Structure, Mechanical Properties, and Self-Assembly of Viral Capsids

Antoni Luque Santolaria

Structure, Mechanical Properties, and Self-Assembly of Viral Capsids

Antoni Luque Santolaria

Departament de Física Fonamental
Facultat de Física
Universitat de Barcelona

Facultat de Física

Departament de Física Fonamental

Structure, Mechanical Properties, and Self-Assembly of Viral Capsids

Programa de doctorat: *Física*

Línia d'investigació: Física de la matèria condensada

Director de tesi: David Reguera López

Memoria presentada per Antoni Luque Santolaria per a optar al títol de Doctor en Física (8 de Juny de 2011).

Agraïments

Probablement, aquestes siguin les línies més complicades d'escriure de tota la tesi. Fer el doctorat ha estat una etapa d'una significació molt important per a mi, i sortosament l'he poguda compartir amb gent molt especial, tant a dins com a fora de la universitat.

El meu primer agraïment va inevitablement dirigit cap al David Reguera. De la seva mà he tingut el privilegi d'iniciar-me en el món de la recerca i crèixer dia a dia com a investigador, gaudint d'una llibertat envejable, i a l'hora d'una atenció i disponibilitat per part seva excepcionals. A més vull donar-li les gràcies per tot el suport que m'ha donat durant tot el doctorat, especialment en aquesta darrera etapa de la tesi.

Dono les gràcies també a tot el personal del departament, per la seva actitud i cordialitat al llarg d'aquest període. En particular volia agrair a tots aquells doctorands, postdocts i professors (tant del Departament de Física Fonamental com del d'Estructura i Constituents de la Matèria) que, a partir del seu interés i participació, han fet possible els *Condensed Matter Seminars*. Sé que a vegades he estat una mica pesat amb els anuncis de les xerrades, però crec que ha valgut la pena i ha estat una activitat enriquidora. També volia destacar els valuosos consells que he rebut en diferents moments per part del Fèlix Ritort, la Marta Ibañes, el Jordi Ortín, el Javier Buceta i l'Ignasi Pagonabarraga. Gràcies també a l'Eduard, l'Helena, l'Olga, i, en especial, a la Bea, per la paciència i atenció que sempre han tingut amb mi quan he hagut de fer gestions al departament.

Així mateix, vull destacar el gran record que guardo de la meva experiència a Los Angeles al grup d'en Robijn Bruisnma. Gràcies Robijn per la teva generositat. I gràcies també Joseph Rudnick, Howard Reiss, Will Klug, i el "virus research groupâ de UCLA (en particular: Bill Gelbart, Chuck Knobler, Ajay Gopal, Walter, i Mauricio Comas) per la vostra gran acollida. Tots vosaltres vau contribuir a que la meva estada científica a UCLA fos excepcional.

Gràcies també al grup de Pedro de Pablo i la gent del CNB de Madrid (JR, Daniel, Carmen, i companyia) per la vostra hospitalitat i col·laboració durant la

tesi.

He de remarcar que la recerca duta a terme en aquesta tesi no hauria estat possible sense el suport econòmic que he rebut al llarg d'aquests anys per part de la Generalitat de Catalunya a través de l'AGAUR, que sempre m'ha ofert, a més, un tracte excel·lent.

No puc oblidar tampoc aquells que no només hem caminat plegats durant el doctorat, sinó que, a més, vam començar la singladura d'aquest cim el primer dia de la llicenciatura: el David "ay" López, el Camps, el Pau, la Belén, el Tavi, el Capde, el Ricard, l'Álvaro, l'Escartín, l'Albert talladit, el Richi, el Tarrús... sens dubte aquesta ha estat una promoció, i sobretot, un grup humà excepcional. Ha estat un plaer haver passat aquests anys amb vosaltres, i és una llàstima que no hàgim tornat a visitar Berlín! Em perdonareu, però al Pau, l'hi he dedicar unes paraules a part. No exagero si dic que ell és un dels responsables que jo ara estigui escrivint aquestes línies. La seva amistat a l'institut va ser un punt d'inflexió no només en el pla personal, sinó també en el vocacional. I així hem acabat: parlant de gens i proteïnes entre "hamiltonians".

A la Noela, el Ricard, el Silvio, el Paolo, i el recent emigrat Dr. Camps, us he de donar les gràcies per tots els bons moments que hem compartit al 3.45, i també demanar perdó per totes les hores de feina que us he robat parlant del Barça, d'*scripts* de linux, o de qualsevol anècdota no excessivament transcendent. Molts ànims Noela en aquest tram final de l'escriptura! Espero que ben aviat puguis tornar gaudir relaxadament del sol i un bon llibre al balcó de Gràcia. Per cert, Ricard, això s'acaba i al final no hem aconseguit muntar cap empresa. Vull mencionar també als recentment doctorats Álvaro i David que tantes hores de tertúlia hem compartit a les escales, així com als "cafeters" Nico, Martin, María i companyia per tots els dinars que hem passat al "ranxo". De fet, la María mereix una menció especial perquè m'ha d'aguantar als "groups meetings", i tot i que em posi pesat quan treballem plegats (com en l'estudi del Capítol 4 d'aquesta tesi), sempre està de bon humor i amb ànims de fer una canya! Xavi, perdona, la tesi no m'ha deixat venir a Florència, sempre em passa igual... Javier, gràcies de nou per la teva generositat oferint-me la teva habitació "santuari" de Madrid, i molts ànims en aquesta nova etapa com a divulgador. Doro i Jan, gràcies per la vostra amistat, i tots els consells i el suport que m'heu donat els pocs dies que ens hem pogut veure aquests mesos.

No vull oblidar tampoc a la Lynne, el Jesse, la Sasha i el Marshall que em van adoptar a la seva família durant la meva estada a Los Angeles, i que juntament amb el Mauricio, la Veta, la Charlotte, l'Ashley, el Peter, el Noah, l'Allegra, la Shellie, el Jon, la Laura, el Joan i la penya culé del Busby's em van fer sentir

com a casa, i viure uns mesos fabulosos.

Una abraçada gegant als de sempre: Albert, Joan, Diego, Àlex i Tolo. Espero que brindem aviat fent un vermut a la Barceloneta, en un concert d'en Jortx el "ráfagas", o en el "dancefloor" de la Gaviota amb el Nacho, el Txepe, els peluts, i la penya de les Corts, si cal. Aquests darrers mesos us he trobat a faltar. I don't like you, I love you!

Gràcies papa i Montse per aquest regal anomenat Aitana. Gràcies iaia, Laura i mama pel vostre suport incondicional durant tota la vida. I gràcies *my Butzie* pel present i el futur que tenim entre mans.

A ma mare

Contents

Introduction

In our globalized world, where information and diseases can spread quickly, many tend to regard viruses as a threat to humanity. The truth is that, in fact, viruses have had an important impact throughout history. For instance, the influenza of 1918 (*Spanish Flu*) was an especially severe strain that, combined with the devastating situation of many countries after World War I, resulted in a pandemic that took over 40 millions of lives (more than WWI itself). Other *tragic* examples can also be mentioned like the virulent, but almost eradicated, smallpox, or the more recent cases of Ebola and AIDS [1, 2]. However, viruses are more than just parasites, since they play a major role in the balance of ecosystems [3], and recently humanity is starting to take advantage of their *abilities*, using them as insecticides [4], as an alternative to antibiotics [5], or even in more sophisticated applications, like gene therapy [6].

