with their basic residues [46, 40, 20, 59, 27, 108]. The dominant role of electrostatics is supported by *in vitro* encapsidation experiments with a variety of polyanionic non-cognate cargo, ranging from other nucleic acids to nanoparticles [91, 36]. Despite a relatively good understanding of this process [24, 40, 103, 7, 29], the origin of selectivity of cognate RNA *in vivo* encapsidation has remained a mystery because of the presence of other non-genomic polyanionic species. One explanation involves certain packaging sequences which confer a higher coat protein binding affinity to cognate viral RNA [89, 97]. However, it has been pointed out that such packaging sequences might not be the sole source of specificity, with other factors being likely at play [15]. Recently, in discussing experiments that entailed a combination of charge detection mass spectrometry and cryo-electron microscopy, Bond *et al.* suggested capsid elastic stress at assembly as a pathway selection mechanism. In their experiments, brome mosaic virus (BMV) coat proteins were found to readily form capsids around multiple oligonucleotides. Those capsids were smaller, non-icosahedral, and less stable than the native ones. Based on these findings, Bond *et al.* speculated that shells do form around small RNAs and other polyanionic cargo but might easily fall apart, with their fragments re-assembling around cognate RNAs when they become available. This hypothesis suggests a narrowing of the possible virus assembly pathways, which would actually take advantage of the presence of non-cognate polyanions [5]. To test theoretically the possibility of this hypothesis, we have

modeled capsid growth under the assumption of an elastic lattice formed of subunits having a preferred radius of curvature.

According to the quasi-equivalence principle, the number of proteins forming icosahedral shells is $60T$ where the T number is a structural index for viral shells and is equal to $T = h^2 + k^2 + kh$ with h and k nonnegative integers [10]. Previous *in vitro* assembly studies of small icosahedral ($T = 3$) viruses found that capsid proteins can assemble *in vitro* around a variety of anionic cargoes to form isometric shells [37, 15]. In the absence of high-resolution structural or stoichiometric data, it was assumed that even if the size of the shells changes because of the size of their cargoes, the symmetry of the capsids is still the same as that of the native shell, *i.e.*, there are usually 12 pentamers, sitting at the vertices of an icosahedron, in a shell built otherwise from hexamers. The experimental results of Bond *et al.* noted above were unexpected as they found that nucleic acid oligomers and capsids proteins form structures with D_{6h} and D_{5h} symmetries [5]. Figure 3.1 shows the cryo-EM images of two different structures obtained in Ref. [5].

In this chapter, we study the range of elastic and geometric subunit parameters for which strained, lower-symmetry non-icosahedral structures similar to those reported by Bond *et al.* form. We calculate the elastic stress distribution across these shells and discover that they should be less stable than the wild type $T = 3$ shells, being strained. We also show that one of the low symmetry structures (Fig. 3.1C) constitutes the minimum energy structure when smaller size cargoes are encapsulated, while the other one (Fig. 3.1A) could be obtained only through non-equilibrium simulations [5]. Our theoretical study thus confirms the possibility of a pathway for the packaging process inside cells whereby capsid

proteins may start assemble around non-cognate small RNAs in the cell during infection, forming shells that are strained and less stable.

Since there is no systematic study about the impact of RNA on the capsid size and morphology as observed in the above experiments, in this chapter we also explore under what conditions RNA can change the size and structure of viral shells. We find that the interplay of the mechanical properties of capsid proteins and their interaction with genome is quite intriguing and that RNA can significantly modify the symmetry of viral shells giving rise to the formation of structures with different symmetries that were not found in the previous theoretical studies on empty capsids.

This study lays out the systematic comparison of theory and experiments, which will allow us to gain a better understanding of the role of RNA in the capsid assembly pathway, stability and structure, both *in vivo* and *in vitro*. A deeper understanding of the role of genome in virus assembly mechanisms could lead to the design principles for alternative antivirals and facilitate the fabrication and design of precise synthetic nanostructures.

3.2 Method

To study the formation of capsids around different genome sizes, we use triangular subunits similar to chapter 2 and associate a monomer to each triangle vertex (Fig. 3.2a). As we mentioned in chapter 2, the elastic energy of a growing capsid, is the sum of the stretching E_s and bending E_b energies that are defined by equations 2.1 and 2.2.

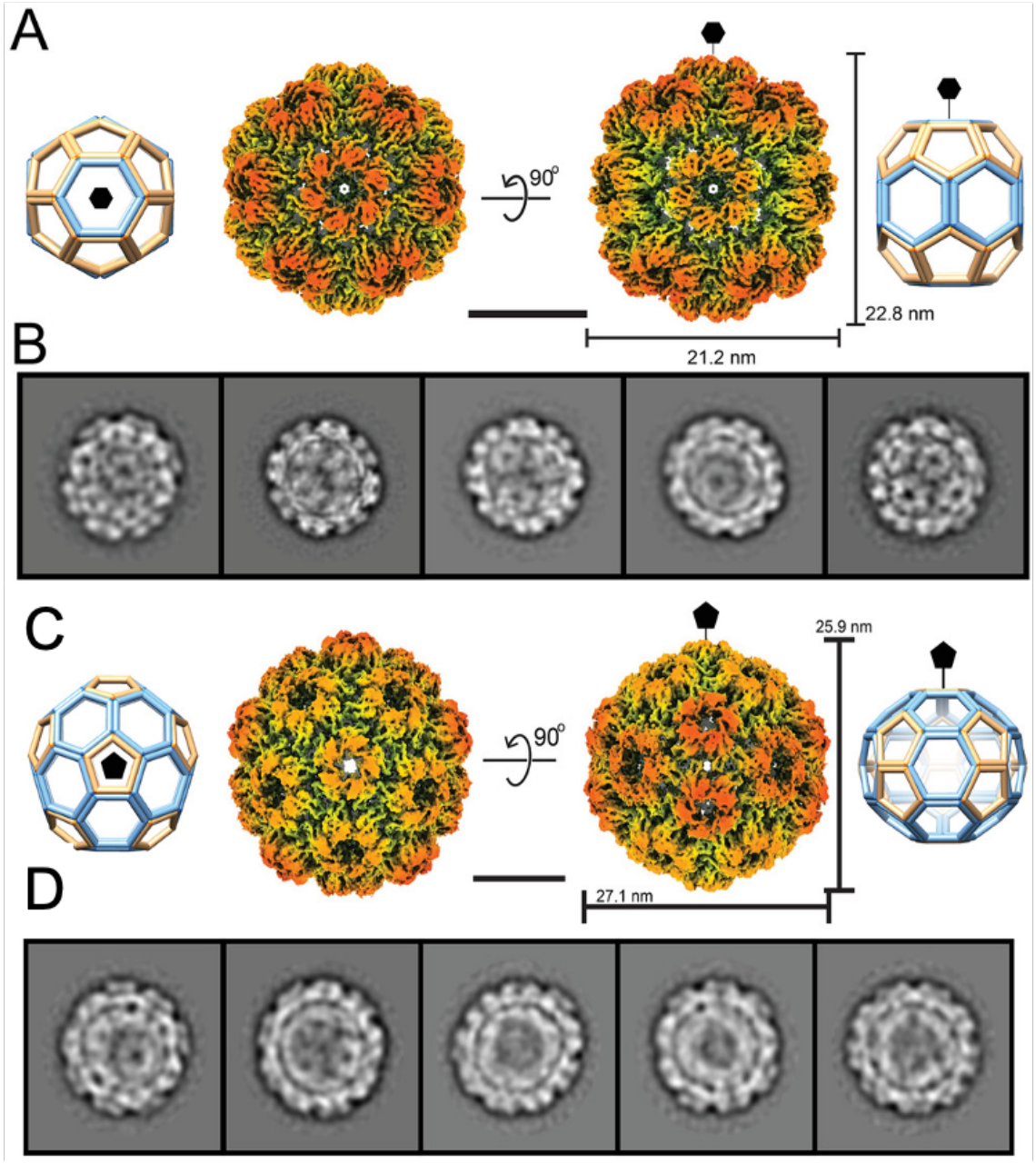

Figure 3.1: Cryo-EM of E and F^* structures with D_{6h} and D_{5h} symmetries obtained in the experiments of Ref. [5]. Note that in [5] the structures E and F^* are called $H8$ and $H15$, respectively. A) Isosurface views of the E structure at different angles. The axis of sixfold symmetry is marked. Scale bar is 10 nm. B) Selected reference-free 2D class averages indicate different orientations of the particle. C) Isosurface views of F^* structure at different angles. The fivefold symmetry axis is marked. D) Selected reference-free 2D class averages show different orientations of the particle. Scale bar is 10 nm. Adapted with permission from [5].

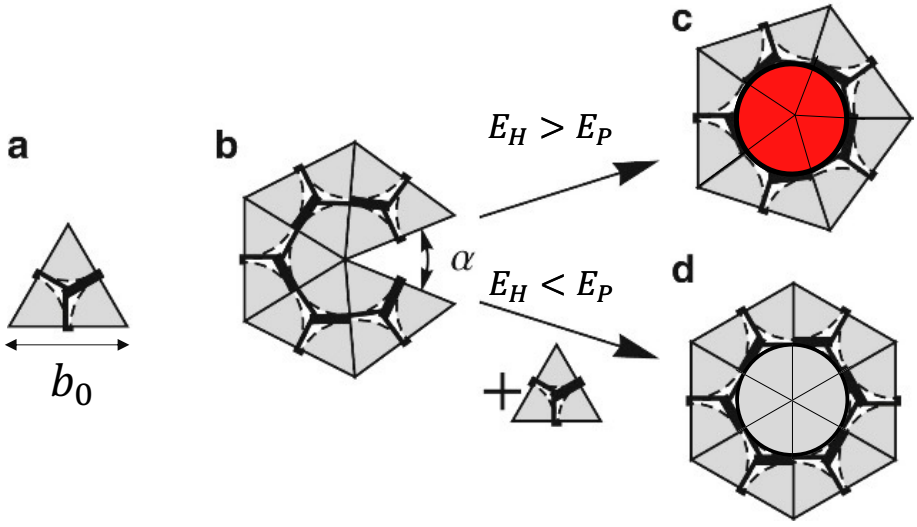

Figure 3.2: (a) A protein subunit built from three monomers illustrated as an equilateral triangle in its equilibrium shape. As triangular subunits are added to the growing shells, because of the curvature of the shell, the subunits can be stretched or compressed. (b) The subunits bind together edge to edge, and the growth of the shell proceeds by adding a subunit to the location with the smallest opening angle, α. As the shell grows, the two unbound edges can either bind together to form a pentamer (c), or a new subunit can be added to form a hexamer (d). The choice between forming a pentamer and a hexamer is based on which leads to a lower energy per subunit in the growing shell. If the energy per subunit becomes lower for the formation of a hexamer (E_H) compared to that of a pentamer (E_P), then a hexamer forms; otherwise a pentamer assembles. Adapted with permission from [99].

To grow a capsid, we consider that the assembly follows the locally minimum energy path, which, under many circumstances, is reasonably close to the most probable growth pathways and thus can yield a meaningful representation of the final assembly products. During the simulations at each growth step a new subunit is added to the growing edge of the incomplete capsid at a position in which the number of neighbors at the vertices of the newly added subunit will be maximized. Hence, the new subunit will be added to a position with the smallest opening angle α, see Fig. 3.2b. A defining step in the assembly process corresponds to the formation of pentamers illustrated in Fig. 3.2. If there are already 5 triangles attached to a vertex on the rim of an incomplete capsid, then the growth can proceed in two different pathways: (i) a pentamer can be formed by attaching

the two neighboring edges, see Fig. 3.2c, or (ii) a hexamer can be assembled by inserting a new subunit, Fig. 3.2d. Depending which structure has a lower energy per subunit in the growing shell, a pentamer or a hexamer forms [74].

Note that the algorithm discussed above for the virus growth corresponds to the physical situation in which the line tension of incomplete capsid is high or subunit-subunit interaction is weak enough such that the proteins can easily move around to explore the energy landscape and then attach themselves to a position where their number of neighbors are maximized [57]. Furthermore, we allow the shell to relax between the addition of every subunit assuming that the relaxation time of the elastic stresses in the incomplete capsid is shorter than the time it takes for the shell to grow. This is compatible with the assumption of the high line-tension or weak protein-protein interaction [60, 105]. This approach is also consistent with the results of Ref. [23], in which the self-assembly simulations of $T = 1$ structures at different protein concentrations show that slowing the adsorption of free subunits into the core decreases frustration [23].

We emphasize that our simulations are deterministic and thus irreversible, mimicking the situation in which the strength of protein-protein interaction is such that once a pentamer or hexamer forms, it can no longer dissociate [74, 57]. Quite interestingly, in our simulations, we do not see aberrant particles as observed in the assembly of $T = 1$ structures in Ref. [24]. This might be due to the fact that we follow the most probable pathway [57].

To find the locations of high stress points across the capsid, we employ the virial formula and calculate the stress component for vertex v_j as follows,

$$\sigma_{v_\alpha}^{ij} = \frac{1}{2A}\sum_{v_\beta}(x_{v_\beta}^i - x_{v_\alpha}^i)f_{<v_\alpha,v_\beta>}^j, \tag{3.1}$$

where $A = Z\frac{1}{3}\frac{\sqrt{3}}{4}b_0^2$ is the area of vertex v_α with Z the coordination number and $x_{v_\alpha}^i(x_{v_\beta}^i)$ is the i^{th} component of the position of vertex $v_\alpha(v_\beta)$. The summation is over the nearest neighbor vertices. The quantity $f_{<v_\alpha,v_\beta>}^j$ is the j^{th} component of the force on vertex v_α in the $< v_\alpha, v_\beta >$ bond and is equal to

$$f_{<v_\alpha,v_\beta>}^j = k_s(r_{<v_\alpha,v_\beta>} - b_0)\frac{x_{v_\beta}^j - x_{v_\alpha}^j}{r_{<v_\alpha,v_\beta>}}, \tag{3.2}$$

where k_s denotes the stretching modulus, $r_{<v_\alpha,v_\beta>}$ is the distance between v_α and v_β, and b_0 is the equilibrium length of the bond. Combining Eqs. 3.1 and 3.2, we obtain

$$\sigma_{v_\alpha}^{ij} = \frac{k_s}{2A}\sum_{v_\beta}(r_{<v_\alpha,v_\beta>} - b_0)\frac{(x_{v_\beta}^i - x_{v_\alpha}^i)(x_{v_\beta}^j - x_{v_\alpha}^j)}{r_{<v_\alpha,v_\beta>}}. \tag{3.3}$$

the elastic stress distribution in the shell which is trace of the stress tensor thus becomes,

$$\sigma_{v_\alpha} = \sum_{v_\beta}(\frac{k_s}{2A})r_{<v_\alpha,v_\beta>}(r_{<v_\alpha,v_\beta>} - b_0) \tag{3.4}$$

where the summation is over all vertices sharing a bond with the vertex v_α. The $r_{<v_\alpha,v_\beta>}$ is the distance between vertices α and β.

To model the genome, we consider that viral RNA in solution is believed to be folded in such a way that its radius of gyration only changes by a factor of 3 inside the capsid [102]. Moreover, several *in vitro* virus self-assembly experiments indicate that RNA collapses into its final size right at the beginning of the assembly after a small amount of proteins are

adsorbed to it, suggesting that the time scale associated with genome condensation is much faster than the overall time scale for the growth process [6, 14, 18]. Therefore, we assume that capsid growth occurs in presence of a globular RNA modeled by a spherical core, which reduces considerably the many degrees of freedoms involved in the problem. This will make the assembly process more tractable, allowing us to explore how the shell size and structure depends on the genome size and the subunit-genome interaction. Thus, we consider a spherical core interacting through Lennard-Jones (LJ) with the capsid proteins

$$E_{lj} = \sum_{v_\alpha} 4\epsilon_{lj}[(\frac{\sigma}{r_{<v_\alpha,g>}})^{12} - (\frac{\sigma}{r_{<v_\alpha,g>}})^{6}] \tag{3.5}$$

where ϵ_{lj} is the potential strength, σ is the distance at which the LJ potential between the core and triangle vertex (monomer) is zero and $r_{<v_\alpha,g>}$ is the distance between the core center and triangle vertex v_α.

We start the growth process by adding one subunit to the core and then we monitor the growth of the shell. We assume that the solution condition is always such that the proteins prefer to assemble rather than to stay free in the solution. We consider that the concentration of free proteins in solution is such that the capsid proteins nucleate only in one location along the genome, and the rest of capsid grows around the nucleating site.

In the results and discussion section, we will discuss the conditions under which if the spontaneous radius of proteins is smaller than the size of the core, the shell might not grow around the genome but will form an empty shell.

3.3 Results and Discussion

To explore the impact of RNA on the growth pathway, morphology and stability of viral shells, we construct a "structure" phase diagram as a function of the protein spontaneous curvature R_0/b_0 and the FvK number γ, for various genome size and the strength of RNA-protein interactions. We seek to determine how the phase diagram of genome-containing particles differs from the phase diagram of the empty shells in 2.8 and 2.7b in the absence of RNA. Note that figure 2.8 shows the results of the deterministic simulations as a two dimensional phase diagram as a function of R_0/b_0 and γ in the absence of genome. Figure 2.8 clearly indicates that at higher γs, *i.e.*, for the shells that are easier to bend than to stretch, we observe only a transition from $T = 1$ (structure A) to $T = 3$ (structure G) as a function of R_0/b_0. However, at lower γs, in addition to $T = 1$ and $T = 3$ shells, several structures with other symmetries form as R_0/b_0 varies. All structures have 12 pentagons, even the aberrant ones assembled in the white region of the figure.

As discussed in the method section, to make our calculations more tractable while keeping the important physical features of the system, we model the genome as a spherical core that interacts attractively with the protein subunits through LJ potential. During the simulations, we have considered that the attractive interaction between protein subunits and the genome is uniform and does not depend on the specific sequence of RNA. However, at least for certain viruses, the binding affinity is thought to vary along the genome with segments containing the packaging signals showing higher affinity than the rest [95]. This is important for the initial stages, but it is not clear that that the packaging signals are still playing a role after the initial nucleo-protein complex has formed and the genome has

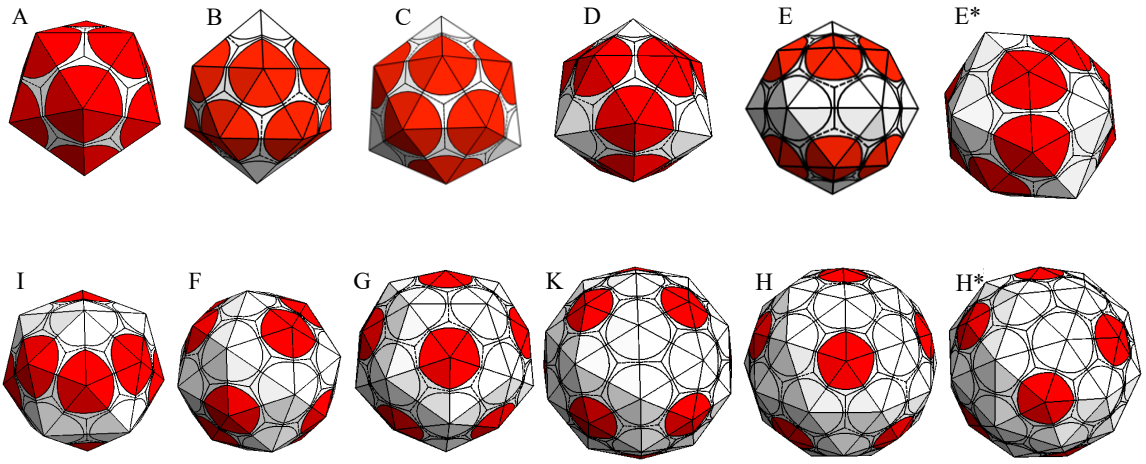

Figure 3.3: The symmetric structures obtained in the deterministic and Monte Carlo simulations corresponding to the labeled regions of the phase diagrams illustrated in Fig. 2.8 and Fig. 2.7b. The shells from left to right and top to bottom have 20, 24, 26, 28, 36, 36, 40, 44, 50, 60, 76, 80 and 80 subunits or 12, 14, 15, 16, 20, 20, 22, 24, 27, 32, 40, 42 and 42 capsomers respectively. Symmetries are icosahedral ($T = 1$), D_6, D_3, tetrahedral, D_{6h}, D_2 (tennis ball), D_2, D_2, D_3, icosahedral ($T = 3$), tetrahedral, icosahedral ($T = 4$) and D_{5h}, respectively. Red and white colors correspond to the location of pentamers and hexamers respectively [99, 77].

assumed a globular shape. Since the focus of the simulations is on the assembly process after the genome has assumed the globular shape, and not on the kinetics, the simplifying assumption of a spherical core is reasonable as confirmed by the fact that we obtain the same structures as observed in the experiments. When examining the influence of the core on the particle phase diagram, it is particularly interesting to focus on particle sizes that are compatible with $T = 1$ and $T = 3$ structures. Note that $T = 1$ was observed in the *in vitro* experiments with Bromo Mosaic Virus (BMV), whose native morphology is a $T = 3$ structure [65].