Viruses are submicroscopic biological entities that need to infect a host cell in order to replicate. They have probably existed since the origin of life. But, they do not fossilize well, so *paleovirology* is only starting to track their history using modern genomic techniques and strategies [7]. In any case, historical records from ancient Egypt and Greece, centuries B.C, indicate that nowadays viruses like smallpox virus, polio virus, or rabies virus were already present at that time. However, the development of rationalized methods to fight these diseases took longer. The first records of primitive types of vaccination (variolation) come from China and India in the 11th century. But it was not until the 1790s when Edward Jenner used cowpox, a mild version of smallpox, as a vaccine for smallpox in Europe, a method that was improved and generalized in 1885 by Louis Pasteur. Curiously, all this progress was achieved without knowing the actual agents of such diseases. Dimitrii Ivanovsky was the first to report in 1892 that the causative agent of tobacco mosaic disease was not retained by the filters that usually removed bacteria from extracts and cultures. Subsequently, in 1898 Martinus Beijerinck made the same observations independently, and, more importantly, made the conceptual leap of identifying the responsible party for

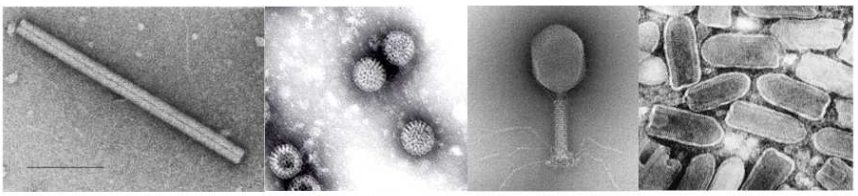


Figure 1: Electron micrographs of viruses. (a) The helical rod-like particle of tobacco mosaic virus (TMV) [10, 11]. (b) Icosahedral human rotavirus particles [12]. (c) Elongated bacteriophage T4, with tail and fibers [13]. (d) Bullet-shaped particles of the rhabdovirus vesicular stomatitis virus [14]

the disease as a distinctive agent, smaller than any known bacteria. These new entities were called ultrafilterable viruses, and eventually simplified to viruses[1] [1].

The structure and composition of viruses started to be unveiled in 1935, when Wendell Stanley crystallized tobacco mosaic virus (TMV) [8], a process facilitated by its structural simplicity. In addition, the introduction of the electron microscope (EM) in the 1930s rapidly revolutionized virology. The power of amplification of EM ($\times 100,000$-fold) confirmed the relatively simple architecture of viruses (see Fig. 1), and paved the way to the first rational classification, based on their morphology. Later on, studies on bacteriophages[2] (or phages), by Max Delbrück and coworkers, showed that phages are stable, self-replicating entities characterized by heritable traits. Remarkably, these features turned out to be common for all viruses. Moreover, in 1952 Alfred Hershey and Martha Chase [9] showed that among the two main components of viral particles, i.e., protein and nucleic acid, the latter corresponded to the genetic material of the virus, and was ultimately responsible for its molecular replication in the host cell. At this point, the basic aspects of viral particles were identified. But, rather than being the endpoint of virology, it gave rise to an outbreak of a field that has many questions to solve and is still growing today.

Interestingly, the study of viruses, due to their relative simplicity in the biological scale and their interrelation with the host cells, has led to important breakthroughs in other fields of biology. In particular, the research on bacteriophages was crucial to establish the foundations of molecular biology. For instance, in 1961 S. Brenner, F. Jacob and M. Meselson discovered messenger

[1]The term *virus* came from the Latin word for *poison* or *venom*.
[2]Viruses that infect bacteria.

RNA (mRNA) [15], the genetic blueprint for protein production in the cells; and, F. H. Crick, L. Barnett, S. Brenner, and R. J. Watts-Tobin elucidated that the information of DNA was encoded in *codons* of three nucleotide basis [16]. Furthermore, the research on viruses also led to important landmarks in cellular biology. For example, the study of oncoviruses, which can cause cancer, like Rous sarcoma virus in chickens, eventually unveiled the molecular basis of cancer. A clear demonstration of the scientific importance of the discoveries and advances derived from the research on viruses is the fact that they have been recognized by multiple Nobel Prizes [1].

Nowadays, it is well known that viruses infect all type of organisms, from bacteria to mammals. Surprisingly, some of them, called *virophages*, infect even other viruses, once these have taken control of the cell [17]. Thus, viruses represent one of the most clear examples of the selfish gene concept [18]: a trait that perpetuates over time by replication. Despite the fact that their *reproduction* strategy could seem *primitive* and *unsophisticated*, viruses are the most abundant biological entity on the planet [3, 19], and they play a fundamental role in regulating ecosystems, especially in marine environments where microbial communities are abundant and essential for the nutrient cycling [3,20]. Interestingly, viruses show a rich spectrum of replication strategies that sometimes do not require killing or even harming the host. Indeed, *cryptic viruses* replicate in organisms without causing any apparent symptoms or disease [21]. Remarkably, some viruses incorporate their genome into the host and can remain latent for a certain period of time, like the well known λ phage or herpes virus [1]. Indeed, viral sequences can be integrated in the organism genome even permanently, becoming a part of it [22].

A viral species is usually specialized in the infection of only one type of organism (or few related ones), and shows well defined morphological and molecular traits. These characteristic traits can vary greatly among different viral species, but there are two ingredients that are always present in virions[3] [1]: the infective genetic material, and the protective protein shell, called the *capsid* (see Fig. 2).

All viruses carry a piece of genetic material that contains the necessary information to hijack the molecular machinery of the cell, and to produce new viral particles under the right cellular conditions. The virus genome can be encoded in the traditional double-stranded DNA (dsDNA), but also in other type of polynucleotides, such as single-stranded DNA (ssDNA), single-stranded RNA (ssRNA), or even double-stranded RNA (dsRNA) [1]. The pathway of replication in the host cell depends very much on the type of genome. For instance, many dsDNA

[3] A virion is an infectious virus particle.