We consider four different core radii $R_c/b_0 = 0.95, 1.2, 1.4, 1.5$ that commensurate with the size of $T = 1$, "E and E^*", "F and F^*", and $T = 3$ structures, respectively. Note that E and E^* have the same size, as it is the case for F and F^*. We also monitor the assembly of capsids around a core with the radius $R_c/b_0 = 1.3$ that is not compatible with

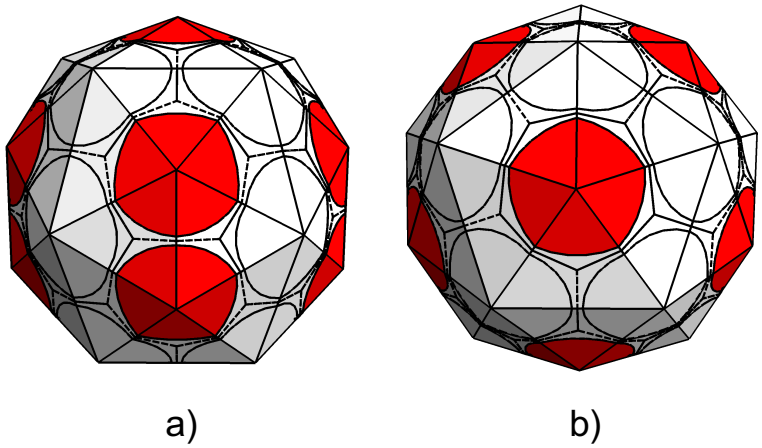

Figure 3.4: Structure F^* with D_{5h} symmetry has 50 subunits and 27 capsomers. a) Side view of F^*. It has five pairs of pentamers on the side, one single pentamer at the top and another one at the bottom. b) Top view of the F^* structure.

the size of any of the symmetric shells in Fig. 3.3. We emphasize that F^* (see Fig. 3.4) is exactly the structure that has been discovered in the recent self-assembly experiments of Bond *et al.* [5] (see Fig. 3.1C) and is of particular interest to our investigation of the role of genome in the structure of capsids and the virus assembly pathway in the presence of non-native RNAs. One of the interesting results of this study is that the presence of the core is essential for the formation of F^*, which did not appear in the previous simulations of equilibrium [77] and non-equilibrium [99] structures for empty shells that we studied in chapter 2.

Figure 3.5 shows the results of the simulations as two dimensional phase diagrams as a function of γ and R_0/b_0. From the top to the bottom, the rows correspond to the core sizes of $R_c/b_0 = 0.95, 1.2, 1.3, 1.4, 1.5$, respectively. The figure also illustrates the impact of the strength of genome-protein attractive interaction. The left column in Fig. 3.5 corresponds to the structures for which the strength of LJ is $\tilde{\epsilon}_{lj} = \epsilon_{lj}/k_b = 0.005$, and

the right column shows the results for stronger LJ interaction, $\tilde{\epsilon}_{lj} = 0.1$. Note that we divide ϵ_{lj} by k_b to make it dimensionless. Similar to Fig. 2.8, each color corresponds to a specific symmetric structure. In the dotted regions of the figure, the shells assemble without encapsulating the genome because the cost of its packaging is too high. The white areas show the regions in which non-symmetrical shells are formed.

The phase diagrams on the left hand side of Fig. 3.5 are very similar to the phase diagram in Fig. 2.8, revealing that the elastic energy of the growing shell has the dominant contribution to the formation of the symmetric shells in the weak LJ regime ($\tilde{\epsilon} = 0.005$). The simulations are designed such that if the size of the core is smaller than the spontaneous curvature of proteins, the strength of LJ potential is such that the shell will always assemble around it. As the size of the core increases, the subunits whose spontaneous radius of curvature is smaller than the core radius will either assemble into an empty shell based on the phase diagram of Fig. 2.8 for the relevant γ and R_0/b_0 or into a filled shell but with a different symmetry than the empty one for the same γ and R_0/b_0. Thus, the main difference between the phase diagrams on the left hand side of Fig. 3.5 ($\tilde{\epsilon}_{lj} = 0.005$) and the ones for the empty shells (Fig. 2.8) arises from the size of the core. At $\tilde{\epsilon}_{lj} = 0.005$, the elastic energy still controls the growth pathway, and as a result the transition lines between different structures move together without influencing the shape varieties. Similar to Fig. 2.8, we observe two main structures form ($T = 1$ and $T = 3$) at high γ values (Fig. 3.5), while different structures form as R_0/b_0 changes for lower γs.

By contrast, if the strength of subunit-genome interaction is strong ($\tilde{\epsilon}_{lj} = 0.1$), the presence of RNA can completely change the structure of the capsids as illustrated in

the right column of Fig. 3.5. The tendency of trimers to aggregate on the core surface because of the LJ potential will compete with the bending energy of the shell, resulting in a considerable change in the phase diagram. Most structures in the phase diagram then will have the shapes instructed by the core size. For example, when the core size is $R_c/b_0 = 0.95$, the optimal size for a $T = 1$ shell, Fig. 3.5a shows that for the weak LJ potential, $T = 1$ covers a small region in the phase diagram. However, for the strong LJ potential the $T = 1$ structure covers a very wide region of the phase diagram (Fig. 3.5b).

Further, Fig. 3.5b reveals that for $R_0/b_0 > 1.12$ at low γs, only irregular shells without any specific symmetry form, in contrast to Fig. 3.5a where various symmetric structures form. The irregular shells have larger number of subunits compared to a $T = 1$ structure in Fig. 3.5b. At low γs when the subunits are less rigid, they squeeze each other to make room for more subunits because of the strong attractive interaction between the proteins and the core. As we keep increasing γ at $R_c/b_0 = 0.95$, the structures D, C, B and $T = 1$ start appearing, respectively. It is important to note that while these shells have different symmetries and number of subunits, they all have more or less the same size due to the strong attraction of the subunits to the core. Figure 3.5b indicates that the largest symmetric capsid that forms for $R_c/b_0 = 0.95$ is the structure D with tetrahedral symmetry and 28 triangular subunits. As γ increases, the number of subunits decreases and $T = 1$ becomes the dominant structure, regardless of the value of R_0/b_0. Without the core, the $T = 1$ structure would not form if $R_0/b_0 > 1.28$.

Increasing the genome size to $R_c/b_0 = 1.2$, we observe empty $T = 1$ structures at weak $\tilde{\epsilon}_{lj}$ and $R_0/b_0 < 1.2$ (Fig. 3.5c). However, for the strong $\tilde{\epsilon}_{lj}$ the most dominant

symmetric structures are E and E^* (Fig. 3.5d) with hexagonal barrel and tennis ball symmetries, respectively. Both structures have 36 protein subunits. Figure 3.5f shows that while for $R_c/b_0 = 1.3$ at strong $\tilde{\epsilon}_{lj}$ the structures J, E, E^*, F and F^* form, for the weak $\tilde{\epsilon}_{lj}$ several additional symmetric structures form 3.5e. We note that the capsids with the number of subunits less than 36 assemble without packaging the genome because the encapsidation of genomes whose size is much larger than the capsid size is energetically forbidden due to the high cost of elastic energies.

Figure 3.5g shows the phase diagram for the core size $R_c/b_0 = 1.4$. For this case, all the capsids with less than 40 subunits are empty. In contrast, Fig. 3.5h shows that at $R_c/b_0 = 1.4$, if $\tilde{\epsilon}_{lj}$ is strong, all the capsids encapsidate the genome, and only symmetric structures F and G form. As the core size increases to $R_c/b_0 = 1.5$, Fig. 3.5j shows that for strong $\tilde{\epsilon}_{lj}$, the structure G with $T = 3$ symmetry covers a large area of the phase diagram. However, at weak $\tilde{\epsilon}_{lj}$, structures with different symmetries form.

The results presented in Fig. 3.5 clearly show that the genome can completely modify the structure of the capsid. In particular, the structure F^* in Fig. 3.5d that has not been previously observed in the empty shell simulations, is exactly the structure called H15 obtained in the self-assembly studies of BMV shown in Fig. 3.1C and D [5]. The structure E^* also covers a broad region of the phase diagram (Fig. 3.5d), while this structure appears in a tiny region of the phase diagram for the empty shells (Fig. 2.8). We will discuss these structures and their relation with the experiments [5] in the summary section.

One of our other intriguing results corresponds to the structure D, the so called mini-coat that appears in the assembly studies of clathrin shells too [32]. Quite interestingly,

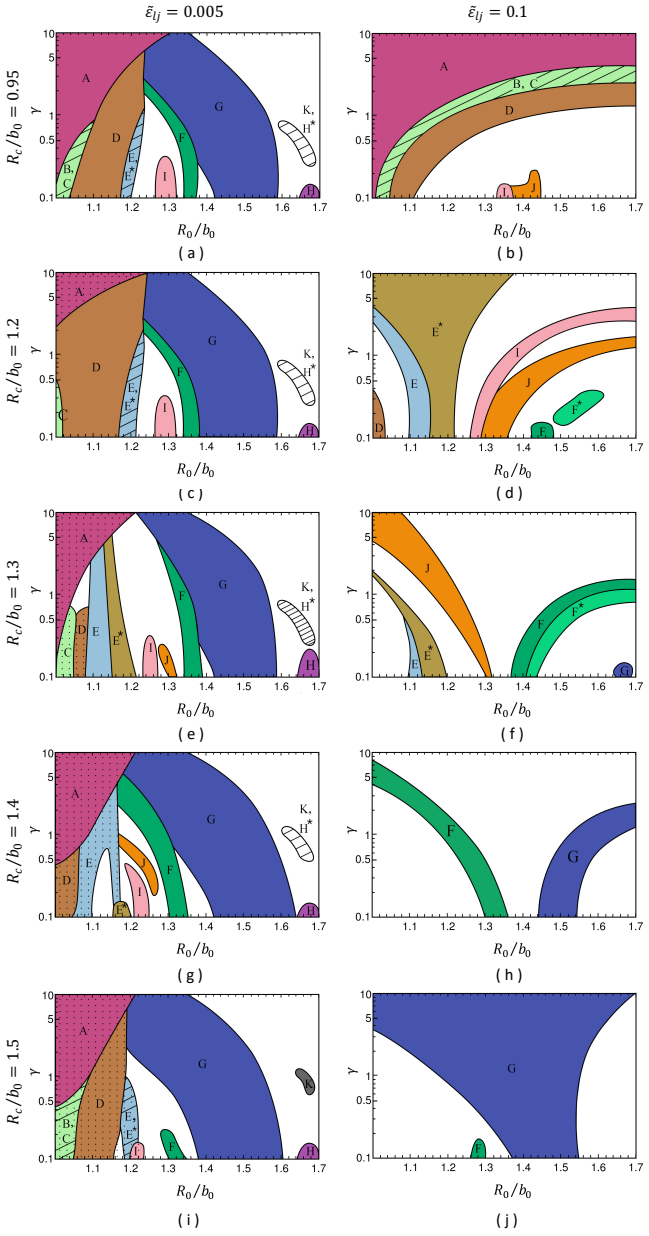

Figure 3.5: Phase diagrams of the structures obtained in the deterministic simulations in the presence of genome with five various sizes and two different strengths of the protein-genome attractive interactions. From the top to the bottom row, the core sizes are $R_c/b_0 = 0.95, 1.2, 1.3, 1.4, 1.5$. We note that the core sizes $R_c/b_0 = 0.95, 1.2, 1.4, 1.5$ are the optimal sizes to grow $T = 1$, E and E^*, F and $T = 3$ structures, respectively. The core size $R_c/b_0 = 1.3$ does not correspond to any of the symmetric structures. The left column corresponds to the weak core-protein interaction $\tilde{\epsilon}_{lj} = 0.005$ and the right column to the strong one $\tilde{\epsilon}_{lj} = 0.1$. Each color corresponds to a structure with specific symmetry (see Fig. 3.3). Dotted regions correspond to the phases in which the capsids form without packaging the core. A careful review of the phase diagrams reveals that the icosahedral structures prevail even if both the spontaneous curvature and core size are optimal for the formation of another type of symmetric structure.

even when the size of cargo is optimal for the formation of $T = 1$ icosahedral structure, depending on the spontaneous curvature of proteins and their mechanical properties, the structure D prevails a large region of phase space (Fig. 3.5b). This might explain to some extent why minicoats form in the case of clathrin shells, rather than an icosahedral structure.

As noted above, the white regions of the phase diagrams in Figs. 2.8 and 3.5 corresponds to the structures with no specific symmetry. In the next section, in which we present the equilibrium structures, we will show that the irregular shells in the white regions do never constitute the minimum energy structures. They basically correspond to the kinetically trapped capsids. In fact, all the small shaded regions in the deterministic phase diagrams surrounded by the white region will expand in the equilibrium phase diagrams (see Fig. 3.7) and the white regions will completely disappear.

Another interesting point of the phase diagrams is that the genome can move the transition point between two icosahedral structures $T = 1$ and $T = 3$ even when the genome-protein interaction strength is weak ($\tilde{\epsilon}_{lj} = 0.005$). For example, at $\gamma = 10$ when there is no core, the transition point from $T = 1$ to $T = 3$ structure occurs at $R_0/b_0 = 1.31$ (Fig. 2.8). However, even when the attraction is weak, for $R_c/b_0 = 1.5$, the empty $T = 1$ forms at very small R_0/b_0 without packaging the core since packaging of a big core will cost a lot of elastic energy. Nevertheless, as can be seen in Fig. 3.6 increasing R_0/b_0 to 1.24 (Fig. 3.5i) results into the formation of $T = 3$ with the genome enclosed.

In general, in the presence of weak LJ, the phase boundaries remain more or less vertical showing that the spontaneous curvature of proteins has the key role in defining the structure of capsids. However, the boundary phases become more horizontal as we increase

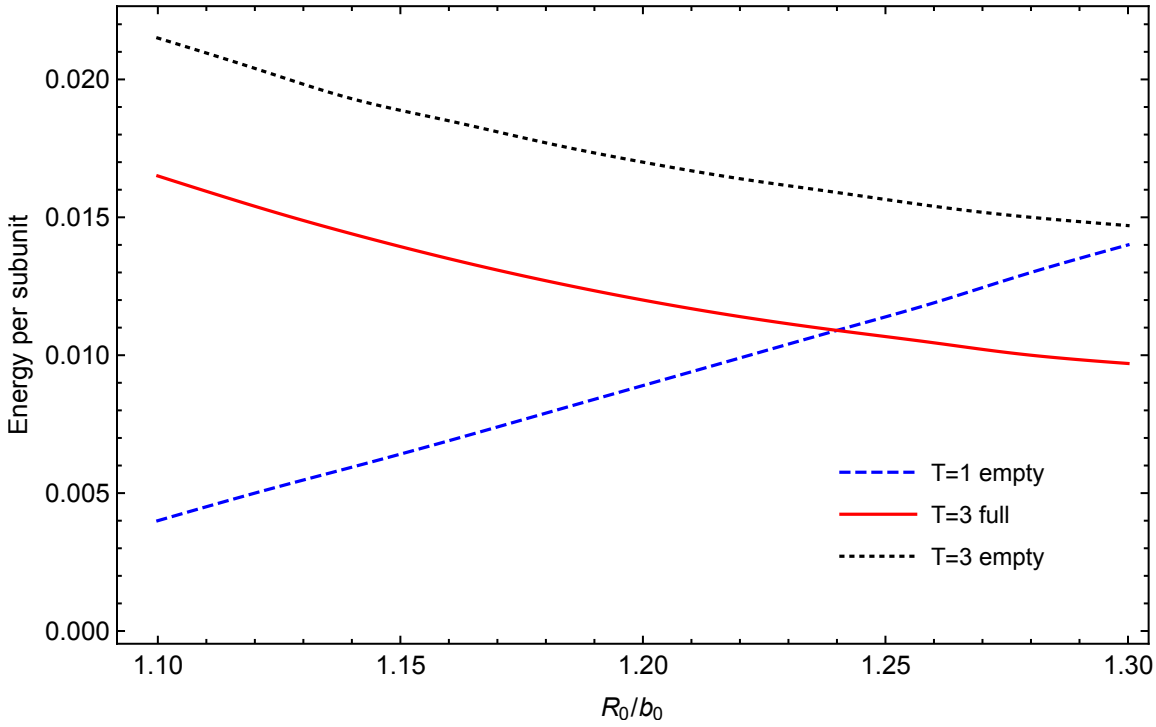

Figure 3.6: At $\gamma = 10$ when there is no core, the transition point from $T = 1$ to $T = 3$ is at $R_0/b_0 = 1.31$. However, in the presence of a core whose size is $R_c/b_0 = 1.5$ (perfect for a $T = 3$ shell), the boundary shifts towards smaller $R_0/b_0 = 1.24$. When the core-protein attraction is weak ($\tilde{\epsilon}_{lj} = 0.005$), $T = 1$ forms for small values of $R_0/b_0 < 1.24$ without packaging the core as it will cost a lot of elastic energy. However, increasing R_0/b_0 to 1.24, $T = 3$ structures form with the genome inside.

LJ strength, indicating that the core determines the radius of capsid while FvK number modifies the shell symmetry.

3.3.1 Equilibrium structures

As noted previously, the deterministic simulations involve some irreversible moves, and as such the structures obtained in Fig. 2.8 and Fig. 3.5 do not necessarily constitute the free energy minimum structures. To obtain the equilibrium size and symmetry of empty capsids as a function of γ and R_0/b_0, in chapter 2, we carried out Monte Carlo (MC) simulations combined with the bond-flipping method [77]. Unexpectedly, the equilibrium structure phase diagram and the deterministic one presented in Fig. 2.8 were very similar. It was found that the striking similarities are the result of the strong affinity for the formation of pentamers at specific positions in the shell, which depends on the spontaneous radius R_0/b_0 and the mechanical properties (γ) of the capsid proteins [60, 77].

Since the previous works have already confirmed that similar structures appear in the deterministic and equilibrium phase diagrams albeit small differences in the boundary between different regions, to obtain the minimum energy structures in the presence of genome, we compared the total energy per subunit of all the structures formed in Fig. 3.5 including the irregular ones for various core sizes as a function of R_0/b_0 and γ.

To construct the equilibrium phase diagrams, we assumed that the genome is located in the center of mass of the capsid and is interacting with the triangular subunits through the LJ potential. The elastic energy of the capsid as before is the sum of the stretching E_s and bending E_b energies as presented in Eqs. 2.1 and 2.2. Keeping the genome at the center of mass, we let different capsids relax and then we calculate the total

energy including the core-capsid, the stretching and bending energies per subunit to find the minimum energy structures. It is worth mentioning that instead of constructing the phase diagrams based on the minimum energy calculations, we could have performed simulations in the grand canonical ensemble framework and monitor the growth of the shells [76]. That approach would be useful if we were interested in the impact of the protein concentration, temperature and the strength of hydrophobic interaction between the proteins too (subject of chapter 4). We note that all these parameters can modify the final assembly product, see Ref. [79]. However, this would have added many more degrees of freedom to the problem and make the interpretation of the results less transparent. To this end, we assumed that the protein concentration and the strength of the hydrophobic interactions are such that only the competition between the elastic energies, the core size and the strength of core-protein interactions define the final structure of capsids. Thus we constructed the equilibrium phase diagrams based on the strength of LJ interactions and elastic energies.

Figure 3.7 shows the equilibrium phase diagrams for the same core sizes (R_c/b_0) as the ones used in Fig. 3.5, from top to bottom $R_c/b_0 = 0.95, 1.2, 1.3, 1.4, 1.5$. Similar to Fig. 3.5, the left and right columns correspond to the weak ($\tilde{\epsilon}_{lj} = 0.005$) and strong ($\tilde{\epsilon}_{lj} = 0.1$) LJ interactions, respectively. In the dotted regions, only empty shells can form. Comparing the equilibrium phase diagrams with the non-equilibrium ones, we find that several structures assembled in the deterministic simulations do not represent the equilibrium structures, such as B, C and E. A quite interesting difference is that the structure F^*, discovered in the experiments of Ref. [5], appears in wide regions of the phase diagrams in Fig. 3.7. However, the structure E that was also obtained in the recent

experiments, only appears in the deterministic simulations performed under non-equilibrium conditions. Further, all the regions related to irregular shapes in the non-equilibrium phase diagram disappear in the equilibrium ones, most of them are replaced by the structure J with a two fold symmetry (see 3.3). The icosahedral structure $T = 3$ occupies a larger area in the equilibrium phase diagrams, in agreement with the observation that many small ssRNA viruses form $T = 3$ structures.

To verify the growth pathway suggested in Ref. [5] and described in the introduction, we closely study the assembly of $T = 3$ and F^* in the presence of different core sizes. Figures 3.7h and j show that at high γs, for the core size $R_c/b_0 = 1.4$, F^* is the equilibrium structure; however, increasing R_c/b_0 to 1.5, $T = 3$ becomes the minimum energy structure in the same region of the phase diagram (for the same γ and R_0/b_0). Figure 3.9a shows the growth pathways as the plots of the energy per subunit, for F^* and $T = 3$ structures, as a function of number of subunits at $R_c/b_0 = 1.4$ and $R_c/b_0 = 1.5$, respectively. The other parameters in the figure are the spontaneous radius of curvature $R_0/b_0 = 1.4$, $\gamma = 5$ and $\tilde{\epsilon}_{ij} = 0.1$. The figure clearly shows that up to 25 subunits the energies of the two structures are very close to each other. Note that the total number of subunits in F^* structure is 50 subunits, and thus 25 is when the half of the capsid is formed. From this point to the end of the growth though, the energy of F^* increases sharply in comparison to $T = 3$.

We emphasize that as shown in Figs. 3.7h and j, the optimal spontaneous radius of curvature for the formation of F^* structure is $R_0/b_0 = 1.4$, explaining why if the core size is also $R_c/b_0 = 1.4$, F^* is the minimum energy structure. However, if we set the radius of the core to $R_c/b_0 = 1.5$ while keeping the preferred spontaneous curvature of proteins the same

as before ($R_0/b_0 = 1.4$), then we obtain $T = 3$ structures even though R_0/b_0 is the optimal size for the formation of F^*. The phase diagrams in Figs. 3.7h and j reveal the regions in which an increase in the core size modifies the optimal structure from F^* to $T = 3$. Quite interestingly, this indicates that if two different genome sizes are in solution, the capsid proteins energetically prefer to assemble around the larger one to form $T = 3$ structure, and this happens, to some extent, regardless of the preferred curvature of proteins.

Figure 3.9b shows the stress distribution of some incomplete structures corresponding to the growth of F^* and $T = 3$ shells in Fig. 3.9a. See Method section for the details of the stress calculation [61]. The bar legend on the right side of Fig. 3.9b shows the color code corresponding to the different stress level. While the violet color corresponds to the maximum stretching tension, the red shows the maximum compression stress in the bonds. The first incomplete structure on the left of the figure with $n_t = 25$ corresponds to both $T = 3$ and F^*. Up to that point, the growth pathways of both structures overlap as is illustrated in Fig. 3.9a too. From $n_t = 25$ onward, the stress distribution in the structures packaging $R_c/b_0 = 1.4$ and $R_c/b_0 = 1.5$ will become different although their growth pathways are still the same *i.e.*, their incomplete structures are almost identical regardless of the stress distribution. Finally, at $n_t = 30$, the incomplete structures forming around $R_c/b_0 = 1.4$ and $R_c/b_0 = 1.5$ start to deviate and follow two different pathways. The upper row shows the pathway corresponding to the formation of the F^* structure with 50 subunits. The figure shows that the highest compression stress distribution is along the pairs of pentamers, where the axis of two fold symmetry are located. However, the bottom row corresponds to $T = 3$ with 60 subunits. The level of stress in the $T = 3$ structure is much lower than

the F^* structure, confirming that F^* can easily break into large fragments to provide the intermediate structures for the formation of $T = 3$ with lower stress when cognate larger RNA becomes available.

Another puzzling question to address corresponds to as why the structure F^* appears only in a small region in the deterministic simulations but covers large regions in the equilibrium phase diagrams. For example, at $R_0/b_0 = 1.1$, $\gamma = 3$, $R_c/b_0 = 1.4$ and the strong LJ potential ($\tilde{\epsilon}_{lj} = 0.1$), F^* is the minimum energy structure, see Fig. 3.7h. However, in this region, we obtain the structure F in our deterministic simulations as shown in Fig. 3.5h. A careful review of the step by step assembly of both structures in Fig. 3.8 shows that both structures F and F^* are built from 50 subunits, and both follow exactly the same pathway till 25 subunits are assembled (the half of the capsid).

After the assembly of the first 25 subunits (Fig. 3.10a), F and F^* follow two different pathways. For the structure F, the five subunits sitting at the edge of the growing shell merge and form a pentamer (Fig. 3.10b). However, in the case of F^*, one more subunit is added and the capsomer will close as a hexamer (Fig. 3.10c). Even though a complete F^* structure has a lower energy than the F one, the assembly of a hexamer is energetically more costly than a pentamer after the half of the capsids is assembled. To this end, there is an energy barrier to the formation of F^* and thus in the deterministic simulations, the shells follow the pathway resulting into the formation of F structures.

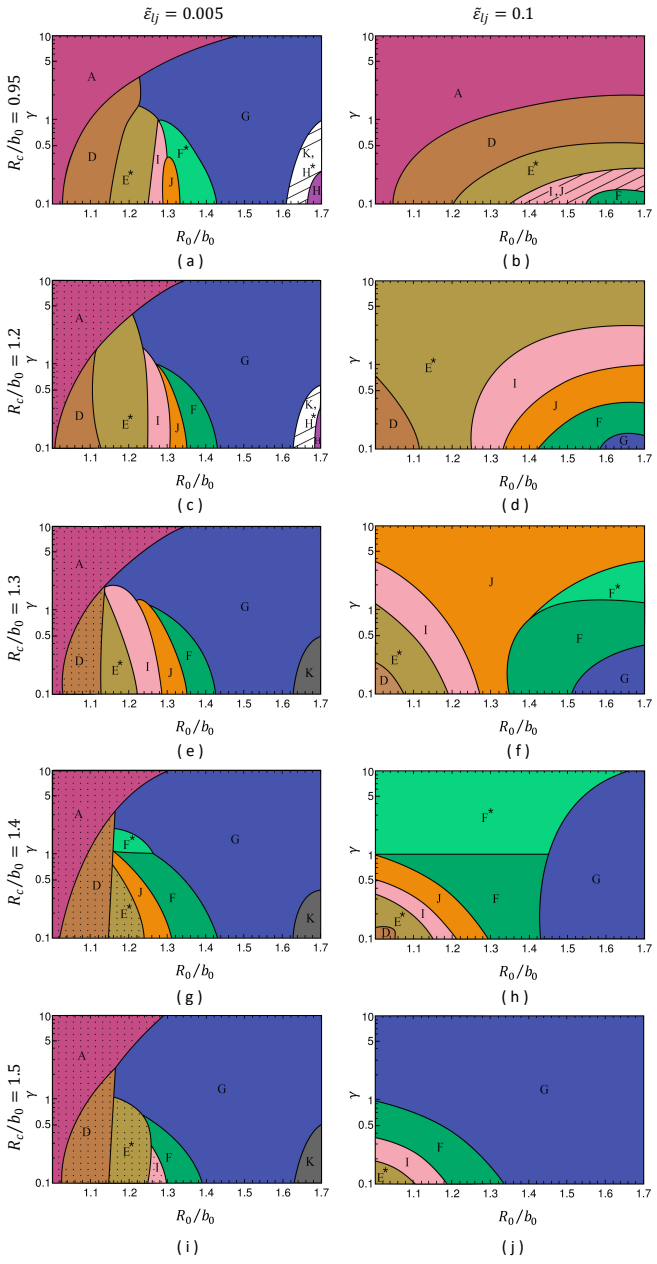

Figure 3.7: Phase diagrams of the equilibrium structures obtained in the presence of genome. The plots have the same color code as in Fig. 3.5. The dotted colors correspond to the regions in which the empty shells are the equilibrium structures. From the top to the bottom row, the core sizes are $R_c/b_0 = 0.95, 1.2, 1.3, 1.4, 1.5$. We note that the core sizes $R_c/b_0 = 0.95, 1.2, 1.4, 1.5$ are the optimal sizes to grow $T = 1$, E and E^*, F and F^*, and $T = 3$ structures, respectively. The core size $R_c/b_0 = 1.3$ does not correspond to any of the symmetric structures. The left column corresponds to the weak core-protein interactions ($\tilde{\epsilon}_{lj} = 0.005$) and the right column to the strong one ($\tilde{\epsilon}_{lj} = 0.1$). In the hashed regions, two equilibrium structures form. Each color corresponds to a structure with specific symmetry (see Fig. 3.3). Similar to the phase diagrams presented in Fig. 5, the icosahedral structures prevail even if both the spontaneous curvature and core sizes are optimal for the assembly of other symmetric structures.

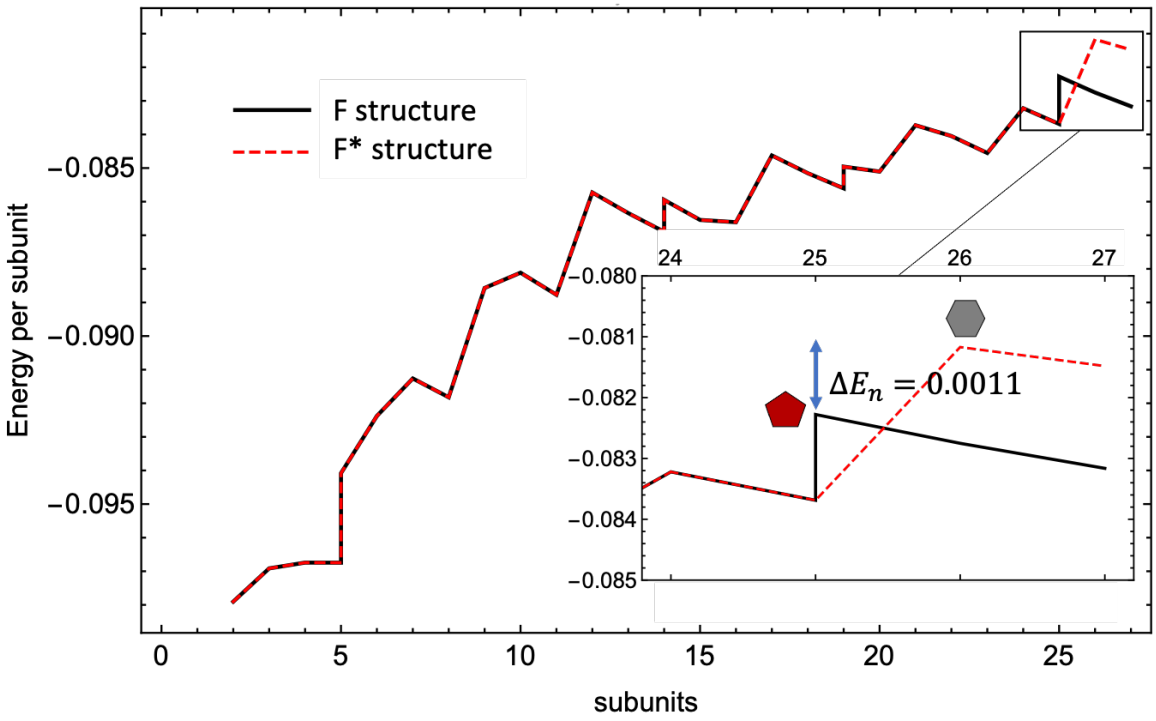

Figure 3.8: Plots of the energy per subunit for the structures F (the black solid line) obtained from the deterministic simulations and F^* (the red dashed line) as a function of the number of subunits at $R_0/b_0 = 1.1$, $\gamma = 3$, $R_c/b_0 = 1.4$ and the strong LJ potential $\tilde{\epsilon}_{ij} = 0.1$. Both structures F and F^* follow the same pathway until one half of the capsids (25 subunits) are assembled. Then, the two structures follow two different pathways. For the F structure, the 5 subunits at the edge merge and form a pentamer. However, for the F^* structure, one more subunit will be added and the capsomer will close as a hexamer with higher energy (see Fig. 3.10). The energetic cost of the formation of a hexamer is higher at this stage, and as such, the capsid follows the black line pathway and forms the F structure while the energy of complete F^* structure is lower.

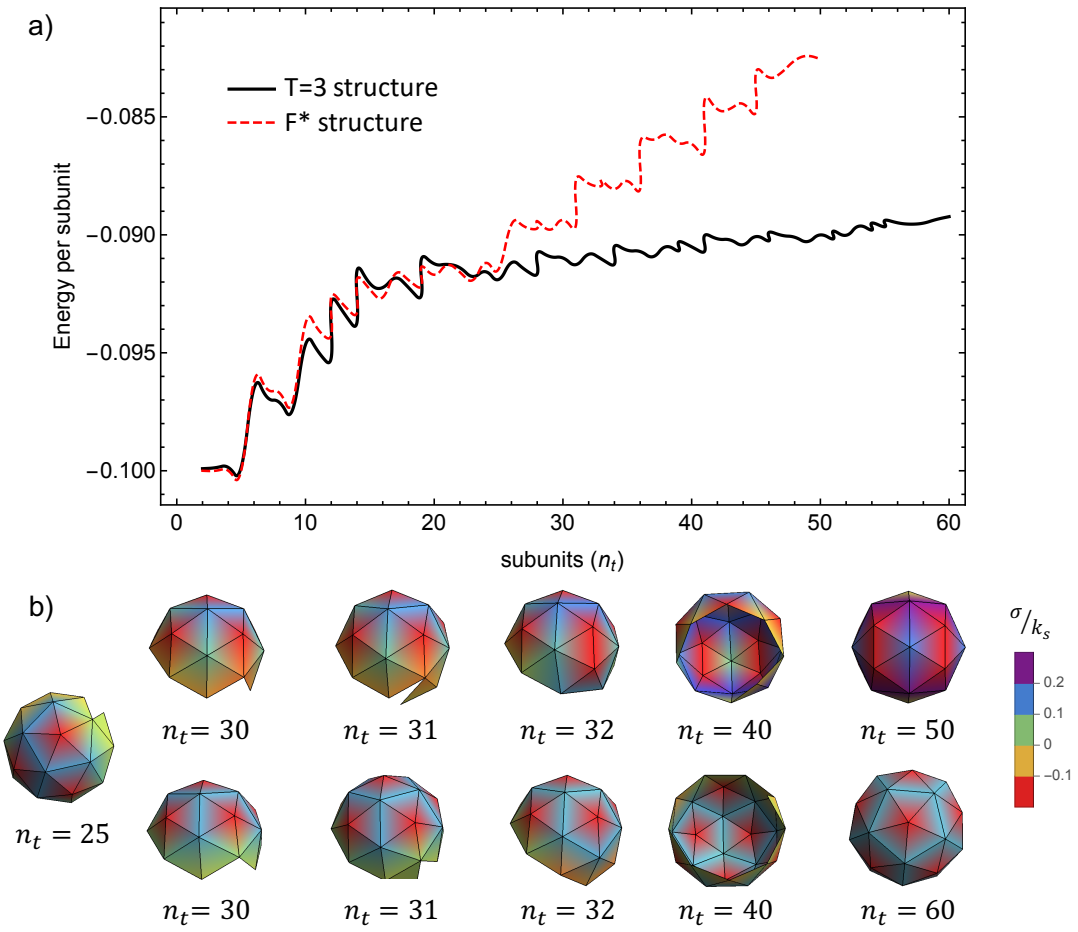

Figure 3.9: a) Plots of the energy per subunit *vs.* number of subunits for the structures F^* (the red dashed line) and $T = 3$ (the black solid line) at $\gamma = 5$ and the strong LJ potential $\tilde{\epsilon}_{ij} = 0.1$. The spontaneous radius of curvature of proteins in both structures is $R_0/b_0 = 1.4$, which is the preferred curvature for the formation of F^*. If the size of the core is also $R_c/b_0 = 1.4$, F^* constitutes the minimum energy structure, as expected. However, if the size of the core is $R_c/b_0 = 1.5$, $T = 3$ becomes the equilibrium structure despite the fact that the spontaneous radius of proteins is $R_0/b_0 = 1.4$, which commensurates with an F^* structure. b) Map of the stress distribution in several incomplete shells growing within the pathways shown in a). The top and bottom rows show some intermediate structures of the F^* and $T = 3$ shells, respectively. Up to 30 subunits, the intermediate structures of both F^* and $T = 3$ look exactly the same. However, their energies become different after 25 subunits are assembled. The color bar shows the different levels of stress. Violet color corresponds to the maximum of stretching stress and the red one to the maximum of compression. The F^* structure is under much more stress compared to $T = 3$.

3.4 Summary

Very little is known as to why inside a cell, capsid proteins predominantly package their cognate RNA. Many factors contribute to this process, but our knowledge in this area is still rudimentary. In this chapter, we have investigated the impact of genome on the structure and excess elastic stress of viral shells, which are tightly connected to the capsid stability and function. In particular we have focused on the recent self-assembly experiments of BMV capsid proteins around short oligonucleotides, using a combination of charge detection mass spectrometry and cryo-EM techniques [5]. Quite interestingly, these experiments have revealed that the capsid proteins can assemble to form symmetric structures other than icosahedral ones [5]. Some of the structures formed around the small pieces of oligonucleotides have the D_{6h} hexagonal barrel symmetry (the E structure in Fig. 3.1A and Fig. 3.3) and some others have D_{5h} symmetry (the F^* structure in Fig. 3.1C and Fig. 3.4).

To explore under what conditions the F^* and E structures could form, we have constructed several phase diagrams for various genome sizes as a function of spontaneous curvature and mechanical properties of protein subunits, see Figs. 3.5 for the results of the deterministic simulations and 3.7 for the equilibrium structures. We found that the structure F^*, which was called $H15$ in Ref. [5], can only form in the presence of genome.

Our results also show that the size of genome plays an important role for the assembly of F^*, and that the optimal core size for the formation of F^* is smaller than that for $T = 3$. As illustrated in the plots of Fig. 3.7h, the shell F^* constitutes the global minimum energy structure at the smaller core size ($R_c/b_0 = 1.4$) for a large region of the

phase diagram. Figure 3.9 reveals that the formation of F^* and $T = 3$ follow almost the same pathway until 30 triangular subunits are assembled. However, their energies and stress distributions start differing after the assembly of the first 25 subunits. Figure 3.9a shows that after the assembly of the first 30 triangular subunits, while F^* is an equilibrium structure for $R_c/b_0 = 1.4$, if a larger RNA becomes available, the shell can easily switch the pathway and form a $T = 3$ structure because of the difference between the energy of a $T = 3$ and F^* for larger RNAs.

We also calculated the stress distribution across F^* and $T = 3$ shells as they grow (see Fig. 3.9b), which confirms our conclusion based on the energy of the systems. As can be seen in the figure, the stress between every two adjacent pentamers in F^* structure is very high despite the fact that it is an equilibrium structure, Fig. 3.7h. The implication is that F^*s are stabilized by encapsidating short RNAs, but they can become metastable and split into large fragments, which later grow to form $T = 3$ structure with much lower stress distribution if larger cognate RNAs are accessible. Thus, our work suggest an alternative pathway for the assembly of $T = 3$ shells inside the cell in the presence of many smaller non-viral RNAs.

Another interesting point of the *in vitro* experiments of Bond *et al.* [5] corresponds to the co-existence of E and F^* structures. Our work clarifies the conditions under which the two different structures can coexist. Since the capsids assembled in the experiments are built from the same types of proteins, we explored if it is possible to obtain both structures (E and F^*) for the same γ and R_0/b_0 (related to the mechanical properties of capsid proteins), albeit different core sizes. A careful review of Figs. 3.5 and 3.7 shows

that we can obtain several structures simultaneously in various cases if different genome lengths are packaged inside the capsids. These results are consistent with the findings of the experiments of Ref. [5], which show that the number of packaged short oligonucleotides are not the same in each structure. Thus if the number of packaged oligomers are different, it is possible to obtain the structures E and F^* simultaneously in solution.

It is important to note that the E structure can only be obtained in the deterministic simulations; it never appears in the equilibrium phase diagrams, and as such it constitutes a minimum energy structure neither for the filled nor for the empty capsids. The Es are probably trapped in a local minimum free energy and are less stable than other structures in the experiments, see also the stress distribution across the E structure in Fig. 3.11. These structures might form in the cell but can easily disassemble and reassemble to form a stable icosahedral structure. It is worth mentioning that our simulations were performed under the assumption that the genome assumes an isotropic shape inside the capsid. However, the structure E is slightly elongated and an anisotropic genome distribution inside the capsid can render E a minimum energy structure. Thus, more theoretical and simulations are needed to study the stability of E structures.

In summary, our work shows that the interplay of protein mechanical properties, genome volume, and strength of the genome-capsid protein interaction can significantly modify the symmetry, structure, and stability of the assembly products, and hence kinetic pathways of assembly. Further, our work also supports the presence of an alternative pathway for the formation of $T = 3$ icosahedral structures around their cognate RNAs in the presence of oligonucleotides. More experiments to confirm the existence of this pathway

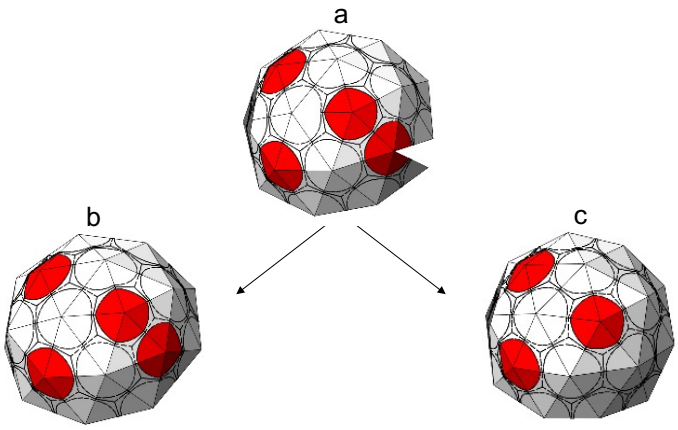

Figure 3.10: Both structures F and F^* follow the same pathway until the half of the capsids (25 subunits) is assembled (a). Then, the two structures follow two different pathways. For the F structure, the 5 subunits at the edge merge and form a pentamer (b). However, for the F^* structure, one more subunit will be added and the capsomer will close as a hexamer (c). The energetic cost of the formation of a hexamer is higher at this stage, and as such, the capsid will follow the F structure pathway while the energy of a complete F^* structure is lower. Red and white colors correspond to the location of pentamers and hexamers respectively. See also Fig. 3.8 for the plot of energy per subunit *vs.* the number of subunits.

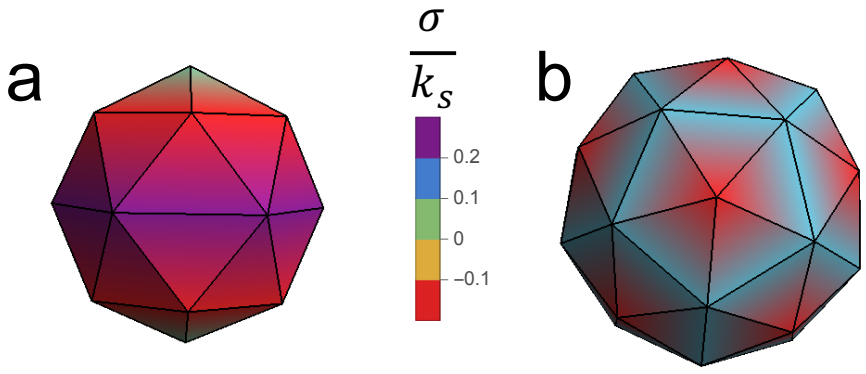

Figure 3.11: Map of the stress distribution of the structures E and $T = 3$ (a) Shell with E symmetry forms in the deterministic simulations for $R_c/b_0 = 1.2$, $\tilde{\epsilon}_{lj} = 0.1$, $R_0/b_0 = 1.04$ and $\gamma = 3$. The highest positive stress is along the hexamer belt in the middle of the shell. (b) When we increase the size of the core to $R_c/b_0 = 1.5$ keeping all the other parameters as in part (a), the structure $T = 3$ forms which has much lower stress compared to the structure E.

would be required, at different salt concentrations, and genome lengths. Such studies are needed in order to understand ssRNA virus assembly in presence of polyanionic, non-cognate potential encapsulation competitors.

Chapter 4

How a Virus Circumvents Energy Barriers to Form Symmetric Shells

4.1 Introduction

The process of formation of virus particles in which the protein subunits encapsidate genome (RNA or DNA) to form a stable, protective shell called the capsid, is an essential step in the viral life cycle [30]. The capsid proteins of many small single-stranded (ss)RNA viruses spontaneously package their wild type (wt) and other negatively charged polyelectrolytes, a process basically driven by the electrostatic interaction between positively charged protein subunits and negatively charged cargo[91, 86, 26, 27]. Understanding the phenomena of formation of viral particles is of great interest due to their potential applications in nanomedicine and biomaterials. Capsids can be employed as nanocontainers, biosensors and drug or gene delivery to name just a few [98, 109, 81, 27, 71].