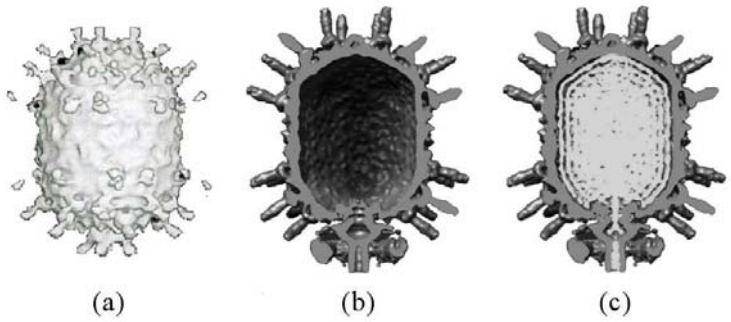


Figure 2: Essential components of a virus: the capsid and the genetic material. (a) Cryo-electron microscopy of the exterior of the bacteriophage $\phi 29$ capsid. Two slices showing the interior of an empty capsid (b) and a capsid full of dsDNA (c). Pictures from M. G. Rossmann's Lab. (Purdue University).

viruses, like herpes virus, replicate their genome using a host polymerase; instead, negative-sense ssRNA virions, e.g., rabies virus, require a viral polymerase to transcribe first the genome into the reciprocal positive-sense ssRNA that is then *readable* for the cell. In fact, the Baltimore classification groups viruses into families depending on their type of genome and method of replication (see Fig. 3). Remarkably, the origin of the viral genomes types and their methods of replication are still in debate, and it has not been ruled out that the viral strategies of infection could even play an important role in the emergence of the actual three domains of life, i.e., bacteria, archaea, and eukaryote, or in their genetic mechanisms [23–25].

In any case, nucleic acids are usually very sensitive to environmental conditions, and easily degradable. Thus, between the formation of the viral particle and the infection of a new host, viruses rely on the capsid to protect their genomes. Viral capsids have sizes from tens to several hundreds of nanometers and are made of multiple copies of one or a few different proteins that spontaneously self-assemble, without consumption of chemical energy (in terms of ATP). Given the limited size of the viral genome, this strategy is much more efficient than encoding just one *huge* single macromolecule. The coat proteins from which they are formed have a few nanometers in size, and, despite having completely disparate sequences for different viruses, they end up building a set of common architectures for the capsid. In fact, it has been recently shown that viral coat proteins form only few structural lineages. Thus, viruses that are apparently unrelated can be grouped by the structural similarities in their coat proteins, and this idea

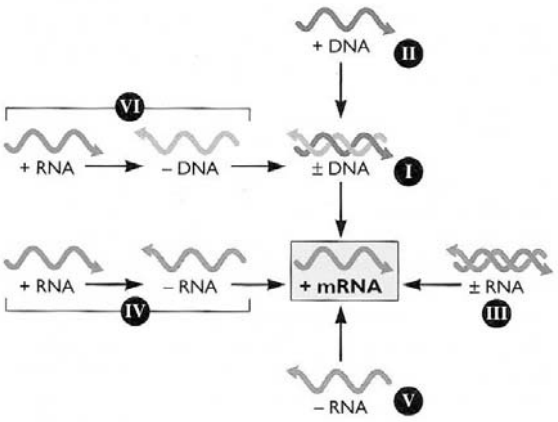


Figure 3: The Baltimore classification (extracted from Ref. [1]). All viruses must generate messenger RNA (mRNA), which is required for protein production in a cell. The different types of viruses in this classification are grouped in classes indicated by the Roman numbers, and defined by the pathways taken to obtain the mRNA, which are mainly determined by the nature and polarity of the viral genomes.

is having an important impact in viral evolution [26] (see Fig. 4). Typically, viral capsids come in different shapes that tend to be unique for each virus *in vivo* (see Fig. 1), but that can be changed by the assembly conditions *in vitro*. The first virion observed under the electron microscope, tobacco mosaic virus (TMV), showed a rod-like structure with coat proteins organized in a helical sheath, opened at both ends [11] (see Fig. 1a). However, most known viruses show closed structures instead. In particular, quasi-spherical capsids seem to be the most abundant, and despite their variability in sizes, their coat proteins are usually organized in a very precise way following icosahedral symmetry [27–29] (see Fig. 1b). The geometrical rules of construction of these quasi-spherical viruses were established by Caspar and Klug in the early sixties [28], and since then they have served as a structural classification, and as the basis of modern structural virology. Nevertheless, there is an important number of viral structures that, despite being relevant, have not been completely characterized yet. These types of virions usually present bacilliform (or prolate) capsids, like many bacteria or plant viruses (see Figs. 1c), or conical shells, like the well known human immunodeficiency virus (HIV). Other less abundant viral structures have been also observed, like the bullet-shaped particles of some animal viruses (see

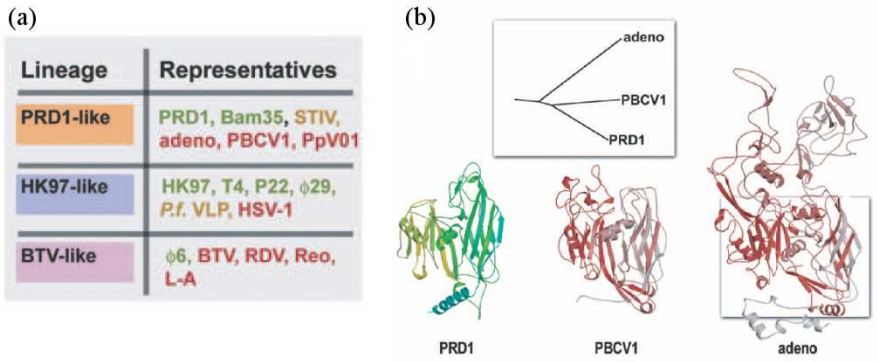


Figure 4: Structure-based viral lineages (extracted from Ref. [26]). (a) Table listing different lineages obtained comparing high-resolution structures of viral coat proteins. (b) Comparison of X-ray-based coat protein structures of PRD1-like viruses. The different colors indicate the domain of life of the respective hosts: green (bacteria), red (eukarya). The structures are presented in side view, where the lower end points to the virus interior. The core of the proteins in this lineage shows a characteristic jelly-roll made of two β sheets (boxed for adenovirus). The inset above is the relevant portion of the structure-based phylogenetic tree, calculated based on these structures.

Fig.1d), or the astonishing *zoo* of morphologies reported for several virions that infect extremophiles archaea [30]. The existence of these common set of structures and the underlying mechanisms responsible for their appearance is one of the many open questions in the field.

Besides the coat proteins and the genetic material, there are other structural elements present in many viral capsids. For instance, many bacteriophages have a proteic tail attached to the capsid that is crucial to inject the viral genome into the bacterial host [31]. Some virions have spikes located at specific locations on the shell (see Fig. 2), or in the tail (see Fig. 1c), to enhance the recognition of, and attachment to the host [32, 33]. Viral particles can also have phospholipid membranes. Sometimes they are placed between the genetic material and the protein capsid, most likely helping the assembly process or modifying the mechanical properties of the shell, like for most viruses in the PRD1-like structural lineage [26]. In other cases the membrane is an envelope involved in virus maturation, and in the infection of the new host, like for HIV [34]. On the other hand, some viral proteins act as transient scaffold during the capsid assembly, and

then are ejected during the maturation process, like for bacteriophage P22 [35]; alternatively, some proteins have an indirect structural role, for instance in the compaction of the genetic material in some dsDNA viruses [36]. These structural elements play a role in the viral *life* cycle, even though they are usually dispensable for the integrity of the viral shell.