Regardless of the virion size and assembly procedures, most spherical viruses adopt structures with icosahedral symmetry [100, 48]. The total number of proteins in an icosahedral capsid is equal to $60T$ where T is the triangulation number assuming only certain integers (T=1,3,4,7,...).[10] Some number of viruses follow a generalized principle of quasi-equivalence, that can be modelled *via* a combination of hexamers, pentamers and trimers[96]. We do not discuss this class of viruses as our focus is basically on $T = 3$ structures in this chapter. Independently of the number of proteins, there are always 12 pentagonal defects in a shell built from a triangular lattice sitting on the vertices of an icosahedron to preserve the symmetry of the capsid.

How exactly capsid proteins (CPs) assemble to assume a specific size and symmetry have been investigated for over half a century now [3, 11]. Since the self-assembly of virus particles involves a wide range of thermodynamics parameters, different time scales and an extraordinary number of possible pathways, the kinetics of assembly has been remained elusive, linked to Levinthal's paradox for protein folding[1, 58, 21]. The role of the genome on the assembly pathways and structure of capsid is even more intriguing [67, 87, 86, 94, 25]. The kinetics of virus growth in the presence of RNA is at least 3 orders of magnitude faster than that of empty capsid assembly, indicating that the mechanism of assembly of CPs around RNA might be quite different. Some questions then naturally arise: What is the role of RNA in the assembly process and by what means then, does RNA preserve assembly accuracy at fast assembly speed?

Two different mechanisms for the role of genome have been proposed: (i) en masse assembly, and (ii) nucleation and growth [54, 40, 80]. Several years ago, McPherson sug-

gested the "en masse" model in which the nucleic acid attracts CPs in solution to its surface through long-range electrostatic interactions. Note that the assembly interfaces in many CPs are principally short-ranged hydrophobic in character, while there is a strong electrostatic, nonspecific long-ranged interaction between RNA and CPs. To this end, the positively charged domains of CPs associate with the negatively charged RNA quite fast and form an amorphous complex. Hydrophobic interfaces then start to associate, which leads to the assembly of a perfect icosahedral shell. Based on the en masse mechanism, the assembly pathways correspond to situations in which intermediates are predominantly disordered.

More recently Chevreuil et al. [14] studied en masse assembly by carrying out time-resolved small-angle X-ray scattering experiments on cowpea chlorotic mottle virus (CCMV), a $T = 3$ single-stranded RNA plant virus. They found that at neutral pH, a considerable number of CPs were rapidly ($\sim$28 ms) adsorbed to the genome, which more slowly ($\sim$48 s) self-organized into compact but amorphous nucleoprotein complexes (NPC) (see also [38]). By lowering the pH, they observed a disorder-order transition as the protein-protein interaction became strong enough to close up the capsid and to overcome the high energy barrier ($\sim 20 k_B T$ with k_B the Boltzmann constant and T the temperature) separating NPCs from virions. It is important to note that a marked difference between *in vitro* and *in vivo* assemblies is that capsid proteins form ordered icosahedral structure at physiological pH (*in vivo* conditions) but till now ordered structures are observed only at acidic pH in the *in vitro* self-assembly studies [36, 8, 18].

While because of the electrostatic interaction between the genome and CPs, the en masse assembly is expected to be the dominant growth pathway, a set of more recent experiments, however, point to the nucleation-elongation mechanism. Garmann et al. have employed interferometric scattering microscopy to measure the assembly kinetics of single MS2 virus particles around MS2 RNA strands tethered to the surface of a coverslip [39, 18]. Comparisons of individual assembly pathways indicated that most trajectories exhibit a sigmoidal time-dependence with plateau scattering intensity values similar to those acquired from complete particles. According to their experimental data, each trajectory is characterized by a lag time after which the assembly takes off, a signature of the nucleation-elongation mechanism. The measurements of Garmann et al. reveal that the lag time is not due to the diffusion of the CPs but it is directly related to the nucleation and growth mechanism [39].

While the experiments of Chevreuil et al. point to the en masse assembly mechanism, it is hard to explain how a transition from a disordered amorphous RNA-CPs complex to a highly ordered icosahedral capsid can take place where the location of pentamers is very precise. Further, the assembly through a nucleation and growth mechanism is not bound to follow an ordered pathway at all times *i.e.*, a disorder-order transition might also occur as the shell grows. The molecular dynamic simulations of Elrad and Hagan have shed light on the assembly pathways of virus particles in the presence of the genome [40, 22]; however, the simulations were designed such that only $T = 1$ structures could form while the kinetic pathways to the formation of slightly larger viruses such as $T = 3, 4$ are still elusive.

In this section, we study the kinetic pathways of virus assembly using MC simulations and specifically elucidate the role of elastic energy in the disorder-order transition in $T = 3$ viruses formed through either en masse or nucleation and growth mechanism. We in particular focus on the impact of protein concentration on the disorder-order transition. Through MC simulations and a set of experiments employing small-angle X-ray scattering and cryo-transmission electron microscopy, we explore the role of protein concentrations in the formation of $T = 3$ structures of CCMV particles. Our findings shed light on the *in vivo* assembly of icosahedral structures and can explain to some extent why virion forms at neutral pH *in vivo* while the acidic pH was so far required in the *in vitro* self-assembly studies of $T = 3$ structures.

This chapter is organized as follows. After introducing the simulation approaches in the Method section we present the results of numerical simulations indicating the importance of elastic energy in the disorder-order transition. We also study the impact of protein concentration on the assembly pathway and compare it with our SAX experiments. Several details of calculations and experimental methods are relegated to the Appendix.

4.2 Method

Similar to chapter 3, to study the growth of viral shells, we use a coarse-grained model in which we focus on the fact that viral capsids are formed basically from hexamers and pentamers. In this chapter, the total energy of a growing shell is the sum of the

attractive interaction between subunits E_{hp}, and the stretching E_s and bending E_b energies as follows,

$$E_s = \sum_{l_i} \frac{1}{2} k_s (b_{l_i} - b_0)^2 \tag{4.1}$$

$$E_b = \sum_{<t_i,t_j>} k_b [1 - \cos(\theta_{<t_i,t_j>} - \theta_0)] \tag{4.2}$$

$$E_{hp} = \sum_{v_i} \epsilon_{hp} [nt_{v_i} \cdot (nt_{v_i} - 1)], \tag{4.3}$$

where the stretching energy sums over all edges l_i with b_0 the equilibrium length set equal to one. Thus b_{l_i} is a dimensionless quantity and k_s the stretching modulus has the unit of energy. The bending energy involves the dihedral angle $\theta_{<t_i,t_j>}$ between two connected trimers indexed by t_i and t_j such that $cos\theta_{<t_i,t_j>} = \hat{n}_{t_i}.\hat{n}_{t_j}$ with $\hat{n}_{t_i}$ and $\hat{n}_{t_j}$ the unit normal vectors of the trimers. The bending rigidity is k_b and $\theta_0 = 2\arcsin(\frac{b_0}{\sqrt{12R_0^2 - 3b_0^2}})$ is the preferred dihedral angle related to the spontaneous radius of curvature R_0 [45]. The attractive energy due to the hydrophobic interaction sums over all vertices v_i with ϵ_{hp} the strength of monomer-monomer interaction. The range of hydrophobic interaction is considered to be small so that each vertex only interacts with its nearest neighbor vertices. The number of connected trimers is denoted by nt_{v_i} for any vertex v_i. We note that the interaction between subunits is due to both the hydrophobic interactions and electrostatic repulsion between subunits [49]. Since the resultant force is attractive for the viruses to assemble, for brevity, we call it, the hydrophobic interaction, in the rest of this chapter.

We emphasize that the hydrophobic interaction is the reason why proteins aggregate and attract each other and thus it is negative. However, the stretching energy is due to the deviation of subunits from their equilibrium shape and is positive. We note that

other parameters like the geometry of the building blocks could also have an impact on the growth pathway but our focus in this study is mainly on the impact of elastic energy.

Using Monte Carlo (MC) simulations, we monitor the growth of a viral shell through reversible steps [85]. We consider three different MC moves: 1. Diffusion, 2. Attachment or detachment and 3. Merging or disjointing.

During the diffusion process, the number of capsid subunits remains fixed, corresponding to a canonical system. At each MC step, either a trimer or a vertex is chosen randomly, see Figs. 4.1 and 4.2. The move is accepted with the probability $p_m = \min(1, e^{-\beta E})$ with $\beta = 1/k_B T$ and E the change in the total energy of the system after relaxation. The diffusion of trimers in Fig. 4.1 could be considered as a combination of detachment and growth for trimers, the result of which should be qualitatively the same as diffusion in the canonical ensemble. The diffusion stops based on the convergence of the energy of the system. The arrows in Fig. 4.2 indicate the reversibility of each action. After the diffusion step, a subunit will be added or removed with the probability p_g and p_d, respectively. The probability of growth follows the detailed balance of the grand canonical ensemble with $p_g = \min(1, \frac{T_g}{T_d} e^{\beta\mu - \beta\Delta E})$ and the probability of removal is $p_d = \min(1, \frac{T_d}{T_g} e^{-\beta\mu - \beta\Delta E})$[35]. The possible growth and detachment sites are denoted by T_g and T_d respectively and μ is connected to the chemical potential of free proteins in solution. For a dilute system, $\mu = k_B T \log \frac{\rho}{\rho_0}$ with ρ the density of free proteins in solution and ρ_0 the reference density.

We note that in most experiments monitoring the assembly of proteins around the genome till now, RNA condensates into its final size much faster than the other time scales involved in the process [6, 14, 18]. Thus we replace the genome with a spherical

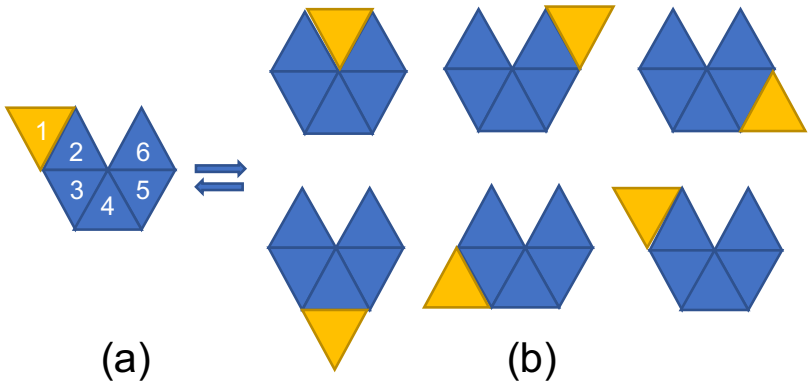

Figure 4.1: Schematic presentation of diffusion of a trimer around the shell edge. (a) Trimer 1 (yellow) is chosen randomly to diffuse around the edge. In this case, only trimers 1 and 6 are allowed to move during the diffusion mode. Trimers 2-5 are connected such that a few bonds need to be broken before they can diffuse. (b) Possible locations for trimer 1 to diffuse.

core as the goal of this study is basically on how the shell overcomes many energy barriers to form symmetric structures and not on the genome configurations. We note that if the experiments were performed in the diffusion limited regime, we could not simply replace RNA with a spherical core.

Replacing genome by a spherical core will decrease many degrees of freedoms in the system allowing us to focus carefully on the energetics involved in the assembly of viral shells. Thus, we consider a spherical core interacting through Lennard-Jones (LJ) with the capsid proteins $E_{lj} = \sum_{vi} \epsilon_{lj}[(\frac{\sigma_{lj}}{r_{<vi,g>}})^{12} - 2(\frac{\sigma_{lj}}{r_{<vi,g>}})^{6}]$ where ϵ_{lj} is the potential strength and $r_{<vi,g>}$ is the distance between the core center and the triangle vertex v_i.

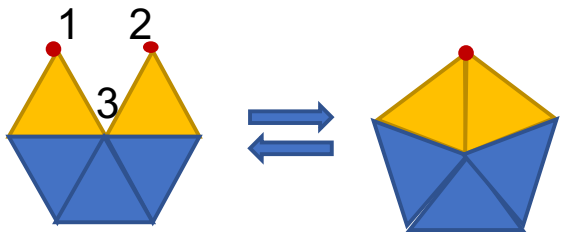

Figure 4.2: Schematic of vertex diffusion: Vertices 1 and 2 merge to form a pentamer. The action is reversible in that both merging (left to right arrow) and disjointing (right to left arrow) are possible in the Monte-Carlo moves.

4.3 Results and Discussion

As noted in the introduction two different mechanisms are proposed for the role of genome in the assembly of viral shells: (1) Nucleation and growth, and (2) en masse assembly. We will decipher each in the following sections. Because of the extraordinary number of degrees of freedom involved in the assembly process, we focus only on understanding a number of recent experiments in which genome has assumed different roles. We study the experiments of Garmann et al. [39] to elucidate the process of nucleation and growth in the presence of the genome and then investigate the experiments of Chevreuil et al. to explore how the genome promotes the assembly through en masse or multi-nucleation mechanism [14]. In order to obtain a better understanding of the disorder-order transition, we perform a set of x-ray scattering experiments at neutral pH to study the role of protein concentration in the disorder-order transition and compare them with the results of our simulations. In what follows we first present the results of our simulations for the single nucleation assembly and then show our experimental data and simulations corresponding to the multi-nucleation or en masse mechanism.

We emphasize that the main difference between the en masse and nucleation and growth mechanisms correspond to their relevant timescales. In the en masse process, subunits bind to the genome in a short time frame forming multiple patches or nuclei, and the structure keeps growing in a disordered manner. When enough subunits are adsorbed, they undergo cooperative rearrangements to assemble into an ordered capsid. By contrast, in the nucleation-elongation mechanism, the longer timescale corresponds to the formation of the first nucleus, and the shell grows by sequential binding of subunits to its edge. The growth phase is always shorter than the nucleation time. The assembly starts at one point and continues until an ordered shell is formed. Both en masse and nucleation-elongation pathways can involve a disorder-order transition.

4.3.1 Nucleation and Growth

We first investigate the assembly of empty capsids where the nucleation and growth is the dominant mechanism and then present our results on the single nucleation mechanism in the presence of the genome.

Empty Capsids

Based on Eqs. 4.1-A.3 in the Method section, the total energy of an empty capsid can be written as

$$E_{tot} = E_{el} + E_{hp} = (\epsilon_{el} + \epsilon)N_T, \tag{4.4}$$

where N_T is the total number of trimers, $E_{el} = E_s + E_b$ is the total elastic energy of the capsid, E_{hp} is the protein-protein interaction (basically hydrophobic). The average elastic

energy and attractive protein-protein interaction per subunit are denoted by ϵ_{el} and ϵ, respectively.

Figure 4.3(a)III illustrates the snapshots of simulations for the growth of an empty shell through the reversible pathway described in the Method section. As the cap grows, at some point, the formation of pentamers becomes energetically unavoidable because of the spherical geometry of the shell. The first pentamer always forms in the vicinity of the cap edge, and then the shell grows around the pentamer such that at the end, the pentamer is in the middle of the cap. The following pentamers appear as illustrated in the figure and the shell grows around them. We note that the hydrophobic interaction always prefers that each subunit has the maximum number of neighbors, six. However, because of the shell curvature this is not always possible. The stretching energy defines the position of pentamers and thus the symmetry of the shell.

Quite interestingly the position of pentamers follows very well the prediction of continuum elasticity theory where the ground state energy of a spherical cap was calculated to obtain the optimal position of disclinations [61, 60], see Fig. A.1. This confirms that the system is completely reversible, equilibrated, and is able to find the minimum energy structures along the pathway.

The fact that most spherical viruses adopt structures with icosahedral symmetry reveals the important role of elasticity in the energetics of viral shells [99, 60, 77]. Nevertheless, it has remained a mystery how an error-free shell formed out of 90 dimers or 60 trimers grows with perfect icosahedral symmetry under many different *in vitro* assembly conditions. One expects that when a considerable part of a shell is formed, the pentamers formed well

inside the cap far from the edge will become more or less frozen at their locations. Thus if a pentamer is formed in a "wrong" position, the icosahedral symmetry should permanently be broken. However, the reversible simulations show that as a shell grows, the elastic energy can become strong enough to easily break the bonds for a wide range of hydrophobic interactions and repair the positions of pentamers formed in the locations not consistent with icosahedral subgroup symmetries.

If the elastic energy is too small or too large compared to the attractive interaction between subunits, the symmetry of the shell will also be broken. To explore the interplay of elasticity and the hydrophobic interaction between subunits in the final structure of shells, we construct a phase diagram for the spring modulus k_s ranging from 20 to 1200 and $\epsilon_{hp} \in [1, 2.2]$ as illustrated in Fig. 4.3b. The value of bending rigidity was chosen based on the experimental values obtained previously [84, 106]. Note that as long as the ratio of the stretching to bending modulus is larger than one, the final assembly products do not depend on the bending modulus, which we have kept it equal to $k_b = 200$ for all the simulations [84, 106].

In region I of the phase diagram in Fig. 4.3, the hydrophobic interaction is so weak that no capsid could nucleate. In region II the elastic energy is weak compared to the hydrophobic interaction, and the shell easily gets kinetically trapped and irregular structures form with total number of subunits less than 60 (a $T = 3$ structure)(Fig. 4.3a). One can see that for a large region of the phase diagram (region III), the competition between elastic and hydrophobic energies is such that the final product is always a $T = 3$ structure. As the elastic energy becomes stronger in region IV, even for very strong hydrophobic interactions,

the shell is not able to close but assumes a structure with cylindrical symmetry and grows in a "branchy" way, forming a messy structure (Fig. 4.3a). While $T = 3$ constitutes the minimum energy structure for the entire phase diagram, messy structures with no symmetry can readily form in different regions as shown in Fig. 4.3. It is not always possible to obtain $T = 3$ structures without using different MC techniques to avoid local traps. We calculate the energy barrier to the formation of $T = 3$ structures in the next section to explain why in some regions the shells can easily get trapped in a local minimum energy structure.

Figure 4.4 shows the snapshots of simulations of a cap in which two pentamers are formed next to each other at the beginning of the assembly. The reversible growth allows the bonds to break and a pentamer to change to a hexamer as the assembled shell grows to have about 25 trimers. Note that the dynamic of system specifically depends on the stretching modulus of subunits, k_s. Lower stretching rigidity slows down the pentamer-hexamer transition, emphasizing the role of elasticity in the kinetic pathway of assembly. Figure A.2 shows that as k_s decreases from $600k_BT$ to $200k_BT$, it takes much longer for the system to repair the position of the pentamer.

Figure 4.4 clearly illustrates that the elastic energy of the shell drops (indicated with an arrow) when the bonds of a "wrong" pentamer are broken. During this time the hydrophobic energy increases, which is the penalty for breaking the bonds, resulting into an energy barrier to the formation of icosahedral shells. However in this case the elastic energy of the shell is strong enough to overcome the local energy barriers and assume the symmetric structure. The total energy of the system is plotted in Fig. A.3. It is important to emphasize that the elastic energy diverges if a shell grows without pentamers.

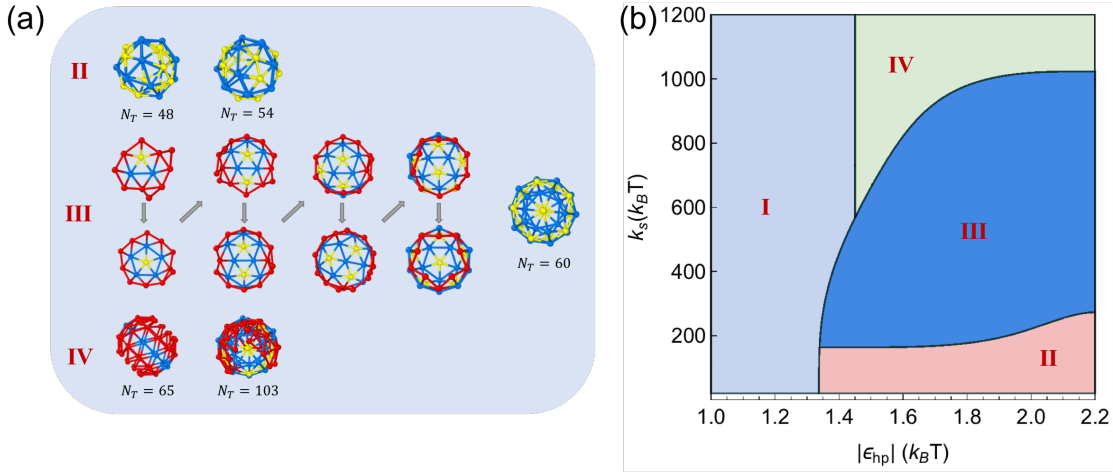

Figure 4.3: Phase diagram for the reversible growth model and snapshots of assembled structures as a function of stretching modulus k_s and monomer-monomer hydrophobic interaction ϵ_{hp}. In region I no shell forms due to lack of nucleation; in region II capsids form irregular shapes with lower symmetries and total trimer number N_T is less than 60; in region III, only $T = 3$ structures form. Snapshots of a representative pathway is illustrated on the panel (a) where pentamers first form at the boundary and then the shell continues to grow around them. Images in panel (a) are made with OVITO[90]. The pathway is consistent and matches very well the ground state calculation corresponding to the locations of pentamers as a shell grows, see Appendix Fig. A.1. In region IV the cost of stretching energy is so high that the structures cannot close but form messy shells. Other parameters of simulations are spontaneous radius $R_0 = 1.5$, bending rigidity $k_b = 200k_BT$ and chemical potential $\mu = -14.6k_BT$.