A crucial aspect of viruses, which makes them fascinating from a biophysical point of view, is that, although they are nanomachines properly *programmed* to infect a host, viral particles do not have their own metabolism. This suggests that during their natural cycle, from the assembly to the infection stage, they must rely on general physical and chemical principles to succeed in most of their tasks, and to resist the different possible extreme physiological conditions of the environment. For example, at the right conditions, viral capsids assemble directly from their building blocks. This means that the interactions among the coat proteins (and also the scaffolding proteins of the genetic material if they are required) spontaneously drive the formation of shells in a process that could be in principle described by basic thermodynamics and kinetics [37–39]. Moreover, the physical properties of the viral genome are different depending on the type of nucleic acid, which has an important influence in the viral strategies of assembly and infection. In particular, ssRNA viruses usually coassemble from their genetic material and coat proteins [40–42]. Thus, it is possible to produce *in vitro* infective particles by mixing just these two constitutive elements at the right conditions, like in the paradigmatic case of TMV [40]. Instead, many dsDNA viruses first form the shell (the *procapsid*), and then pack the genome at high densities using a molecular motor [43]. The internal pressure generated during the process is generally used to initiate the subsequent infection of the host [31, 44].

Thus, in addition to the wealth of biological details that confer specificity to viruses, the study of the underlying physics involved in the assembly of viral particles, their stability, or the gene delivery process, is of great importance to understand the properties of viruses that are common among different species. This knowledge, beyond paving the way to understand how viruses function, may also lead to different useful applications. Viral particles have been subjected to natural selection for hundreds of millions of years, so they have found optimal solutions to perform their basic functions efficiently and that could guide the design of a new generation of nanostructures. In fact, the impressive properties of viral capsids make them good candidates for promising biomedical and nanotechnological applications. For example, adenovirus capsids are used for genetic therapy, replacing the infective viral genome for an alternative non-viral genetic material that encodes a protein of interest [6]; viral capsids have been labeled with fluores-

cent dyes to image the *in vivo* vascular system of organisms [45]; they have been also used as nanoreactors to confine enzymatic reactions [46]; or as scaffolds to template the production of symmetrical inorganic structures [47] with remarkable electrical properties as nanowires [48] or efficient batteries [49]. More importantly, the understanding of the common physical mechanisms that are behind the life-cycle of all viruses could also help the development of broad-spectrum antivirals that might interfere systematically with viruses that use similar strategies of assembly or gene delivery. Finally, those general physical principles governing viral capsids might also play an important role in other well defined biological structures, like gas vesicles [50], vault shells [51], or photosynthetic vesicles [52], and could be applied to produce artificial structures with similar properties [53].

Although there has been some progress in elucidating these underlying physical properties of viruses [54], the research is still in its initial stages. In particular, there are still many open questions regarding the origin of the morphologies adopted by viruses, their mechanical advantages, and the mechanisms that control their assembly, which will be addressed in the context of this thesis.

Scope of the Thesis

The purpose of the present thesis is to investigate the physical mechanisms involved in the formation and stability of viral capsids. The work has been divided into three main blocks, devoted to studying the structure, the mechanical properties, and the assembly of viruses. In each case we have tried to develop a general theoretical framework to understand the different phenomena, and to establish the basis for future investigations. Simulations have played a crucial role in justifying the theory and clarifying its connection with experiments, and have been the driving force of the thesis. A special effort has been placed on the application and interpretation of our findings in light of experimental information. In this sense, we have been able to collaborate with one experimental group in analyzing the mechanical properties of viruses.

Part I of the thesis is devoted to the study of the *Structure of Viral Capsids*. Viruses adopt highly symmetrical structures that can be described by precise geometrical principles. In particular, many viruses have a quasi-spherical capsid whose structure has been intensely investigated from both biological and physical standpoints. However, bacilliform or prolate viruses, despite being abundant, have been less explored. In this first part, we extend the physical and geometrical principles governing the structure of spherical viruses to understand the architecture of elongated capsids. The block is divided into two chapters.

In the first chapter (*Geometrical Construction of Viral Capsids*), we will introduce the geometrical principles involved in the construction of spherical and tubular viral shells. First, we will review the classical Caspar and Klug (CK) construction that classifies icosahedral quasi-spherical particles using the triangulation number T. Special emphasis is placed on the reorganization of the T-capsids in terms of classes, which groups shells sharing similar geometrical properties that in turn translate into their actual physical properties. Then, we develop a generalization of the CK construction for the less studied elongated capsids, for which a complete structural classification was missing. The analysis shows that only a limited set of tubular architectures can be built closed by hemispherical icosahedral caps. In particular, the length and number of proteins adopt a very special set of discrete values dictated by the axial symmetry (5-fold, 3-fold, or 2-fold) and the triangulation number of the caps. Finally, we show how these ideas can be used to unveil the structure of some tubular viruses, by using a few experimental inputs.

In the second chapter of Part I (*Optimal Structures for Spherical and Elongated Viral Capsids*), we show that the origin of the aforementioned symmetrical architectures arises from a fundamental physical principle: the free energy minimization. To this end, we introduce a simple phenomenological model that accounts for the basic ingredients of the interaction between the morphological units of capsids (the capsomers), allowing a systematic exploration of the optimal structures for spherical and elongated shells. In this way, one recovers the CK icosahedral capsids, and also the new architectures described for prolate capsids, confirming the geometrical rules derived before. Interestingly, we also find non-icosahedral structures, and "all-pentamer capsids", which are exceptions to the standard classification that have been observed experimentally. Thus, our simple model reproduces the architecture of spherical and bacilliform viruses both *in vivo* and *in vitro*, and combined with the geometrical constructions constitutes an important step towards the understanding of viral structure, and its potential control for biological and nanotechnological applications.

The second part of the thesis is dedicated to the analysis of the *Mechanical Properties of Viral Capsids*. During the virus *life* cycle, the mechanical properties of its capsid play an important role in several biological processes. For instance, to be infective, many viruses require a maturation stage where the capsid undergoes a buckling transition that tunes its resistance. In addition, capsids of several double-stranded DNA (dsDNA) viruses must withstand high pressures, arising from their densely packed genetic material, which are later on used to initiate the translocation of the DNA into the host. In this second block, we will discuss

some of these remarkable mechanical properties of viral capsids combining theory, simulations, and experiments.

In Chapter 3 (*Theoretical Insights in the Elasticity of Viral Shells*), we will introduce a simple theoretical framework connecting some elastic properties of viral capsids with the microscopic interactions among coat units. This is the basis to rationalize the study of the mechanical properties of viral shells, and also to infer their potential biological advantages.

In the second chapter of Part II (*Buckling and Maturation of Spherical Viruses*), we analyze the role of the organization of coat proteins in the buckling phenomenon observed in spherical capsids. This has been an elusive question in the context of continuum elasticity theory, and here we use our coarse-grained physical model, introduced in Part I, to show that capsid shape and buckling transition are strongly influenced by the icosahedral class P of the viral structure. In particular, we find that capsids of a specific class ($P = 1$) are the most likely to produce polyhedral shapes, minimizing their energy and accumulated stress. On the other hand, we also find that under expansion, spherical capsids always show a buckling transition to polyhedral shells, independently on the T-number, and in consonance with the maturation of many viral particles. Interestingly, the resulting polyhedral shaped-particles are mechanically stiffer, tolerate larger expansions, and withstands higher internal pressures before failing, a trait that might be of particular interest for pressurized dsDNA viruses.

Finally, in the third chapter of the block (*Built-in Stress in Bacteriophage* ϕ29), we investigate the unexpected mechanical properties of bacteriophage ϕ29, a prolate virus that packs dsDNA at high densities. The study was performed in collaboration with the experimental groups of P. J. de Pablo[4] and J. L. Carrascosa[5]. Their experiments, using atomic force microscopy (AFM), revealed that, for empty shells, the equatorial zone was stiffer than the caps, in contradiction with elasticity theory. However, simulations using our model show that the origin of this discrepancy comes from the presence of built-in stress in the structure. Interestingly, the effects of this mechanism are analogous to the pressurization of a bacilliform vessel. Hence, this reinforcement strategy of empty ϕ29 capsids could play a relevant role in accommodating the pressure stored by the dsDNA in the mature particle, which is crucial for the infectivity of the virus.

The last part of the thesis, titled *Self-assembly of Viral Capsids*, is devoted to the kinetics and thermodynamic aspects of the self-assembly of spherical viruses, and it is divided in two chapters.

[4]Universidad Autónoma de Madrid, Spain.
[5]Centro Nacional de Biotecnología - CSIC, Spain.

In the first (*Classical Nucleation Theory of Spherical Shells*), we introduce a general theoretical framework to understand and characterize the assembly of viral particles. In particular, we show that Classical Nucleation Theory (CNT) can be adapted to explain capsid formation, providing a solid thermodynamic and kinetic description of the viral assembly process. This theory consistently explains many features observed in the assembly and disassembly of viral particles, for instance, the steep concentration dependence, the lag time at the first stages of assembly, or the sigmoidal kinetics of capsid production. Furthermore, it clarifies some apparent controversies reported in the literature, e.g., the variability of the critical nucleus size required for the assembly, or the hysteresis between the processes of association and dissociation. Even though here we develop the theory for the case of spherical capsids, the same principles could be extended to other types of structures, or more complex assembly scenarios, e.g., involving the genetic material or scaffolding proteins.

In order to validate some of the theoretical assumptions introduced by CNT, in Chapter 6 (*Simulation of the Constrained Assembly of Spherical Shells*), we simulate the constrained assembly of spherical capsids again using our simple coarse-grained model. Remarkably, this study also unveils some interesting and unexpected phenomena that can have an important role, for instance, during the closure of capsids. The new mechanisms unveiled here open up interesting questions regarding the physical advantages of maturation in viruses, and could also guide the development of novel antiviral strategies focused on promoting a misassembled capsid.

Finally, in the last chapter (*Conclusions and Perspectives*) we will summarize the main achievements of the thesis, emphasizing the potential biological implications of the study, and end up by discussing the new questions and future perspectives derived from our research.

References

[1] S. J. Flint, L. W. Enquist, V. R. Racaniello, and A. M. Skalka, *Principles of virology*. Washington, D.C. ASM Press cop., 2004.

[2] D. H. Crawford, *The invisible enemy: a natural history of viruses*. Oxford university press, 2000.

[3] C. A. Suttle, "Viruses in the sea," *Nature*, vol. 437, pp. 356–61, 2005.

[4] L. K. Miller, A. J. Lingg, and L. A. Bulla, "Bacterial, viral, and fungal insecticides," *Science*, vol. 219, pp. 715–721, 1983.

[5] S. Matsuzaki, M. Rashel, and J. Uchiyama, "Bacteriophage therapy: a revitalized therapy against bacterial infectious diseases," *J. Infect. Chemother.*, vol. 11, pp. 211–9, 2005.

[6] R. C. Mulligan, "The basic science of gene therapy," *Science (New York, N.Y.)*, vol. 260, pp. 926–32, 1993.

[7] M. Emerman and H. S. Malik, "Paleovirology–modern consequences of ancient viruses," *PLoS biology*, vol. 8, p. e1000301, 2010.

[8] W. Stanley, "Isolation of a crystalline protein possessing the properties of tobacco-mosaic virus," *Science*, vol. 81, pp. 644–645, 1935.

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

[10] ICTVdB, "The Universal Virus Database, version 4." http://www.ncbi.nlm.nih.gov/ICTVdb/ICTVdB/index.htm.

[11] A. Klug, "The tobacco mosaic virus particle: structure and assembly," *Philos. Trans. R. Soc. London. [Biol.]*, vol. 354, pp. 531–5, 1999.

[12] F. P. Williams, "U.S. Environmental Protection Agency." http://www.epa.gov/microbes/rota.html.

[13] R. L. Duda, "University of Pittsburgh."

[14] F. A. Murphy, "U.S. Centers for Disease Control and Prevention." http://www.epa.gov/microbes/rota.html, 1975.

[15] S. Brenner, F. Jacob, and M. Meselson, "An unstable intermediate carrying information from genes to ribosomes for protein synthesis," *Nature*, vol. 190, pp. 576–581, 1961.

[16] F. H. Crick, L. Barnett, S. Brenner, and R. J. Watts-Tobin, "General nature of the genetic code for proteins," *Nature*, vol. 192, pp. 1227–32, 1961.

[17] B. La Scola, C. Desnues, I. Pagnier, C. Robert, L. Barrassi, G. Fournous, M. Merchat, M. Suzan-Monti, P. Forterre, E. Koonin, and D. Raoult, "The virophage as a unique parasite of the giant mimivirus," *Nature*, vol. 455, pp. 100–4, 2008.

[18] R. Dawkins, *The selfish gene*. New York City: Oxford University Press, 1976.

[19] R. A. Edwards and F. Rohwer, "Viral metagenomics," *Nat. Rev. Microbiol.*, vol. 3, pp. 504–10, 2005.

[20] S. Yau, F. M. Lauro, M. Z. DeMaere, M. V. Brown, T. Thomas, M. J. Raftery, C. Andrews-Pfannkoch, M. Lewis, J. M. Hoffman, J. A. Gibson, and R. Cavicchioli, "Virophage control of antarctic algal host-virus dynamics," *Proc. Natl. Acad. Sci. USA.*, vol. 108, pp. 6163–6168, 2011.

[21] G. Boccardo, V. Lisa, E. Luisoni, and R. G. Milne, "Cryptic plant viruses," *Adv. Virus. Res.*, vol. 32, pp. 171–214, 1987.

[22] G. Harper, R. Hull, B. Lockhart, and N. Olszewski, "Viral sequences integrated into plant genomes," *Annu. Rev. Phytopathol.*, vol. 40, pp. 119–36, 2002.

[23] M. G. Fischer and C. A. Suttle, "A Virophage at the Origin of Large DNA Transposons," *Science*, vol. 231, 2011.

[24] P. Forterre, "Defining life: the virus viewpoint," *Orig. Life Evol. Biosph.*, vol. 40, pp. 151–160, 2010.

[25] D. Moreira and P. López-García, "Ten reasons to exclude viruses from the tree of life," *Nat. Rev. Microbiol.*, vol. 7, pp. 306–11, 2009.

[26] D. H. Bamford, J. M. Grimes, and D. I. Stuart, "What does structure tell us about virus evolution?," *Curr. Opin. Struc. Biol.*, vol. 15, pp. 655–63, 2005.