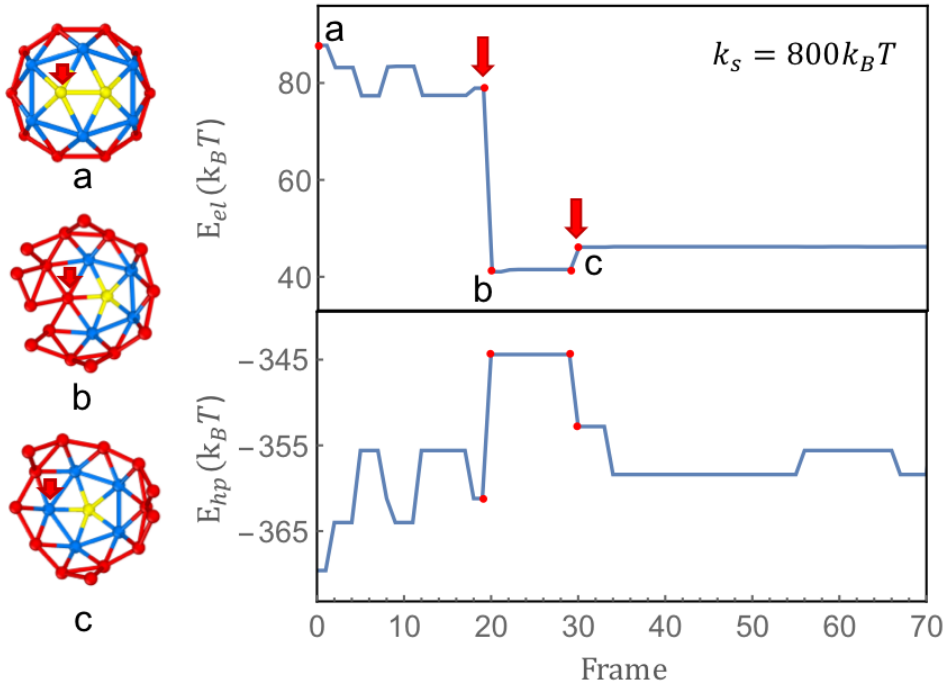

Figure 4.4: Snapshots of simulations for a partially formed capsid with 24 subunits. The right column shows the elastic and hydrophobic energies vs. Frame (number of MC moves) or time. (a) Two pentamers are formed in the vicinity of each other. (b) Around 20 MC steps due to high elastic energy (marked with a red arrow) a bond breaks to decrease the elastic energy while increasing the hydrophobic energy (point b in the energy plots). (c) A hexamer forms and slightly increases the elastic energy while decreasing the hydrophic energy. The protein subunits stretching rigidity is $k_s = 800k_BT$ and bending rigidity $k_b = 200k_BT$. Other parameter used are hydrophobic interaction $\epsilon_{hp} = -1.4k_BT$ and chemical potential $\mu = -14.6k_BT$.

At the end of this section, we emphasize that even though during the simulations one subunit was added at a time, our conclusion remains the same if several subunits join the growing shell simultaneously. The strength of elastic energy as the shell assembles makes the process reversible and the results robust, independent of the number of subunits joining the growing shell.

Classical Nucleation and Energy Barrier

To understand why closed structures without any particular symmetry appear in the phase diagram of Fig. 4.3 where $T = 3$ structure has the lowest energy, we calculate the energy barriers for the formation of $T = 3$ in different regions of the diagram.

For a complete shell, the total hydrophobic interaction is $E_{hp} = 15(Q-4)\epsilon_{hp}$ with Q the total number of subunits and ϵ_{hp} the monomer-monomer interaction as described in the Method section. For a growing shell or a cap, the hydrophobic interaction, Eq. A.3, becomes

$$E_{hp} = \epsilon N_T + \epsilon_l, \tag{4.5}$$

where ϵ_l is the line tension. The average hydrophobic interaction per subunit is $\epsilon = 15\epsilon_{hp}\frac{Q-4}{Q}$, converges to $15\epsilon_{hp}$ as $Q \to +\infty$ (see Hydrophobic Interaction section in SI). Combining Eq. A.4 and Eq. 4.5, we obtain the free energy of the growing capsid as,

$$\frac{E_{tot}}{|\epsilon|} = AN_T + a\sqrt{N_T(Q-N_T)}, \tag{4.6}$$

where $a = \frac{\sqrt{\alpha Q}}{Q-4}$ is a geometric factor, $\alpha = \frac{9\sqrt{3}\pi}{25}$ and $A = \frac{\epsilon_{el}+\epsilon-\mu}{|\epsilon|}$ measures the difference between the chemical potentials of free subunits in solution and in full capsids (see the

appendix for more details). To make it dimensionless, we have divided it by the hydrophobic energy per subunit ϵ. For a $T = 3$ structure

$$A = \frac{\epsilon_{el} + 14\epsilon_{hp} - \mu}{|14\epsilon_{hp}|}. \tag{4.7}$$

The height of energy barrier and the nucleation size can then be calculated through the maximum of the free energy as follows,

$$\begin{aligned} N_T^* &= \frac{1}{2}Q\left(1 + \frac{A}{\sqrt{A^2 + a^2}}\right) \\ {E_{tot}}^* &= \frac{1}{2}|\epsilon|Q\left(A + \sqrt{A^2 + a^2}\right). \end{aligned} \tag{4.8}$$

Figure 4.5 shows the total free energy (Eq. A.5) vs. number of subunits for $A \in (-0.5, 0.5)$ (Eq. 4.7) at three different values of the monomer-monomer hydrophobic interaction, $\epsilon_{hp} = -1.0, -1.4, -1.8$. The figure reveals the impact of different thermodynamics parameters on the height of energy barriers. As illustrated in the figure, for a fixed $\mu = -14.6k_BT$ and a given hydrophobic interaction ϵ_{hp}, the energy barrier increases when the elastic energy per subunit ϵ_{el} increases, see Eq. 4.7. The results shown in Fig. 4.5 is consistent with the phase diagram in Fig. 4.3 in that for $\mu = -14.6k_BT$ if factor A is large because of the strong elastic energy and the small hydrophobic $\epsilon_{hp} = -1k_BT$, then the energy barrier is too high for the formation of capsids and no structure nucleates. In fact, for $\epsilon_{hp} = -1.0k_BT$ regardless of stretching modulus, no capsid nucleates unless the chemical potential is at least $\mu = -10k_BT$. On the other hand, when the hydrophobic interaction is too strong compared to the elastic energy, the nucleation size is less than five subunits resulting in the formation of kinetically trapped smaller structures than $T = 3$.

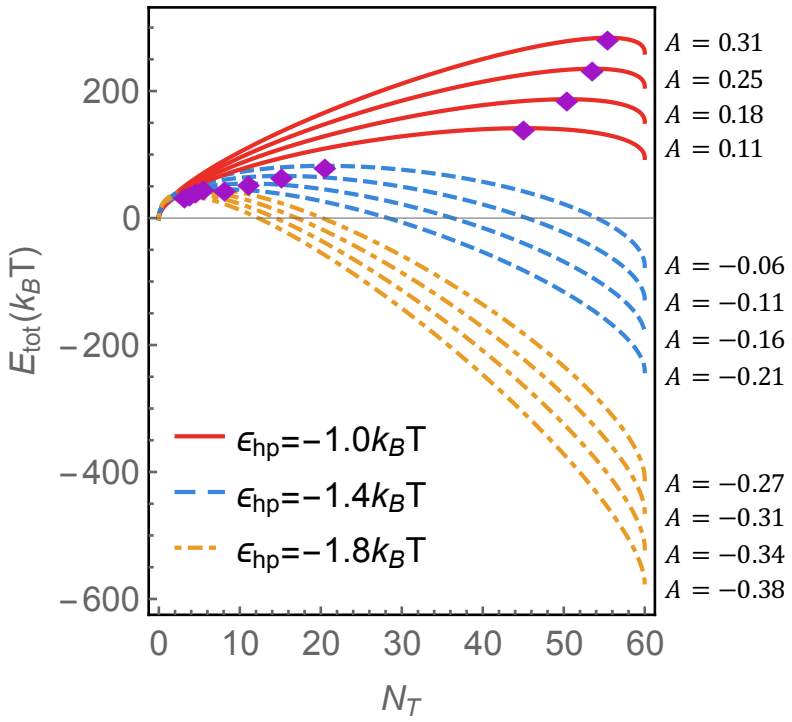

Figure 4.5: Total energy E_{tot} as a function of subunits. The energy barriers are indicated with diamonds. The value of coefficient A (Eq. 4.7) is indicated next to each curve. The stretching moduli used are $k_s = 800, 600, 400, 200 k_BT$ (from top to bottom for each color). The chemical potential is $\mu = -14.6 k_BT$ and the hydrophobic interaction $\epsilon_{hp} = -1.0$ (solid), -1.4k_BT (dashed) and -1.8 (dotted dashed).

In the next section we present the growth of capsids in the presence of the genome through nucleation and growth mechanism.

Assembly in the Presence of the Genome

The self-assembly experiments of Garmann et al. indicate the assembly proceeds through nucleation and growth mechanism with genome present [39]. The process of nucleation and growth on the surface of a gold nanoparticle or condensed genome is very similar to that of an empty capsid. In the former case, we start the simulations assuming that one trimer is sitting on the core interacting attractively with it.

Mimicking the experiments of Garmann *via* we assume that the concentration of proteins is such that the process is not diffusion limited, *i.e.* there are enough proteins in the vicinity of genomes to form capsids. More specifically, we consider that the subunits diffuse from the solution to the core but the duration of attachment of a single subunit to the core is smaller than the nucleation time. Thus the delay in the assembly is just because of nucleation. To this end, we only focus on the assembly of proteins on the core and do not study the diffusion of proteins to the core. The total energy of the complexes of core and proteins can be written as,

$$E_{tot} = E_{lj} + E_{el} + E_{hp} = (-\epsilon_{lj} + \epsilon_{el} + \epsilon)N_T, \tag{4.9}$$

with ϵ_{lj} the depth of the LJ potential representing the core-protein interaction on the core surface. For simplicity we always set $R_{core} = R_0$.

The results of simulations in the presence of the core are presented in Fig. 4.6. For these sets of simulations we have used experiments of Garmann et al. to calibrate our

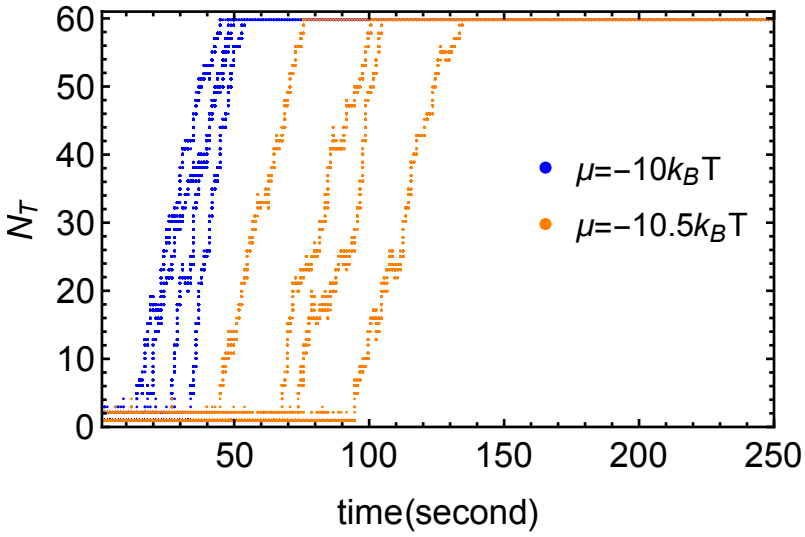

Figure 4.6: Several pathways for the formation of $T = 3$ structures. The plot shows the number of subunits in the assembled caps as a function of time (second) for two different chemical potentials $\mu = -10\ k_BT$(Blue color) and $\mu = -10.5k_BT$(Orange color). For each color four different runs are shown revealing fluctuations in the pathways of formation of $T = 3$ icosahedral structures. The other parameters in simulations are k_s=600k_BT, k_b=200k_BT, ϵ_{hp}=-1k_BT and the genome-protein interaction ϵ_{lj}=1k_BT. For these parameters, the assembly only proceeds through the nucleation and growth mechanism.

MC steps to second [39]. During the simulations, we observe that after about four or five subunits are assembled, the growth becomes very fast. As shown in the figure, it takes about 70s on average for $\mu = -10.5k_BT$ (Orange lines) the capsid to nucleate, which is followed by a rapid growth phase of 33s. As the protein concentration goes higher ($\mu = -10k_BT$, blue lines), the nucleation time becomes shorter, about 24s, and the rapid growth phase becomes about 25s during which the capsid assembles to almost its final size. Note that in these cases, in the absence of genome, no capsids could form. Changing μ from -10.5k_BT to -10k_BT, we increase the protein concentration by 64% and observe that the nucleation time decreases about 65%. This is consistent at least qualitatively with the results of Gramann et al.[39], which according to their experiments, as the protein concentration changes from 1.5μM to 2μM, the nucleation time decreases from about 160s to about 95s. The results of our simulations also shows that the growth time decreases with increasing protein concentration, but with much slower rate than the nucleation time as seen in the experiments.

If we increase the protein concentration even further to $\mu = -8k_BT$, the chance of formation of another nucleus will be high and the process might become multi-nucleus or en masse assembly, as will be explained in the next section. All these simulations show that the interaction of capsid proteins with genome facilitates the formation of capsids, *i.e.* lowers the energy barrier.

The above results reveal that in the presence of the genome, even if the protein-protein interaction is not strong enough for the formation of empty capsids and the protein-genome interaction is weak such that the rate of detachment is larger than the rate of capsid

nucleation, the combined effect of both interactions can give rise to the assembly of filled capsids. Note that if the presence of genome decreases the energy barrier only by one or two k_BT, because of the associated Boltzmann factor the rate of formation of capsid will increase by a factor of three or four.

In the next section, we will study the en masse or multi-nucleation mechanism. First we present our recent experiments and then the results of our simulations assuming that the protein-genome interaction is so strong that the proteins can get easily adsorbed to the genome and then nucleation occurs in different positions on the core.

4.3.2 En Masse Assembly

Experimental Evidence of a Disorder-Order Transition at Neutral pH

The kinetic studies of Chevreuil et al. [14] performed by mixing CCMV subunits (dimers of CPs) and the full (single-stranded RNA) genome at neutral pH, showed that the assembly proceeded through an en masse rather than through a single nucleation and growth process, and the final objects were amorphous nucleoprotein complexes (NPCs). Quite intriguingly, when genome was replaced by a flexible, linear polyelectrolyte, namely, poly(stryrene sulfonic acid) [46, 93], small but closed structures were formed, again *via* the en masse pathway [14]. Why then cannot subunits form ordered structures in the presence of genomic RNA at neutral pH, which is yet physiologically relevant?

Figure 4.7 illustrates the effect of an increase in the concentration of subunits and genomic RNA on the structure of NPCs at neutral pH. SAXS patterns of NPCs (Fig. 4.7a) made with the full CCMV genome exhibit an increasingly pronounced interference peak in

the medium q-values as the concentrations are raised. The peak is even sharper when the full genome is replaced solely by *in vitro* transcribed RNA 2. The peak located around 0.04 $Å^{-1}$ corresponds to length scales around 15 nm and is associated with the spherical symmetry of virus-like objects. Lowering pH yields a similar effect with the onset of two peaks, which are found at the same q-values on the scattering curve of native virions at acidic pH (see appendix Fig. A.6). The reconstructed shapes and their associated sphericity indexes (Fig. 4.7b) suggest that the structures become more ordered and gradually acquire a spherical symmetry at high concentrations. Notice that the reconstructed shape of sample **IV** displays a hollow core (see inset of Fig. 4.7b) like native virions [33]. The reconstructed shapes must be considered with caution though, as they only provide ensemble-averaged, low-resolution structures and cannot reflect the actual variability between individual objects. Note that the analysis of the SAXS patterns by a polydisperse vesicle model (see appendix Fig. A.5) also supports a higher degree of order at high concentrations.

In order to get a better insight into the morphology of the assembled objects, we perform cryo-transmission electron microscopy on a highly concentrated sample of NPCs made with *in vitro* transcribed RNA 2 because the SAXS patterns reveal objects with the highest sphericity index (Fig. 4.7b). Quite consistently, the imaged objects (Fig. 4.8) are mostly ordered and morphologically similar to native virions in these conditions, with a diameter around 30 nm. Some objects are still slightly disordered or aggregated. No empty spherical capsids are found but we can see a few hollow nanotubes (see appendix Fig. A.7) in agreement with the phase diagram of CCMV CPs at neutral pH [55].

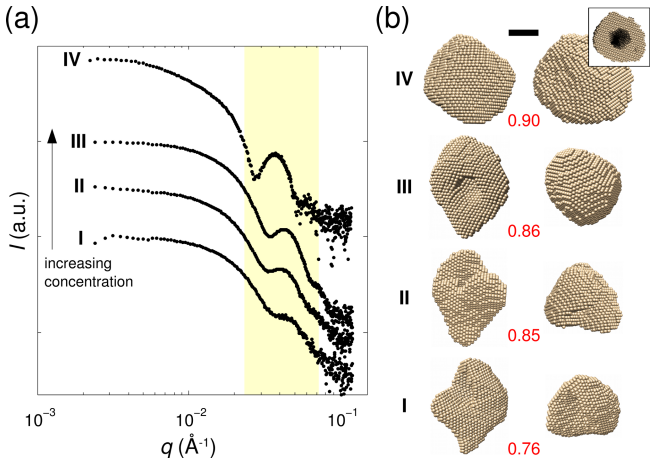

Figure 4.7: Equilibrium structure of NPCs at neutral pH and increasing concentrations in subunits and genomic RNA. (a) SAXS patterns obtained at a fixed subunit-to-RNA mass ratio of 6. Samples **I**, **II** and **III** are made with the full CCMV genome at subunit concentrations of 0.5, 1.0 and 2.1 g.L^{-1}, respectively, whereas sample **IV** contains only *in vitro* transcribed CCMV RNA 2 for a subunit concentration of 2.1 g.L^{-1}. The yellow area highlights the growth of an interference peak in the curves due to an increasingly well-defined length scale in the scattering objects. The scattering curves are shifted for clarity. (b) *Ab initio* shape reconstructions carried out with the scattering curves in (a). The numbers in red are the sphericity indexes calculated for the corresponding structures. The inset is a cross-sectional view of the structure obtained for the sample **IV**. Scale bar is 10 nm.

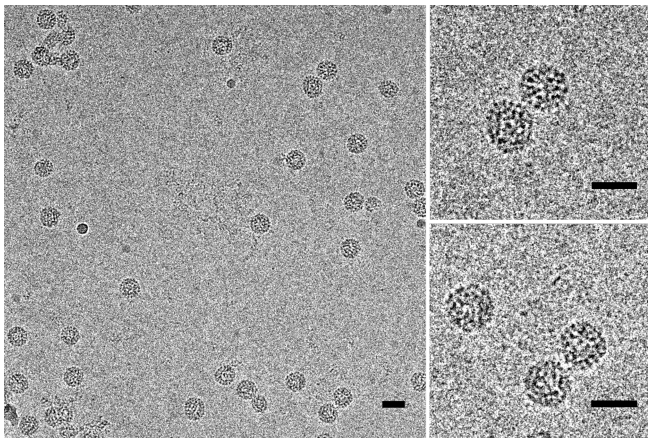

Figure 4.8: Cryo-transmission electron microscopy images of NPCs prepared at neutral pH. CCMV subunits are at a concentration of 3.9 g.L^{-1} and *in vitro* transcribed CCMV RNA 2 is mixed at a subunit-to-RNA mass ratio of 6. Scale bars are 30 nm.

In summary, our experimental studies show that a disorder-order transition occurs within NPCs upon an increase in subunits and genomic RNA concentrations, and that closed spherical shells packaging viral RNA can be assembled at neutral pH. These findings motivate us now to investigate theoretically the ability of the chemical potential to lower the energy barrier between NPCs and assembled virions.

Simulations

The snapshots of the growth of a capsid through en masse is illustrated in Fig. 4.9. Right at the beginning of simulations many subunits get adsorbed to the core and a messy structure forms (Fig. 4.9a), similar to the NPCs observed in previous experiments [38, 14] at pH 7.5 and the recent ones presented in the previous section. Special care is taken that there is no overlap between subunits. We note that as for the simulations in which the nucleation and growth is the dominant mechanism, we do not study the diffusion of proteins into the genome in case of en masse (or multi-nucleations) assembly either. More specifically, we consider that the assembly is reaction limited and thus the important time scale in the problem involves the addition of subunits to each other in order to form larger patches. Due to strong protein-genome interaction, we start simulations with about 10 to 15 subunits attached to the genome and consider that the growth proceeds basically through joining of the new subunits to other subunits already attached to the genome. Figure 4.9b shows a snapshot of the simulations when the interaction is weak ($\epsilon_{hp} = -0.7$ k_BT), and the number of trimers is 40 with multiple nuclei already formed. We continue adding the subunits randomly to the edge of one of the existing subunits and accept the move as explained in the Method section. The structure is still disordered.

The previous self-assembly experiments[38, 14] showed that when the pH is lowered, the attractive interaction between the protein subunits becomes stronger. Following the experimental steps, we also decrease ϵ_{hp} from -0.7 k_BT to -1 k_BT, when there are 45 trimers on the genome. Around 5000s, 54 trimers are relaxed to their minimum energy positions in Fig. 4.9c. The shell will form a $T = 3$ structure around 14000s (Fig. 4.9d). The kinetics of assembly through en masse pathway is illustrated in Fig. 4.10 in the form of the number of subunits vs. time. The behavior of the growth curves obtained in the en masse assembly is very similar to those of Chevreuil et al. in that upon decreasing pH, a disordered to ordered transition occurs. We have employed the results of Fig. 3a in Ref. [14] to calibrate the MC steps with time.

In addition to the pH change, the experiments presented in the previous section reveal the important role of protein concentrations in the disordered to ordered transition. To explore the impact of protein concentration and protein-protein interaction and to explain the experiments presented in Fig. 4.7, we build a phase diagram for the formation of closed shells as a function of chemical potential and hydrophobic (attractive) interaction, see Fig. 4.11.