[27] F. H. C. Crick and J. D. Watson, "Structure of small viruses," *Nature*, vol. 10, pp. 473–475, 1956.

[28] D. L. D. Caspar and A. Klug, "Physical principles in the construction of regular viruses," in *Cold Spring Harbor Symp. Quant. Biol.*, vol. 27, pp. 1–24, Cold Spring Harbor Laboratory Press, 1962.

[29] T. S. Baker, N. H. Olson, and S. D. Fuller, "Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs," *Microbiol. Mol. Biol. Rev.*, vol. 63, pp. 862–922, 1999.

[30] D. Prangishvili, P. Forterre, and R. A. Garrett, "Viruses of the Archaea: a unifying view," *Nat. Rev. Microbiol.*, vol. 4, pp. 837–48, 2006.

[31] W. M. Gelbart and C. M. Knobler, "Pressurized viruses," *Science*, vol. 323, pp. 1682–1683, 2009.

[32] H. Lortat-Jacob, E. Chouin, S. Cusack, and M. J. van Raaij, "Kinetic analysis of adenovirus fiber binding to its receptor reveals an avidity mechanism for trimeric receptor-ligand interactions," *J. Biol. Chem.*, vol. 276, pp. 9009–15, 2001.

[33] Y. Xiang and M. G. Rossmann, "Structure of bacteriophage ϕ29 head fibers has a supercoiled triple repeating helix-turn-helix motif," *Proc. Natl. Acad. Sci. USA.*, vol. 2011, pp. 1–5, 2011.

[34] B. K. Ganser-Pornillos, M. Yeager, and W. I. Sundquist, "The structural biology of HIV assembly," *Curr. Opin. Struc. Biol.*, vol. 18, pp. 203–17, 2008.

[35] P. E. Prevelige, D. Thomas, and J. King, "Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells," *Biophys. J.*, vol. 64, pp. 824–35, 1993.

[36] W. C. Russell, "Adenoviruses: update on structure and function," *J. Gen. Virol.*, vol. 90, pp. 1–20, 2009.

[37] R. F. Bruinsma, W. M. Gelbart, D. Reguera, J. Rudnick, and R. Zandi, "Viral self-assembly as a thermodynamic process," *Phys. Rev. Lett.*, vol. 90, p. 248101, 2003.

[38] R. Zandi, P. van der Schoot, D. Reguera, W. Kegel, and H. Reiss, "Classical nucleation theory of virus capsids," *Biophys. J.*, vol. 90, pp. 1939–1948, 2006.

[39] S. Katen and A. Zlotnick, *The thermodynamics of virus capsid assembly*, vol. 455, ch. 14, pp. 395–417. Elsevier Inc., 1 ed., 2009.

[40] H. Fraenkel-Conrat and R. C. Williams, "Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components," *Proc. Natl. Acad. Sci. USA.*, vol. 41, pp. 690–698, 1955.

[41] J. B. Bancroft, C. E. Bracker, and G. W. Wagner, "Structures derived from cowpea chlorotic mottle and brome mosaic virus protein," *Virology*, vol. 38, pp. 324–335, 1969.

[42] R. F. Bruinsma, "Physics of RNA and viral assembly," *Eur. Phys. J. E*, vol. 19, pp. 303–10, 2006.

[43] S. Sun, V. B. Rao, and M. G. Rossmann, "Genome packaging in viruses," *Curr. Opin. Struc. Biol.*, vol. 20, pp. 114–20, 2010.

[44] V. Gonzalez-Huici, M. Salas, and J. M. Hermoso, "The push-pull mechanism of bacteriophage ϕ29 DNA injection," *Mol. Microbiol.*, vol. 52, pp. 529–540, 2004.

[45] J. D. Lewis, G. Destito, A. Zijlstra, M. J. Gonzalez, J. P. Quigley, M. Manchester, and H. Stuhlmann, "Viral nanoparticles as tools for intravital vascular imaging," *Nat. Med.*, vol. 12, pp. 354–60, 2006.

[46] M. Comellas-Aragonés, H. Engelkamp, V. I. Claessen, N. a. J. M. Sommerdijk, A. E. Rowan, P. C. M. Christianen, J. C. Maan, B. J. M. Verduin, J. J. L. M. Cornelissen, and R. J. M. Nolte, "A virus-based single-enzyme nanoreactor," *Nature Nanotech.*, vol. 2, pp. 635–9, 2007.

[47] T. Douglas and M. Young, "Viruses: making friends with old foes," *Science*, vol. 312, pp. 873–875, 2006.

[48] C. Mao, D. J. Solis, B. D. Reiss, S. T. Kottmann, R. Y. Sweeney, A. Hayhurst, G. Georgiou, B. Iverson, and A. M. Belcher, "Virus-based toolkit for the directed synthesis of magnetic and semiconducting nanowires," *Science (New York, N.Y.)*, vol. 303, pp. 213–7, 2004.

[49] Y. J. Lee, H. Yi, W.-J. Kim, K. Kang, D. S. Yun, M. S. Strano, G. Ceder, and A. M. Belcher, "Fabricating genetically engineered high-power lithium-ion batteries using multiple virus genes," *Science (New York, N.Y.)*, vol. 324, pp. 1051–5, 2009.

[50] A. E. Walsby, "Gas vesicles," *Microbiol. Rev.*, vol. 58, pp. 94–144, 1994.

[51] J. Querol-Audi, A. Casanas, I. Uson, D. Luque, J. R. Caston, I. Fita, and N. Verdaguer, "The mechanism of vault opening from the high resolution structure of the N-terminal repeats of MVP," *EMBO J.*, vol. 28, pp. 3450–3457, 2009.

[52] M. Sxener, J. Struumpfer, J. A. Timney, A. Freiberg, C. N. Hunter, and K. Schulten, "Photosynthetic vesicle architecture and constraints on efficient energy harvesting," *Biophys. J.*, vol. 99, pp. 67–75, 2010.

[53] J. Zhang, D. Li, G. Liu, K. J. Glover, and T. Liu, "Lag periods during the self-assembly of {Mo(72)Fe(30)} macroions: connection to the virus capsid formation process," *J. Am. Chem. Soc.*, vol. 131, pp. 15152–9, 2009.

[54] W. H. Roos, R. Bruinsma, and G. J. L. Wuite, "Physical virology," *Nature Phys.*, vol. 6, pp. 733–743, 2010.

Part I

Structure of Viral Capsids

Chapter 1

Geometrical Construction of Viral Capsids

1.1 Introduction

A fascinating aspect of viruses is the well defined shape of their capsids. These structures are usually highly symmetrical, and are composed of multiple copies of one or few different coat proteins, which are arranged in a very precise fashion in each virus. The size and morphology of the capsid can vary depending on the virus. However, it is also true that viruses that are apparently unrelated phylogenetically, and infect completely different organisms, can adopt the same structure. In fact, the systematic study of viral capsids show that there are some geometrical rules that can rationalize the structure of an important number of viral capsids.