Figure 4.11 shows that for fixed $k_s = 600k_BT$ and $k_b = 200k_BT$, at high protein concentrations and strong protein-protein attractive interaction, only $T = 3$ capsids form (region III). Note that if we increase both protein concentration and the strength of protein interaction significantly, the structures will be stuck in a local minimum energy and aberrant particles form, not shown in the phase diagram. In contrast at low hydrophobic interactions and low protein concentration no structures nucleate (region I). We note that in the presence

of packaging signals with high affinity for CP, the capsid forms even at very low protein concentration as the packaging signal increases the effective attractive interaction between the genome and proteins [89].

In the purple region (VI) of the diagram in Fig. 4.11, only messy/amorphous structures form as observed in the experiments at pH=7.5. In this regime the protein-protein interaction is not strong enough to form an ordered structure. Nevertheless, the figure shows that for a fixed hydrophobic interaction, upon increasing protein concentrations, a disorder-order transition can occur, see for example the region around $\mu = -14k_BT$ and $\epsilon_{hp} = -0.8kBT$. As illustrated in the figure, if for $\epsilon_{hp} = -0.8k_BT$, the chemical potential increases from -15 to -13 k_BT, one can observe the formation of $T = 3$ structures, consistent with the experimental results shown in Fig. 4.7. It is worth mentioning that the transition from a disordered nucleoprotein complex to an icosahedral virion was reported several times by lowering the pH[14, 38, 36, 8] (see Fig. A.6), which is equivalent to increasing ϵ_{hp}, for a fixed μ in Figure 11. The experiments in this chapter corresponds to the evidence of a disorder-order transition at a fixed ϵ_{hp} – through fixed pH and ionic strength – by decreasing μ *via* the increase of subunit concentration. Consistent with the phase diagram depicted in Fig. 4.11, SAXS measurements show that upon an increase of subunit concentration (the subunit-to-RNA mass ratio is set to 6 in all cases), the nucleoprotein complexes become gradually ordered and cryo-TEM reveals the presence of a number of well assembled virions at high concentration.

Quite interestingly, when we change the size of core to be commensurate with a capsid whose diameter is 23 nm consistently with the assembly experiments in the presence

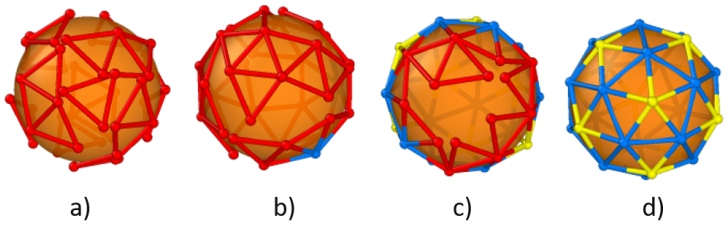

Figure 4.9: Snapshots of simulations in Fig. 4.10. The chemical potential is $\mu = -14k_BT$, k_s=600k_BT, k_b=200k_BT and $\epsilon_{lj} = 11.2k_BT$. (a) View of an amorphous nucleoprotein complex after less than a second for $\epsilon_{hp} = -0.7k_BT$. There are around 20 subunits on the surface of the core.(b) View of a disordered structure. After 96s there are 40 trimers on the genome and multiple nuclei have formed. Later when capsid has 45 trimers, we increase the hydrophobic strenght by changing ϵ_{hp} from $-0.7k_BT$ to $-1k_BT$. This is consistent with the experiments when the pH is lowered from 7.5 to 5.2. (c) View of a partially formed capsid with 54 trimers after 5000s. Subunits are relaxed to their lowest minimum energy positions.(d) The structure of a closed $T = 3$ capsid with 60 trimers. It takes around 14000s for the capsid to complete. Red colors correspond to vertices on the edge. Blue and yellow colors show the position of hexamers and pentamers respectively. The images are made with OVITO[90].

of PSS, we obtain a structure with the symmetry of a tennis ball (see Fig. 3.3), which has been previously observed in self-assembly studies of clathrin shells [99, 77]. The phase diagram for PSS is very similar to that for RNA but the boundary lines between ordered-disordered configurations have moved, see Fig. A.8 in SI. We find that capsid proteins assemble to form closed shells at lower protein concentration that are needed for assembly with RNA. This behavior supports at least qualitatively the experimental results with PSS [14].

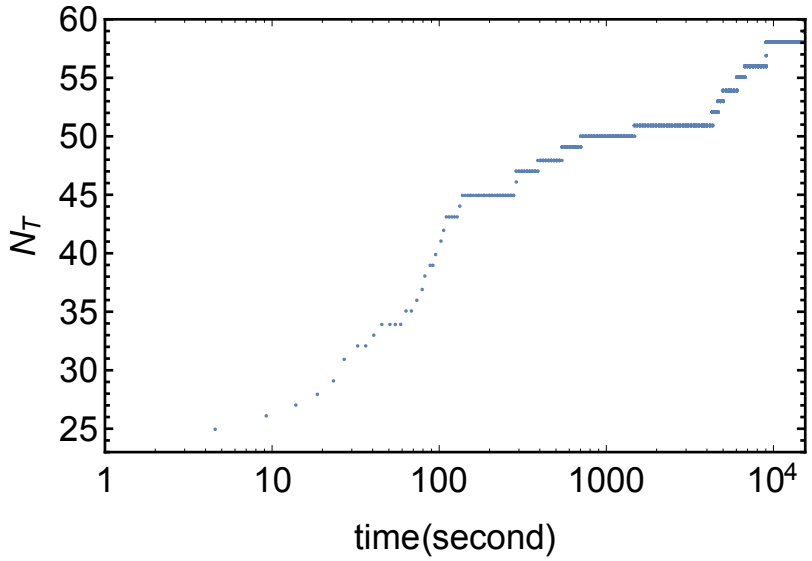

Figure 4.10: The number of subunits assembled around the genome vs. time (second) for en masse (multi nucleation) assembly at $\mu = -14k_BT$, k_s=$600k_BT$, k_b=$200k_BT$ and $\epsilon_{lj} = 11.2k_BT$. The number of subunits increases from 23 to around 45 trimers in the first 130s where the hydrophobic strength is weak ($\epsilon_{hp} = -0.7k_BT$). At $N_T = 45$, Decreasing ϵ_{hp} to $-1k_BT$ partially formed disordered capsid starts relaxing to the final structure with T=3 symmetry. The process of relaxation and reorganization to an ordered capsid takes around 14000s.

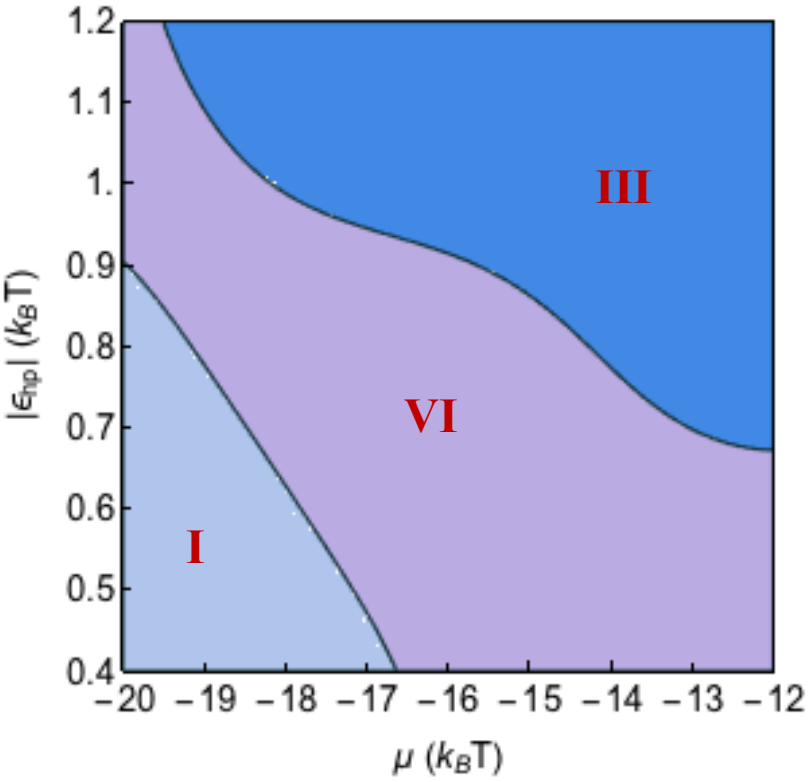

Figure 4.11: Phase diagram of structures obtained from en masse simulations for different values of chemical potentials and the strength of hydrophobic interactions. The strong genome-protein attractive interaction ($\epsilon_{lj} = 11.2\ k_BT$) makes the assembly proceed through en masse pathway. The blue shade corresponds to the region in which $T = 3$ capsids form but the purple one represents the phase where amorphous structures assemble. No capsid nucleates in light blue region. As we move from left to right, the concentration of free subunits increases and we observe the disordered to ordered transition. The other parameters in the simualtions are k_s=$600k_BT$, k_b=$200k_BT$ and the spontaneous radius $R_0 = 1.5$.

4.4 Conclusion

Since during the assembly of viral particles, the intermediate states are transient and as such not easily accessible experimentally, the most fundamental questions about assembly pathways have remained unresolved. In this chapter, we have focused on two different mechanisms observed in recent experiments: one supporting the en masse assembly and the other the nucleation and growth pathway.

Our results suggest at least in part that the elastic energy plays a crucial role in the assembly of error-free symmetric shells in both cases. We have found that as a shell grows, if the pentamers form in the positions that break icosahedral order, through a reversible pathway, the pentamers can move to preserve the symmetry and form a $T = 3$ shell. This is counter-intuitive as many bonds need to break for a pentamer formed far from the edge, well inside the capsid to move. However, our results show that the elastic energy becomes so strong that as many bonds as necessary can be broken to move the pentamer, see Fig. 4.4. Obviously, the relative strength of hydrophobic interaction and elastic energy is very important in this process, see Figs. 4.3 and 4.5. Figure 4.12 shows a messy shell that was initially assembled fast in the presence of the high protein concentration. These sorts of shells can form either through nucleation and growth or en masse pathway. Upon decreasing the protein concentration, we find that the shell relaxes to a $T = 3$ structure through reversible pathway. The elastic energy indeed plays a crucial role in preserving the symmetry in transferring disordered to ordered structures.

The important role of protein concentration in the assembly pathway is summarized in Fig. 4.13 for two different values of genome-protein interaction, ϵ_{lj}. If the genome-

protein interaction is weak $\epsilon_{lj} = 1k_BT$, for the low protein concentration, no shell nucleates. As the protein concentration increases, the shell grows through the nucleation and growth pathway. At higher concentrations, we observe that subunits get adsorbed onto the spherical core and shell grows through en masse assembly into a $T = 3$ structure. In contrast for large $\epsilon_{lj} = 11.2k_BT$ at low protein concentrations, amorphous structures form. Upon increasing protein concentration ordered structures are obtained through en masse assembly.

These results are consistent with our SAXS measurements at neutral pH, which clearly show an increasing degree of order upon the increase of concentration. While even our best curves are not as regular as those obtained with native virions, in marked contrast with the previous experiments we find that some structures look spherical and fully assembled at neutral pH. We note that the concentration may not be high enough or other factors such as crowding effect or divalent cations may be important in the formation of icosahedral capsids as well. Nevertheless, our experimental and numerical results clearly demonstrate that the chemical potential can significantly lower the energy barrier between amorphous complexes and virions, and it must have definitively a role *in vivo* assembly as well.

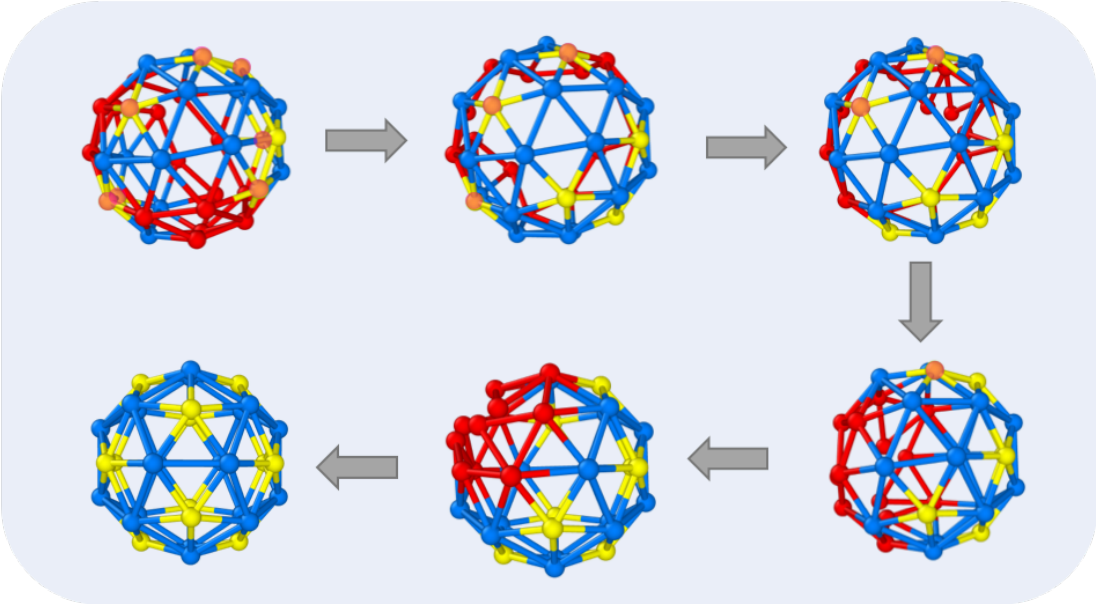

Figure 4.12: Relaxation of a messy shell to an icosahedral structure. At the begining, the shell has 57 trimers with 7 pentamers formed in "wrong" positions (marked with orange color). The shell relaxes to a T=3 capsid upon a change in pH or protein concentration. The blue, yellow and red colors correspond to hexamers, pentamers and edge respectively. The parameters used are stretching rigidity $k_s = 200$, bending rigidity $k_b = 200$, hydrophobic interaction $\epsilon_{hp} = 1.0$ and chemical potential $\mu = -14$. The images are made with OVITO[90].

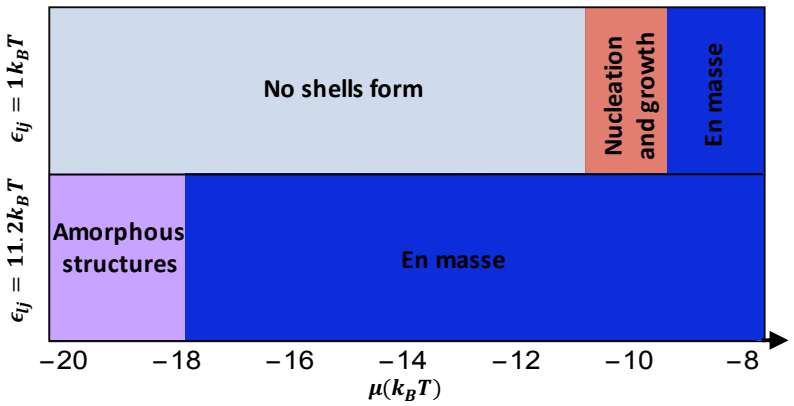

Figure 4.13: Capsids form with different mechanisms based on the strength of protein-genome interaction. At weak Lennard-Jones interaction ($\epsilon_{lj} = 1.0 k_B T$), for the chemical potentials between -11 $k_B T$ and -9 $k_B T$ capsids form through the nucleation and growth mechanism. For higher chemical potentials, the en masse assembly becomes the dominant pathway for the formation of $T = 3$ structures. However, at strong Lennard-Jones interaction ($\epsilon_{lj} = 11.2\ k_B T$) we obtain amorphous structures for low chemical potentials but T=3 structures form through en masse pathway for higher chemical potentials. Other parameters used in the simulations are k_s=600$k_B T$, k_b=200$k_B T$ and $\epsilon_{hp} = -1\ k_B T$.

Chapter 5

Conclusions

Viruses and many protein nano cages in nature spontaneously assemble into symmetric structures. Finding the optimal structure of empty protein nano cages and also investigating the factors contributing to the stability of protein cages and specifically viral shells are of important interests because of their applications in drug design and genome delivery. Our studies show that regardless of growth pathways, viruses mostly form icosahedral structures. We find that mechanical properties of CPs such as stretching and bending rigidity and also the dihedral angle between CPs are key factors in determining the final shape of protein cages. We show that when CPs are very rigid icosahedral structures become dominant structures, however; by decreasing the stretching rigidity of CPs different symmetric shells become equilibrium structures.

The simplest viruses contain an RNA genome that is packaged by a shell called capsid. The driving force for the assembly of CPs around genome is electrostatic interaction. How CPs select and package its cognate RNA inside crowded cellular milieu while there are

many nonviral RNA and other anionic polymers is a mystery. In this thesis, we show that CPs can assemble around a globular RNA and form nonicosahedral cages. However, our calculations shows that these structures are strained and can easily split into large fragments and provide intermediates for the formation of stable icosahedral capsids if the larger RNA becomes available. This is an important result since it confirms the possibility of a pathway for packaging genome inside cells. Besides, our studies show how the genome size and strength of interaction between CPs and genome can affect the morphology of capsids. While interplay between elastic energy of a capsid and attractive interaction between genome and CPs can lead to the formation of strained structures, in some cases CPs may prefer to form empty capsids without packaging genome. Finally, Our assembly simulations around a core show the formation of a structure with specific symmetry that has not been seen in previous theoretical studies and is reported by a experimental research recently.

In addition, this thesis studies the two major kinetic pathways of virus assembly around genome: 'en masse' and 'nucleation and growth' using Monte Carlo simulations in a grand canonical ensemble. Considering that the kinetics of CPs assembly around RNA is much faster than the formation of empty capsids, the question that arises is how symmetry is maintained in capsids in such high speed of assembly. Experiments show that in 'en masse' pathway, CPs are rapidly adsorbed to the genome and form a disordered structures which later form into ordered structures by increasing CPs concentration. Our simulations confirm such disordered-ordered transition and shows that elastic energies play a key role in this transition. We also study the interplay between hydrophobic interactions between CPs-CPs, elastic energy and chemical potential in the formation of error free capsids.

Bibliography

[1] Roi Asor, Lisa Selzer, Christopher John Schlicksup, Zhongchao Zhao, Adam Zlotnick, and Uri Raviv. Assembly reactions of hepatitis b capsid protein into capsid nanoparticles follow a narrow path through a complex reaction landscape. *ACS Nano*, 2019.

[2] William M Aumiller, Masaki Uchida, and Trevor Douglas. Protein cage assembly across multiple length scales. *Chemical Society Reviews*, 47(10):3433–3469, 2018.

[3] J. B. Bancroft. The Self-Assembly of Spherical Plant Viruses. *Adv. Virus Res.*, 16:99, 1970.

[4] Ray H Baughman, Douglas S Galvão, Changxing Cui, Yang Wang, and David Tománek. Fullereneynes: a new family of porous fullerenes. *Chemical physics letters*, 204(1-2):8–14, 1993.

[5] Kevin Bond, Irina B Tsvetkova, Joseph Che-Yen Wang, Martin F Jarrold, and Bogdan Dragnea. Virus assembly pathways: Straying away but not too far. *Small*, 16(51):2004475, 2020.

[6] Alexander Borodavka, Roman Tuma, and Peter G Stockley. Evidence that viral rnas have evolved for efficient, two-stage packaging. *PNAS*, 109(39):15769–15774, 2012.

[7] Robijn F Bruinsma, Mauricio Comas-Garcia, Rees F Garmann, and Alexander Y Grosberg. Equilibrium self-assembly of small rna viruses. *Phys. Rev. E*, 93(3):032405, 2016.

[8] Ruben D Cadena-Nava, Mauricio Comas-Garcia, Rees F Garmann, ALN 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.

[9] Mauricio Carrillo-Tripp, Craig M Shepherd, Ian A Borelli, Sangita Venkataraman, Gabriel Lander, Padmaja Natarajan, John E Johnson, Charles L Brooks III, and Vijay S Reddy. Viperdb2: an enhanced and web api enabled relational database for structural virology. *Nucleic acids research*, 37(suppl_1):D436–D442, 2008.

[10] Donald LD Caspar and Aaron Klug. Physical principles in the construction of regular viruses. *Cold Spring Harbor Symp. Quant. Biol.*, 27:1–24, 1962.

[11] Ting Chen, Zhenli Zhang, and Sharon C. Glotzer. A precise packing sequence for self-assembled convex structures. *Proceedings of the National Academy of Sciences*, 104:717, 2007.

[12] Ting Chen, Zhenli Zhang, and Sharon C. Glotzer. Simulation studies of the self-assembly of cone-shaped particles. *Langmuir*, 23(12):6598–6605, 2007. PMID: 17489618.

[13] 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.

[14] 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.

[15] Mauricio Comas-Garcia. Packaging of genomic rna in positive-sense single-stranded rna viruses: A complex story. *Viruses*, 11(3):253, 2019.

[16] Francis HC Crick and James D Watson. Structure of small viruses. *Nature*, 177:473–5, 1956.

[17] RA Crowther, JT Pinch, and BMF Pearse. On the structure of coated vesicles. *Journal of molecular biology*, 103(4):785–798, 1976.

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

[19] T Dokland. Scaffolding proteins and their role in viral assembly. *Cellular and Molecular Life Sciences CMLS*, 56(7-8):580–603, 1999.

[20] 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(6):062423, 2020.

[21] 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.

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

[23] Oren M Elrad and Michael F Hagan. Mechanisms of size control and polymorphism in viral capsid assembly. *Nano lett.*, 8(11):3850–3857, 2008.

[24] Oren M Elrad and Michael F Hagan. Encapsulation of a polymer by an icosahedral virus. *Phys Biol*, 7(4):045003, 2010.

[25] Gonca Erdemci-Tandogan, Henri Orland, and Roya Zandi. Rna base pairing determines the conformations of rna inside spherical viruses. *Phys. Rev. Lett*, 119(18):188102, 2017.