Among the different shapes adopted by viral capsids, there are four morphologies that have traditionally caught the attention of scientists, due to their relative abundance, their impact on research, or their consequences on human health. These four groups correspond to rod-like, quasi-spherical, bacilliform, and conical viruses (see Fig. 1.1).

The rod-like structures are probably the simplest architectures that viruses can adopt (see Fig. 1.1a). The coat subunits in these capsids form an open helical structure that wraps the genetic material in a channel of a few nanometers. They usually infect plants, and were the first viruses to be observed, once the electronic microscope was developed [1]. Their structural simplicity allowed the first advances in the characterization of viruses, and unveiled the ability of viruses to self-assemble, showing that it was possible to produce infective viral particles

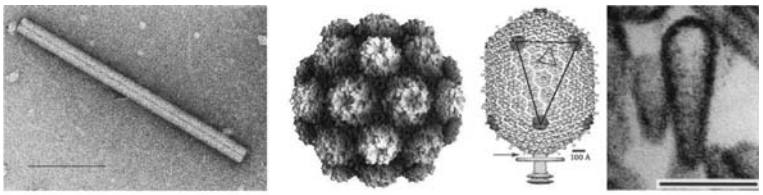


Figure 1.1: Examples of typical viral capsid shapes. (a) Electron micrograph image of tobacco mosaic virus (TMV) [5], a rod-like virus. (b) X-ray reconstruction of turnip yellow mosaic virus (TYMV) [5], which has a spherical shell. (c) Cryo-em reconstruction of a bacteriophage T4 [6], a bacilliform or prolate virus. (d) Electron micrograph image of the nucleocapsid of the human immunodeficiency virus HIV-1 [7], which has a conical shape.

in vitro [2]. The tobacco mosaic virus has been the classic model of study for rod-like viruses [3]. Despite their relative simplicity, the mechanisms involved in the assembly of those viruses are still under investigation [4].

Probably, the spherical capsid is the most common shape adopted by viruses (see Fig. 1.1b), since it is thought that almost half of all viruses are quasi-spherical. The term *quasi-* is used to include also the faceted structures that show a polyhedral rather than strictly a spherical shape, as we will see in more detail in Chapters 3 and 4. The size of these capsids can vary significantly depending on the virus, and their diameters span from $\sim 20\,nm$, e.g., for small plant viruses (made of 60 coat proteins), up to $\sim 500\,nm$, as for giant viruses like the mimivirus (composed of more than $60,000$ subunits), which is comparable to the size of some small bacteria[1]. Despite this polydispersity in sizes, surprisingly, almost all quasi-spherical viruses have a characteristic trait: their capsids show icosahedral symmetry. The ubiquity of icosahedral symmetry in viruses has been a focus of intense research since the mid 1950s, when it was first proposed [8–10], but its origin and biological implications are still being investigated [11–14].

Despite not being as abundant as spherical capsids, bacilliform structures are also a common shape adopted by viruses (see Fig. 1.1c). These elongated morphologies have been extensively reported by electron micrographs showing also a great variety of sizes [5,15]. However, the inherent "anisotropy" of the structures, compared to spherical capsids, has hindered the study of their structure [16,17].

[1] In particular, some members of the bacterial genus Mycoplasma have dimensions around $\sim 300\,nm$.

Fortunately, the progress of non-crystallographic techniques such as cryo-electron microscopy (cryo-EM) has started to unveil the structural details of those capsids [6, 18–20]. The study of the geometrical properties of bacilliform capsids will be precisely the main focus of this chapter, and their main characteristics will be described in more detail along the following sections.

Conical capsids are perhaps the less common shape among the four groups of morphologies that we have enumerated (see Fig. 1.1d). Nevertheless, the interest in these structures comes from the fact that human immunodeficiency virus (HIV), and other related retroviruses, have a conical nucleocapsid protecting the RNA viral genome. However, these viruses can be pleomorphic, meaning that the same virus can easily adopt different possible conical and irregular shapes [7, 21, 22], and it has not been possible yet to reconstruct a full conical viral capsid knowing the precise position of all the coat proteins [23, 24].

Remarkably, the presence of these common architectures in viruses can be justified by combining genetic, physical, and geometrical arguments. We have mentioned above that the size of viral capsids is rather small, on the order of tens to hundreds of nanometers. This in turn restricts the amount of information that can be coded in the viral genome. Therefore, for the sake of genetic *economy* capsids are typically built from multiple copies of one or a few similar small proteins, instead of a single macroprotein [8]. These subunits interact with each other and self-assemble into a regular hollow shell, minimizing the free energy as we will see in detail in Chapter 2. Notice that for a flat 2D surface the hexagonal lattice, or its dual (the triangular lattice), maximizes the packing and the number of interactions of identical subunits. Then, starting from this situation, it is possible to build the four basic capsid shapes listed above. The open helical tube characteristic of rod-like viruses can be obtained by simply wrapping the lattice [10]. On the other hand, to construct closed shells, Euler's theorem states that 12 pentagonal *defects*[2] have to be introduced on the original hexagonal sheet [26]. If those defects are evenly distributed on the surface one will get the polyhedral shell with icosahedral symmetry characteristic of quasi-spherical viruses. In addition, prolate or bacilliform capsids can be made by wrapping the lattice into an helical tube and closing each of its ends with six defects [9, 16]. At last, conical viruses can be built similarly to prolates but introducing a different number of defects in each cap, e.g., five and seven, or four and eight [7].

The physical and geometrical principles leading to the formation of rod-like and quasi-spherical icosahedral viruses are now relatively well understood. How-

[2]It would be also possible to build a closed structure using other type of defects, but they usually introduce a higher degree of elastic frustration [10, 25].

ever, many viruses, including some bacteriophages, such $\phi 29$ or T4, and several fungus, plant, and animal viruses, e.g., in the genera barnavirus, badnavirus, and ascovirus, respectively [15], have a prolate capsid whose geometrical construction is not so well understood.

The main goal of this chapter is to describe the geometrical principles that lead to the construction of such bacilliform viral capsids. We will focus our attention on *closed* elongated viruses, which are also labeled in the literature as prolate, bacilliform, elongated, tubular, or allantoid [15], leaving aside specifically *open-ended* rod-like viruses such as tobacco mosaic virus. Our work is based on the ideas introduced by Caspar and Klug [10] and further extended by Moody [27], and establishes a general geometrical framework to describe icosahedral spherical capsids as well as icosahedrally-capped bacilliform shells. The choice of these particular structures is not arbitrary and will be justified on energetic grounds in the next chapter.

The importance of this geometrical description is that it enumerates and characterizes the set of structures that can be built. We find that prolate capsids adopt a discretized set of lengths, radius, and numbers of proteins. The knowledge of these selection rules can be very useful to infer the structure of a virus from simple experimental data as well as for nanotechnological applications that rely on a proper control of the dimensions and architecture of viral capsids.

This chapter will be distributed as follows. First, in Section 1.2 we review the classical Caspar and Klug construction for icosahedral viral capsids. Section 1.3 is devoted to the general geometrical description of bacilliform capsids using all possible symmetrical constructions of prolates closed by hemi-icosahedral caps. The main geometrical properties of the resulting prolates as well as several applications of the model for specific viruses will also be analyzed. Finally, Section 1.4 discusses some biological implications of our results, and in Section 1.5 we summarize our findings and remark their most relevant consequences.