[26] Gonca Erdemci-Tandogan, Jef Wagner, Paul Van Der Schoot, Rudolf Podgornik, and Roya Zandi. Rna topology remolds electrostatic stabilization of viruses. *Phys. Rev. E*, 89(3):032707, 2014.

[27] Gonca Erdemci-Tandogan, Jef Wagner, Paul van der Schoot, Rudolf Podgornik, and Roya Zandi. Effects of rna branching on the electrostatic stabilization of viruses. *Physical Review E*, 94(2):022408, 2016.

[28] S. Fejer, D. Chakrabarti, and D. Wales. Emergent complexity from simple anisotropic building blocks: Shells, tubes, and spirals. *Nano Lett*, 4:219, 2010.

[29] Szilard N Fejer, Dwaipayan Chakrabarti, and David J Wales. Emergent complexity from simple anisotropic building blocks: Shells, tubes, and spirals. *ACS nano*, 4(1):219–228, 2010.

[30] SJ. Flint, LW. Enquist, RM. Krug, VR. Racaniello, and A Skalka. *Principles of Virology: Molecular Biology, Pathogenesis, and Control*. ASM Press: Washington D.C., 2000.

[31] Juan Fontana, Daniel Nemecek, Colleen A McHugh, Anastasia A Aksyuk, Naiqian Cheng, Dennis C Winkler, J Bernard Heymann, Egbert Hoiczyk, and Alasdair C Steven. Phage capsid-like structure of myxococcus xanthus encapsulin, a protein shell that stores iron. *Microscopy and Microanalysis*, 20(S3):1244–1245, 2014.

[32] Alexander Fotin, Yifan Cheng, Piotr Sliz, Nikolaus Grigorieff, Stephen C Harrison, Tomas Kirchhausen, and Thomas Walz. Molecular model for a complete clathrin lattice from electron cryomicroscopy. *nature*, 432(7017):573, 2004.

[33] J. M. Fox, G. Wang, J. A. Speir, N. H. Olson, J. E. Johnson, T. S. Baker, and M. J. Young. Comparison of the native ccmv virion with In Vitro assembled ccmv virions by cryoelectron microscopy and image reconstruction. *Virology*, 244:212–218, 1998.

[34] Daniel Franke and Dmitri I. Svergun. Dammif, a Program for Rapid ab-initio Shape Determination in Small-Angle Scattering. *J. Appl. Cryst.*, 42:342–346, 2009.

[35] Daan Frenkel and Berend Smit. *Understanding Molecular Simulation: From Algorithms to Applications*, volume 1. Elsevier: San Diego, 2001.

[36] Rees F Garmann, Mauricio Comas-Garcia, Ajaykumar Gopal, Charles M Knobler, and William M Gelbart. The assembly pathway of an icosahedral single-stranded rna virus depends on the strength of inter-subunit attractions. *J. Mol. Biol.*, 426(5):1050–1060, 2014.

[37] Rees F Garmann, Mauricio Comas-Garcia, Charles M Knobler, and William M Gelbart. Physical principles in the self-assembly of a simple spherical virus. *Accounts of chemical research*, 49(1):48–55, 2015.

[38] 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.

[39] 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.

[40] M. F. Hagan. Modeling Viral Capsid Assembly. *Adv. Chem. Phys.*, 155:1, 2014.

[41] M. F. Hagan and R. Zandi. Recent advances in coarse-grained modeling of virus assembly. *Curr. Opin. Virol.*, 18:36–43, 2016.

[42] Michael F. Hagan and Gregory M. Grason. Equilibrium Mechanisms of Self-Limiting Assembly. *Rev. Mod. Phys.*, 93(2):025008, jun 2021.

[43] Jonathan G Heddle. Protein cages, rings and tubes: useful components of future nanodevices? *Nanotechnology, science and applications*, 1:67, 2008.

[44] Stephen D. Hicks and C. L. Henley. Irreversible growth model for virus capsid assembly. *Phys. Rev. E*, 74:031912, Sep 2006.

[45] Stephen D Hicks and CL Henley. Irreversible growth model for virus capsid assembly. *Phys. Rev. E*, 74(3):031912, 2006.

[46] Yufang Hu, Roya Zandi, Adriana Anavitarte, Charles M. Knobler, and William M. Gelbart. Packaging of a Polymer by a Viral Capsid: The Interplay between Polymer Length and Capsid Size. *Biophys. J.*, 94(4):1428–1436, 2008.

[47] Chantal Hulo, Edouard De Castro, Patrick Masson, Lydie Bougueleret, Amos Bairoch, Ioannis Xenarios, and Philippe Le Mercier. Viralzone: a knowledge resource to understand virus diversity. *Nucleic acids research*, 39(suppl_1):D576–D582, 2010.

[48] John E Johnson and Jeffrey A Speir. Quasi-equivalent viruses: a paradigm for protein assemblies1. *Journal of molecular biology*, 269(5):665–675, 1997.

[49] Willem K. Kegel and Paul Competing Hydrophobic Van Der Schoot. Competing Hydrophobic and Screened-Coulomb Interactions in Hepatitis B Virus Capsid Assembly. *Biophys. J.*, 86(6):3905–3913, 2004.

[50] Andrew MQ King, Elliot Lefkowitz, Michael J Adams, and Eric B Carstens. *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*, volume 9. Elsevier, San Diego, 2011.

[51] T Kohyama, DM Kroll, and G Gompper. Budding of crystalline domains in fluid membranes. *Physical Review E*, 68(6):061905, 2003.

[52] Harold W Kroto, James R Heath, Sean C O'Brien, Robert F Curl, and Richard E Smalley. C60: Buckminsterfullerene. *Nature*, 318(6042):162, 1985.

[53] 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.

[54] S. B. Larson and A. McPherson. Satellite Tobacco Mosaic Virus RNA: Structure and Implications for Assembly. *Curr. Opin. Struc. Biol.*, 11:59, 2001.

[55] 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.

[56] Artem Levandovsky and Roya Zandi. Nonequilibirum assembly, retroviruses, and conical structures. *Physical review letters*, 102(19):198102, 2009.

[57] Artem Levandovsky and Roya Zandi. Nonequilibirum assembly, retroviruses, and conical structures. *Phys. Rev. Lett*, 102(19):198102, 2009.

[58] Cyrus Levinthal. How to fold graciously. *Mossbauer Spectroscopy in Biological Systems*, 67:22–24, 1969.

[59] Siyu Li, Gonca Erdemci-Tandogan, Paul van der Schoot, and Roya Zandi. The effect of rna stiffness on the self-assembly of virus particles. *J. Condens. Matter Phys.*, 30(4):044002, 2017.

[60] Siyu Li, Polly Roy, Alex Travesset, and Roya Zandi. Why large icosahedral viruses need scaffolding proteins. *Proceedings of the National Academy of Sciences*, 2018.

[61] Siyu Li, Roya Zandi, Alex Travesset, and Gregory M. Grason. Ground states of crystalline caps: Generalized jellium on curved space. *Phys. Rev. Lett.*, 123:145501, Sep 2019.

[62] Z Li and H A Scheraga. Monte carlo-minimization approach to the multiple-minima problem in protein folding. *Proceedings of the National Academy of Sciences*, 84(19):6611–6615, 1987.

[63] Jack Lidmar, Leonid Mirny, and David R Nelson. Virus shapes and buckling transitions in spherical shells. *Physical Review E*, 68(5):051910, 2003.

[64] H.-K. Lin, P. van der Schoot, and R. Zandi. Impact of Charge Variation on the Encapsulation of Nanoparticles by Virus Coat Proteins. *Phys. Biol.*, 9:066004, 2012.

[65] Robert W Lucas, Yurii G Kuznetsov, Steven B Larson, and Alexander McPherson. Crystallization of Brome Mosaic Virus and T = 1 Brome Mosaic Virus Particles Following a Structural Transition. *Virology*, 286(2):290–303, 2001.

[66] Kai Ma, Yunye Gong, Tangi Aubert, Melik Z Turker, Teresa Kao, Peter C Doerschuk, and Ulrich Wiesner. Self-assembly of highly symmetrical, ultrasmall inorganic cages directed by surfactant micelles. *Nature*, page 1, 2018.

[67] Eric R. May. Recent developments in molecular simulation approaches to study spherical virus capsids. *Mol Simul*, 40(10-11):878–888, 2014.

[68] Colleen A McHugh, Juan Fontana, Daniel Nemecek, Naiqian Cheng, Anastasia A Aksyuk, J Bernard Heymann, Dennis C Winkler, Alan S Lam, Joseph S Wall, Alasdair C Steven, and Egbert Hoiczyk. A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. *The EMBO journal*, 33(17):1896–1911, 2014.

[69] Junseon Min, Soohyun Kim, Jisu Lee, and Sebyung Kang. Lumazine synthase protein cage nanoparticles as modular delivery platforms for targeted drug delivery. *RSC Advances*, 4(89):48596–48600, 2014.

[70] Pepijn Moerman, Paul Van Der Schoot, and Willem Kegel. Kinetics versus thermodynamics in virus capsid polymorphism. *The Journal of Physical Chemistry B*, 120(26):6003–6009, 2016.

[71] Farzaneh Mohajerani and Michael F Hagan. The role of the encapsulated cargo in microcompartment assembly. *PLoS Comput. Biol.*, 14(7):e1006351, 2018.

[72] TT Nguyen, RF Bruinsma, and WM Gelbart. Continuum theory of retroviral capsids. *Physical review letters*, 96(7):078102, 2006.

[73] Jiying 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.

[74] Jiying 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. *Nat. Commun.*, 7:13689, 2016.

[75] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, New York, NY, USA, second edition, 2006.

[76] 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.

[77] Sanaz Panahandeh, Siyu Li, and Roya Zandi. The equilibrium structure of self-assembled protein nano-cages. *Nanoscale*, 10(48):22802–22809, 2018.

[78] Stefan Paquay, Halim Kusumaatmaja, David J. Wales, Roya Zandi, and Paul van der Schoot. Energetically favoured defects in dense packings of particles on spherical surfaces. *Soft Matter*, 12:5708–5717, 2016.

[79] Stefan Paquay, Halim Kusumaatmaja, David J Wales, Roya Zandi, and Paul van der Schoot. Energetically favoured defects in dense packings of particles on spherical surfaces. *Soft Matter*, 12(26):5708–5717, 2016.

[80] 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.

[81] Jason D Perlmutter, Cong Qiao, and Michael F Hagan. Viral genome structures are optimal for capsid assembly. *Elife*, 2, 2013.

[82] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. Ucsf Chimera–A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.*, 25(13):1605–1612, 2004.

[83] A L N Rao. Genome Packaging by Spherical Plant RNA Viruses. *Annu. Rev. Phytopathol.*, 44:61–87, 2006.

[84] WH Roos, R Bruinsma, and GJL Wuite. Physical virology. *Nat. Phys.*, 6(10):733, 2010.

[85] Grant M Rotskoff and Phillip L Geissler. Robust nonequilibrium pathways tofleva microcompartment assembly. *Proceedings of the National Academy of Sciences*, 115(25):6341–6346, 2018.

[86] Antonio Šiber and Rudolf Podgornik. Nonspecific interactions in spontaneous assembly of empty Versus functional single-stranded rna viruses. *Phys. Rev. E*, 78(5):051915, 2008.

[87] 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*, 119:13991, 2015.

[88] 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.

[89] Peter G Stockley, Reidun Twarock, Saskia E Bakker, Amy M Barker, Alexander Borodavka, Eric Dykeman, Robert J Ford, Arwen R Pearson, Simon EV 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(2):277–287, 2013.

[90] Alexander Stukowski. Visualization and Analysis of Atomistic Simulation Data With OVITO-The Open Visualization Tool. *Model. Simul. Mater. Sci. Eng.*, 18(1), JAN 2010.

[91] 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. *PNAS*, 104(4):1354–1359, 2007.

[92] D. I. Svergun. Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *J. Appl. Crystallogr.*, 25:495–503, 1992.

[93] Guillaume Tresset, Mouna Tatou, Clémence Le Coeur, Mehdi Zeghal, Virginie Bailleux, Amélie Lecchi, Katarzyna Brach, Magdalena Klekotko, and Lionel Porcar. Weighing polyelectrolytes packaged in viruslike particles. *Phys. Rev. Lett.*, 113:128305, 2014.

[94] L. Tubiana, A. L. Bozic, C. Micheletti, and R. Podgornik. Synonymous Mutations Reduce Genome Compactness in Icosahedral ssRNA Viruses. *Biophys. J.*, 108:194, 2015.

[95] 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.

[96] Reidun Twarock and Antoni Luque. Structural puzzles in virology solved with an overarching icosahedral design principle. *Nat. Commun.*, 10(1):1–9, 2019.

[97] Reidun Twarock and Peter G Stockley. Rna-mediated virus assembly: Mechanisms and consequences for viral evolution and therapy. *Annu. Rev. Biophys*, 48:495–514, 2019.

[98] Graziano Vernizzi, Rastko Sknepnek, and Monica Olvera De La Cruz. Platonic and archimedean geometries in multicomponent elastic membranes. *Proceedings of the National Academy of Sciences*, 108(11):4292–4296, 2011.

[99] Jef Wagner and Roya Zandi. The robust assembly of small symmetric nanoshells. *Biophysical journal*, 109(5):956–965, 2015.

[100] Peter Weiss. Armor-plated puzzle: Deciphering the code of viral geometry. *Science News*, 168(10):152–154, 2005.

[101] Paul T Wingfield, Stephen J Stahl, Robert W Williams, and Alasdair C Steven. Hepatitis core antigen produced in escherichia coli: subunit composition, conformation analysis, and in vitro capsid assembly. *Biochemistry*, 34(15):4919–4932, 1995.

[102] Aron M Yoffe, Peter Prinsen, William M Gelbart, and Avinoam Ben-Shaul. The ends of a large rna molecule are necessarily close. *Nucleic Acids Res.*, 39(1):292–299, 2011.

[103] Roya Zandi, Bogdan Dragnea, Alex Travesset, and Rudolf Podgornik. On Virus Growth and Form. *Phys. Rep.*, 847:1–102, mar 2020.

[104] 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.

[105] Roya Zandi, Paul van der Schoot, David Reguera, Willem Kegel, and Howard Reiss. Classical nucleation theory of virus capsids. *Biophys. J.*, 90(6):1939–1948, 2006.

[106] 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.

[107] Cheng Zeng, Liam Scott, Andrey Malyutin, Roya Zandi, Paul Van der Schoot, and Bogdan Dragnea. Virus mechanics under molecular crowding. *J. Phys. Chem*, 125(7):1790–1798, 2021.

[108] Ran Zhang and Per Linse. Topological effects on capsomer-polyion co-assembly. *J. Chem. Phys.*, 140(24):244903, 2014.

[109] 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.

Appendix A

A.1 Elastic Energy of a Spherical Cap

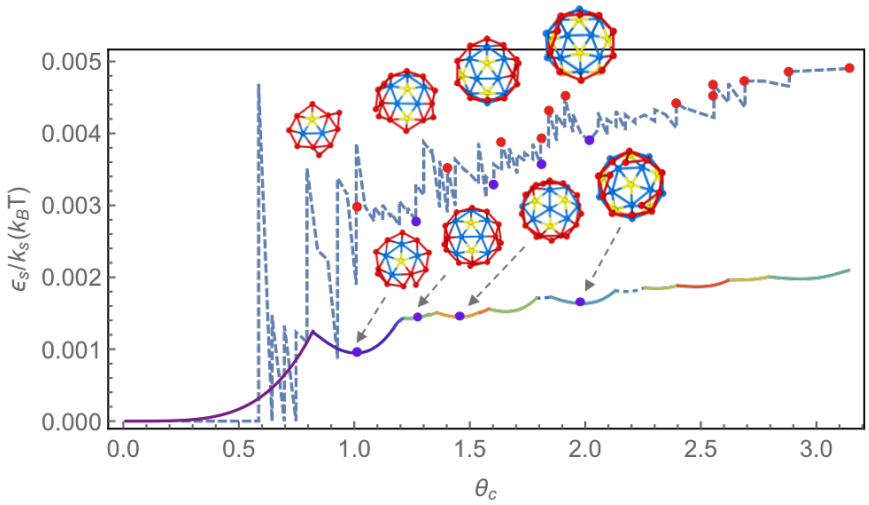

Figure A.1: The elastic energy of a spherical cap as a function of θ_c, which is related to the cap $area = 2\pi R^2(1 - \cos\theta_c)$. The formation of each new pentamer is marked with a red dot. Some snapshots of simulations are shown above the dashed line. Each pentamer first forms at the cap boundary and then the shell grows around it. After relaxation, the pentamers have moved towards the center of the cap. The relaxed configurations are shown below the dashed line and are marked with purple dots. The solid line shows the ground state energy obtained through continuum elastic theory, and the purple dots illustrate the same configurations as the ones found in simulations. There is a perfect match between the structures obtained in the simulations and continuum theory. The parameters used for the simulations are the spring modulus $k_s = 200$, bending rigidity $k_b = 200$, hydrophobic interaction $\epsilon_{hp} = -1.6$ and chemical potential $\mu = -14.6$. The final structure has T=3 icosahedral symmetry.

A.2 Disordered to Ordered Transition and the Role of Elastic Energy

Two examples of -disorder-order transition can be seen in Fig. A.2.

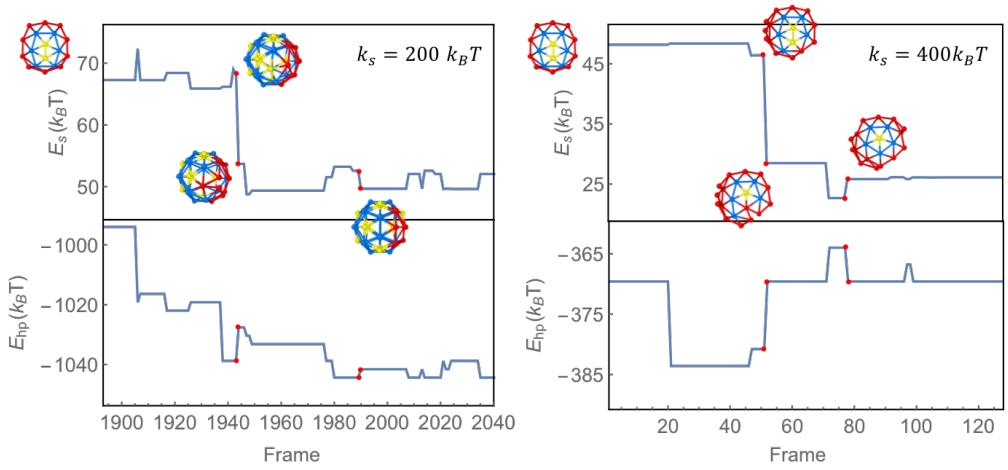

Figure A.2: Two examples of disorder-order transitions with different spring constants $k_s = 200$ and 400. The distribution of pentamers in ordered structures possesses subgroup symmetries of icosahedron. The red dots illustrate the step in which a pentamer formed in a "wrong" position dissociates and becomes a hexamer. For $k_s = 200$, this occurs after 1990 MC steps when the capsid has 56 trimers. The higher elastic energy, the faster pentamers transforms into hexamers if formed in "wrong" positions. For $k_s = 400$, it takes 78 steps for a pentamer to move and "correct" its location such that the cap has only 25 trimers when the pentamer formed in the "wrong" position dissociates. We note that each MC move involves any attempt to move any trimer or vertex through diffusion, growth or detachment. The simulations are performed for the bending rigidity $k_b = 200$, hydrophobic interaction $\epsilon_{hp} = -1.4$, and chemical potential $\mu = -14.6$.

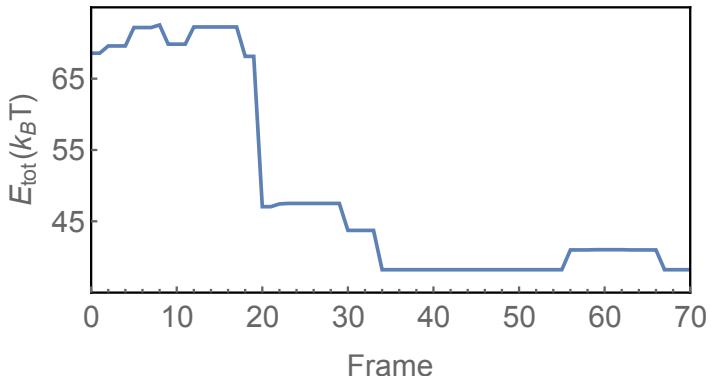

Figure A.3: The plot of the total energy (Eq. A.4) vs. Frame (number of MC moves). The elastic energy and hydrophobic energy of the system are shown in figure 4.4. The stretching rigidity is $k_s = 800$ and bending rigidity $k_b = 200$. Hydrophobic interaction $\epsilon_{hp} = -1.4$ and chemical potential $\mu = -14.6$.

A.3 Hydrophobic Interaction

The total hydrophobic energy of a capsid is

$$E_{hp} = \sum_{v_i} \epsilon_{hp}[nt_{v_i} \cdot (nt_{v_i} - 1)], \tag{A.1}$$

with v_i the vertex index, ϵ_{hp} the strength of the monomer-monomer interaction and nt_{v_i} the number of triangles sharing the vertex v_i. For a complete shell with Q triangles, the total number vertices is $N_v = 0.5Q + 2$ with twelve pentamers and the rest hexamers. The total hydrophobic energy of a complete shell based on Eq. A.1 can thus be written as $E_{hp} = 12\epsilon_{hp} \times 20 + (0.5Q + 2 - 12)\epsilon_{hp} \times 30 = 15\epsilon_{hp}(Q - 4)$. The hydrophobic interaction per trimer subunit is then equal to $\epsilon = 15\epsilon_{hp}\frac{Q-4}{Q}$. For a T=3 capsid with $Q = 60$ trimers, $\epsilon = 14\epsilon_{hp}$. When $Q \to \infty$, almost every triangle is part of a hexamer with 15 monomer-monomer interactions, and as such the hydrophobic interaction then converges to $15\epsilon_{hp}$ per subunit. For an incomplete shell, the vertices are categorized as the pentameric vertex

v_5, the hexameric vertex v_6 and the edge vertex v_e. The total number of trimers N_T then statisfy the constraint $3N_T = 5v_5 + 6v_6 + 3v_e$. Following the results of Ref.[60, 61], we assume that pentamers grow linearly with the capsid area, so the pentameric vertex number is $v_5 = \frac{12N_T}{Q}$. The edge vertex number v_e and the perimeter l_p are related through $v_e = \frac{l_p}{l_0}$ with l_p equal to

$$\begin{aligned} l_p &= 2\pi R \sin\theta_c \\ &= 2\pi R\sqrt{1-(1-A/(2\pi R^2))^2} \\ &= 2\sqrt{\pi}\sqrt{A - A^2/4\pi R^2} \\ &= 2\sqrt{\pi}\sqrt{N_T a - (N_T a)^2/(Qa)} \\ &= 2\sqrt{\pi}\sqrt{a}\sqrt{\frac{N_T}{Q}(Q-N_T)}, \end{aligned} \tag{A.2}$$

and $a = \frac{\sqrt{3}}{4}l_0^2$ the area of one trimer. l_0 is the length of each trimer side. The number of edge vertices and hexameric vertices then becomes $v_e = \sqrt{\sqrt{3}\pi}\sqrt{\frac{N_T}{Q}(Q-N_T)}$ and $v_6 = \frac{1}{6}(3N_T - 5v_5 - 3v_e)$, respectively. Finally we can obtain the hydrophobic energy of a partially formed shell by assuming each vertex at the edge on average has three neighbors. The total hydrophobic energy of an incomplete shell will then be

$$\begin{aligned} E_{hp} &= \epsilon_{hp}(5\times 4v_5 + 6\times 5v_6 + 3\times 2v_e) \\ &= \epsilon_{hp}\left(20v_5 + 5(3N_T - 5v_5 - 3v_e) + 6v_e\right) \\ &= \epsilon_{hp}(15N_T - 5v_5 - 9v_e) \\ &= \epsilon_{hp}\left(15N_T - \frac{60N_T}{Q} - 9\sqrt{\sqrt{3}\pi}\sqrt{\frac{N_T}{Q}(Q-N_T)}\right) \\ &= 15\epsilon_{hp}\frac{Q-4}{Q}N_T - 15\epsilon_{hp}\sqrt{\alpha}\sqrt{\frac{N_T}{Q}(Q-N_T)} \\ &= \epsilon N_T + \epsilon_l, \end{aligned} \tag{A.3}$$

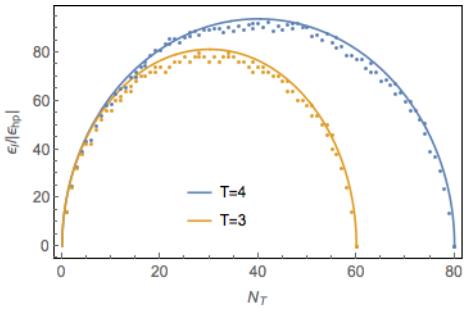

Figure A.4: Plots of line tensions vs. number of subunits N_T. The figure compares the line tension ϵ_l obtained from Eq. A.3 (solid lines) and the results of simulations (dots) for T=3 ($Q = 60$) and T=4 ($Q = 80$) capsids. $T = 4$ structure has higher line tension.

with $\alpha = \frac{9\sqrt{3}\pi}{25}$ and $\epsilon = 15\epsilon_{hp}\frac{Q-4}{Q}$. The quantity ϵ_l is called the line tension in classical nucleation theory. The comparison of the line tension calculated through Eq. A.3 and obtained through our simulations is illustrated in Fig. A.4 for $T = 3$ and $T = 4$ capsids with $Q = 60$ and 80. We find a very good match between the results of the simulations (dotted lines) and the simple theory (Eq. A.3) presented above (dashed lines).

A.4 Total Energy

As explained in chapter 4, the total energy of an empty capsid can be written as

$$E_{tot} = E_{el} + E_{hp} - N_T\ \mu \tag{A.4}$$

Substituting Eq. A.3 in Eq. A.4, the free energy of the growing capsid becomes

$$\begin{aligned} \frac{E_{tot}}{|\epsilon|} &= \frac{\epsilon_{el} + \epsilon - \mu}{|\epsilon|} N_T + \sqrt{\alpha}\frac{Q}{Q-4}\sqrt{\frac{N_T}{Q}(Q - N_T)} \\ &= AN_T + a\sqrt{N_T(Q - N_T)}, \end{aligned} \tag{A.5}$$

where $a = \frac{\sqrt{\alpha Q}}{Q-4}$ is a geometric factor and $A = \frac{\epsilon_{el}+\epsilon-\mu}{|\epsilon|}$ measures the difference between the chemical potentials of free subunits in solution and in full capsids.

The height of energy barrier and the nucleation size in Eq. 4.8 in chapter 4 can also be written as follows

$$\begin{aligned} N_T^* &= \frac{1}{2}Q\left(1+\frac{\Gamma}{\sqrt{\Gamma^2+1}}\right) \\ {E_{tot}}^* &= \frac{1}{2}|\epsilon|aQ\left(\Gamma+\sqrt{\Gamma^2+1}\right), \end{aligned} \tag{A.6}$$

where $\Gamma = \frac{A}{a}$ is the supersaturation.

A.5 Experimental Methods

A.5.1 Sample Preparation

70 g of infected leaf materials were blended with 150 mL of 0.15 M sodium acetate pH 4.8 and 150 mL of ice-cold chloroform. After centrifugation at 10,000× g for 10 min, the virions in the supernatant were precipitated by adding NaCl to a final concentration of 20 mM along with 8% (v.w^{-1}) poly(ethylene glycol) (MW 8000). The solution was centrifuged at 10,000× g for 10 min and the pellet was resuspended in 50 mM sodium acetate pH 4.8. After centrifugation at 8,000× g for 10 min, the supernatant containing virions was ultracentrifuged through a 20% (v.w^{-1}) sucrose cushion in water at 150,000× g for 2 h, and the virions in the pellet were stored at -80 °C until use.

For protein purification, 5 mg of virions were dialyzed against 50 mM Tris-HCl pH 7.5, 0.5 $CaCl_2$, 1 mM ethylenediaminetetraacetic acid (EDTA) pH 8.0, 1 mM phenylmethylsulfonyl fluoride (PMSF), 1 mM dithiothreitol (DTT), and centrifuged at 150,000× g for 18

h. Proteins were collected in the supernatant and stored in 50 mM sodium acetate pH 4.8, 0.5 M NaCl, 1 mM EDTA pH 8.0 at 4 °C until use. Full viral RNA genome comprising the four RNA segments encoding for the complete CCMV genome was purified by mixing 0.1 mL of virions with 1 mL of TRIzol reagent (Life Technologies, France). After addition of 0.2 mL of chloroform, the solution was centrifuged at 12,000× g for 15 min and the aqueous phase was removed prior to the addition of 0.5 mL of isopropanol. After centrifugation at 12,000× g for 15 min, the pellet was washed with 1 mL of 75% ethanol and centrifuged at 7,500× g for 5 min. The RNA pellet was resuspended in RNase-free water and stored at -80 °C until use. CCMV RNA 2 was transcribed in vitro as follows: The plasmid coding for the RNA was transformed in NEB 10-β competent E. coli cells (New England Biolabs, MA, USA), produced and purified by a NucleoBond® Xtra plasmid DNA purification kit (Macherey-Nagel, Germany), and linearized using the XbaI restriction enzyme. The linearized plasmid containing a T7 promoter was then transcribed using a MEGAscript Kit (Thermo Fisher Scientific, MA, USA). Newly produced RNAs were finally purified using a MEGAclear Transcription Clean-Up Kit (Thermo Fisher Scientific, MA, USA). Purity was checked by running an agarose gel electrophoresis and RNA was stored in RNase-free water at -80 °C. Capsid proteins and RNA were checked by spectrophotometry, and verified $A_{280}/A_{260} > 1.65$ and $A_{260}/A_{280} > 1.8$, respectively.

Nucleoprotein complexes (NPCs) were assembled as follows: Capsid proteins were initially dialyzed against capsid disassembly buffer, i.e., 50 mM Tris-HCl pH 7.5, 450 mM NaCl, 1 mM EDTA pH 8.0, 1 mM PMSF and 1 mM DTT. For the measurements with the full genome, capsid proteins and RNA were separately dialyzed against NPC buffer, i.e., 50

mM Tris-HCl pH 7.5, 50 mM NaCl and 1 mM EDTA pH 8.0, and rapidly mixed together with a stopped-flow device at the desired final concentrations. For the measurements with in vitro transcribed CCMV RNA 2, RNA was dialyzed against capsid disassembly buffer, then mixed with capsid proteins at the desired final concentration, and the mixture was dialyzed against NPC buffer overnight.

A.5.2 Small-Angle X-ray Scattering (SAXS)

Scattering patterns were recorded at the ID02 and SWING beamlines of the European Synchrotron Radiation Facility (Grenoble, France) and the SOLEIL synchrotron (Saint-Aubin, France), respectively. Between 10 and 100 frames were collected for each sample with a beam exposure time set to 10 ms. The two-dimensional scattering images were radially averaged and the intensities were converted into absolute units after subtraction of the contribution of the buffer solutions. Ab initio shape reconstructions were performed by using the programs DAMMIF [34] and GNOM [92]. For each shape reconstruction, 20 models were averaged then filtered, and the final structures were rendered with Chimera [82]. Sphericity indexes were computed through a principal component analysis. Let a, b and c be the standard deviations in descending order along the principal axes of a reconstructed structure, the sphericity index is defined as $[c^2/(ab)]^{1/3}$. The sphericity index is positive and tends to one for a sphere.

A.5.3 Cryo-Transmission Electron Microscopy

The sample was frozen on a holey carbon grid (Quantifoil R2/2) by using a FEI Vitrobot and observed with a JEOL JEM-2010 electron microscope equipped with a 200-kV

field emission gun. Images were collected with a Gatan Ultrascan 4K CCD camera with a $\times$50,000-magnification using a minimal dose system. Defocus was set to 2.5 μm.

A.6 Polydisperse Vesicle Model

Figure A.5 depicts the scattering patterns of various NPC samples fitted with a polydisperse vesicle model. This model is not intended to represent realistically the structure of the NPCs but rather to give a rough measure of their size and shape uniformity. The fitted patterns of samples **I**, **II** and **III**, made up with the full CCMV genome, show that upon increasing the subunit concentration from 0.5 to 2.1 g.L^{-1}, the polydispersity index $\Delta R/R$ decreases from 0.56 to 0.35, which suggests an increasing degree of order in the structure of the NPCs, in good agreement with the simulated disorder-order transition. Note that the subunit-to-RNA mass ratio is set to 6, that is, well larger than the value of 4 found in native virions; in other words, there are in principle always more than enough subunits to build up a closed capsid with genomic RNA inside. For NPCs made with only in vitro transcribed RNA 2 (sample **IV**), the degree of order looks even higher since the polydispersity index drops to 0.21, which can be ascribed to a lesser variability of the NPC structure than in the presence of multiple RNA segments.

Table A.1: Sample and fitting parameters for the SAXS patterns on Fig. A.5 using a polydisperse vesicle model.

Sample	Subunit concentration (g.L^{-1})	Subunit-to-RNA mass ratio	R (nm)	$\Delta R/R$
I	0.5	6	40	0.56
II	1.0	6	52	0.41
III	2.1	6	46	0.35
IV	2.1	6	62	0.21

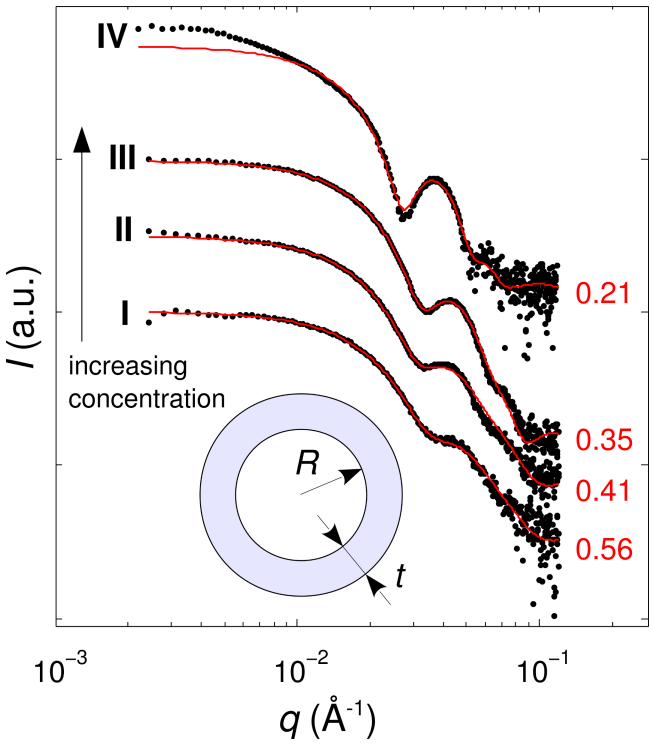

Figure A.5: SAXS patterns of NPCs (black dots) at neutral pH, fitted with a polydisperse vesicle model (red line). The model consists of homogeneous vesicles (see the drawing) with a fixed shell thickness t and radii distributed normally about R with a standard deviation ΔR. The numbers in red next to the curves are the polydispersity indexes $\Delta R/R$. NPCs are made of the full RNA genome (**I**, **II** and **III**) with subunit concentrations of 0.5, 1.0 and 2.1 g.L^{-1}, respectively, and of in vitro transcribed RNA 2 (**IV**) at a subunit concentration of 2.1 g.L^{-1}. In all cases, the subunit-to-RNA mass ratio is about 6. The scattering curves are shifted for clarity. The fitting parameters are given in Table A.1.

A.7 Disorder-Order Transition at Acidic pH

The disorder-order transition of CCMV NPCs upon a decrease of pH was reported in previous studies [36, 38, 14]. Fig. A.6 illustrates the effect of such a transition on the form factors (i.e., the normalized scattering intensities) of NPCs at neutral and acidic pH. Two peaks are clearly visible around 0.045 Å^{-1} and 0.07 Å^{-1} at pH 5.2, which is characteristic of the long-range order of spherically-symmetric objects. A direct comparison with the form factor of purified virions indicates that the size and the overall structure of NPCs at acidic pH are akin to those of native virions.

A.8 Nanotubes

Nanotubes are seen at high concentrations of subunits and in vitro transcribed RNA 2 (Fig. A.7). These nanotubes are probably formed by the excess of free subunits that do not bind on RNA. The presence of nanotubes at neutral pH is consistent with the phase diagram of CCMV CPs [55].

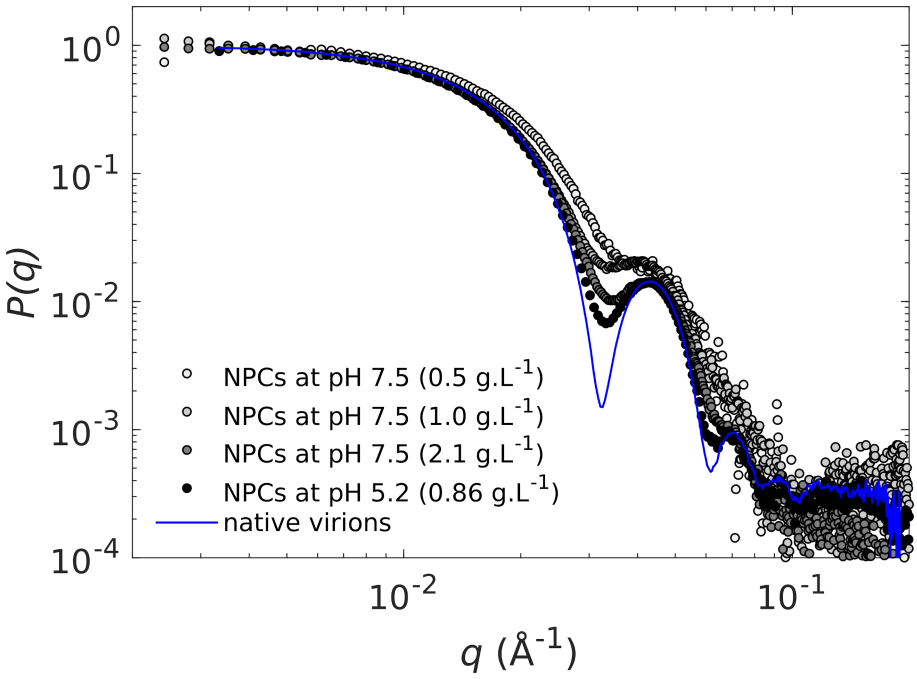

Figure A.6: SAXS form factors $P(q)$ of native virions and NPCs at two different pH. The NPCs at pH 7.5 (light gray, gray and dark gray discs) correspond to the samples **I**, **II** and **III** with the full RNA genome and a subunit concentration of 0.5 g.L^{-1}, 1.0 g.L^{-1} and 2.1 g.L^{-1}, respectively. The NPCs at pH 5.2 (black discs) were formed with the full RNA genome at a subunit concentration of 0.86 g.L^{-1}, a subunit-to-RNA mass ratio of 5.5, and an ionic strength of 55 mM. The curve was collected 4970 s after mixing NPCs initially at pH 7.5 with an acidic buffer. The form factor of native virions (blue line) was measured from a solution of purified virions in a buffer solution at pH 4.8 and ionic strength of 50 mM. The two latter curves are adapted from [14].

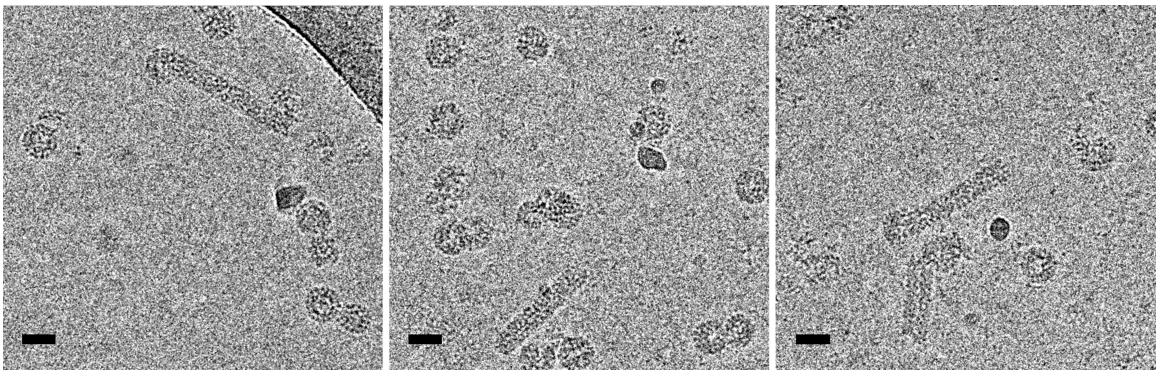

Figure A.7: Cryo-transmission electron microscopy images of nanotubes coexisting with spherical NPCs at neutral pH. The sample contains 3.9 g.L^{-1} of subunits along with in vitro transcribed RNA 2 at a subunit-to-RNA mass ratio of 6. Scale bars are 30 nm.

A.9 Simulations with Smaller Cores

Many experiments show that the capsids assembled around PSS are smaller [14]. To this end, we construct a phase diagram (Fig. A.8) as a function of ϵ_{hp} and μ, similar to the one in Fig. 4.11 in the chapter 4. We use en masse simulations with small cores($R_0 = 1.2$), comparable to the capsids filled with PSS. The smaller structure looks like a tennis ball, previously observed in the self-assembly studies of clathrin shells [32, 99].

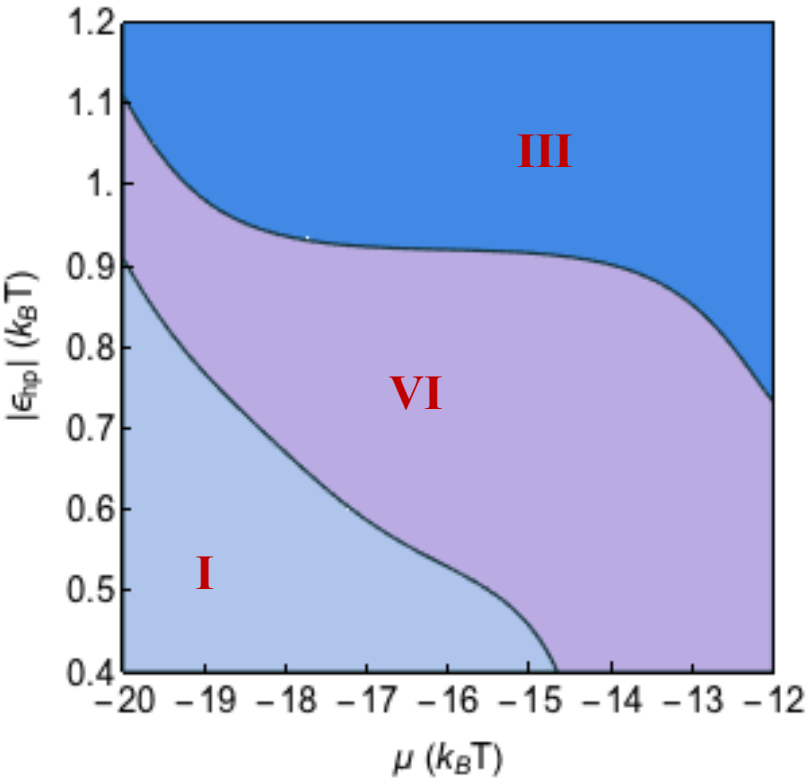

Figure A.8: Phase diagram of structures obtained from en masse simulations in the presence of smaller core size comparing to Fig. 4.11 in chapter 4. The blue shade corresponds to the region in which closed capsids, mostly with tennis ball symmetry form but the purple one represents the phase where amorphous structures form. The light blue region corresponds to region where no shell forms. Parameters that are used in this simulations are $R_0 = 1.2$, $\epsilon_{lj} = 11.2k_BT$, k_s=$600k_BT$ and k_b=$200k_BT$.