1.2 Icosahedral capsids

As mentioned above, the quasi-spherical shape is probably the most common architecture among viral capsids. Despite the variety of infection routes or assembly mechanisms used by spherical viruses, the majority of them adopt a well-ordered structure with icosahedral symmetry, as illustrated by the gallery of Fig. 1.2. This means that the coat proteins are organized in a way that the resultant shell has the same elements of symmetry of an icosahedron, i.e., the presence of 15 2-fold, 10 3-fold and 6 5-fold rotation axes (see Fig. 1.3).

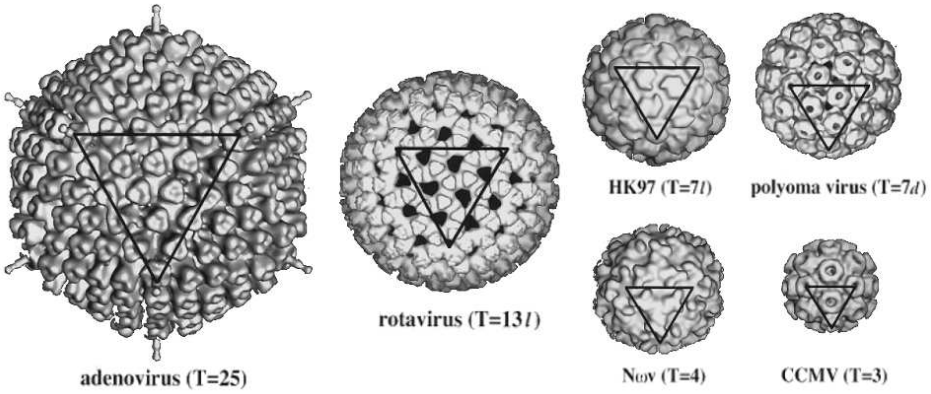


Figure 1.2: Quasi-spherical viruses. The cryo-electron reconstructions of several quasi-spherical viruses are shown. The shapes of capsids can be almost perfectly spherical, like for rotavirus, or strongly faceted adopting a polyhedral shape, as for adenovirus. Despite the variety of sizes of viral capsids, or the different amount of coat proteins composing the shell, most quasi-spherical viruses have icosahedral symmetry and can be labeled by the structural index T. In each case the depicted triangle corresponds to one face of the implicit icosahedron, whose vertexes are associated to 12 pentamers. The viruses shown in the figure infect different hosts. Hong Kong 97 (HK97) is a bacteriophage virus; cowpea chlorotic mottle virus (CCMV) is a plant virus; Nudaurelia capensis omega (NωV) infects arthropods; while humans are the host of polyomavirus, rotavirus, and adenovirus. Adapted from Ref. [11].

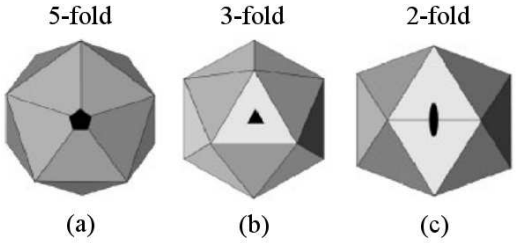


Figure 1.3: The three axes of symmetry of an icosahedron: 5-fold (a), 3-fold (b), and 2-fold (c).

The first insights into the occurrence of icosahedral symmetry in viruses were presented by Crick and Watson in 1956. They suggested that small spherical viruses should adopt a regular capsid made of multiple copies of the same protein [8]. From a biological standpoint, this solution is specially efficient in terms of genetic information, because a regular structure made of multiple copies of the same protein is encoded in less genetic material that an hypothetical single macroprotein shell. Their conclusions were based on different grounds. The two spherical viruses crystallized at that time, bushy stunt virus (BSV) [28, 29] and turnip yellow mosaic virus (TYMV) [30], showed crystals with a unit cell adopting the shape of a cube. Then, since only one single viral particle was contained in each unit cell, this meant that the viral capsid must possess also the point group symmetry of the unit cell. In addition, it was known that the viral coat protein was asymmetric, so this pointed out that, in order to obtain a symmetrical viral particle, several coat proteins had to be involved in the construction of the capsid. The ways to build a regular structure, with cubic point group symmetry, combining asymmetric coat proteins are limited, and are all related to the classical platonic solids (see Table 1.1). Remarkably, of all platonic solids, the icosahedron is the biggest structure, which can accommodate 60 asymmetrical subunits (3 on each face), and maximizes the amount of volume (genetic information) per unit surface (coat protein production).

Interestingly, in the same number of Nature where Crick and Watson published their seminal paper, D. Caspar showed that bushy stunt virus (BSV) adopted the symmetry 532 corresponding to the icosahedral construction [31] (see Table 1.1), and in the subsequent years, other spherical viruses were found to possess also icosahedral symmetry [9].

The icosahedral point group symmetry does not impose that capsids must be made strictly of 60 coat proteins [8]. In fact, studies on icosahedral viruses started to show that the same symmetry was compatible with capsids made of a different number of subunits. In 1961 Horne and Wildy were the first to propose a model that rationalized the possible icosahedral structures of viral capsids [9], but it was not until the seminal work of Caspar and Klug in 1962 [10] when the theory of icosahedral capsids was properly established.

The smallest icosahedral viruses have capsids made of 60 coat proteins that are arranged in an exact regular way on the surface of an icosahedron (see Fig. 1.4). Nevertheless, the situation becomes more complicated for larger viruses, because not all subunits can be placed in fully equivalent environments. So, despite having icosahedral symmetry, larger capsids will not be strictly regular (in terms of the coat proteins). Then, inspired by the geometrical principles

Crystallographic description	No. and type of rotation axes present	No. of asymmetric units	Platonic solid with these symmetry elements
23	3 2-fold 4 3-fold	12	Tetrahedron
432	6 2-fold 4 3-fold 3 4-fold	24	Cube Octahedron
532	15 2-fold 10 3-fold 6 5-fold	60	Dodecahedron Icosahedron

Table 1.1: The three possible cubic point groups for spherical viruses. The number of subunits will be the same as, or a multiple of, the number of asymmetric units. A platonic solid (and its dual) is associated to each crystallographic description. This analysis was first introduced by Crick and Watson in Ref. [8].

applied by Buckminster Fuller in the construction of geodesic domes, Caspar and Klug (CK) reasoned that coat proteins in these larger capsids should occupy also *quasi-equivalent* environments [10]. This quasi-equivalence principle was the key to formulate the precise rules to construct and characterize the possible icosahedral capsids, and, since then, the CK construction has been the basis to understand and classify the structure of spherical viruses [12] (see Fig 1.2).

In the remaining of this section we will introduce in detail the CK model for quasi-spherical viruses, whose ideas will be the basis of the generalization for elongated capsids that will be developed in Section 1.3.

1.2.1 The Caspar and Klug construction and the T-number

CK showed that the requirement of quasi-equivalence leads necessarily to icosahedral symmetry as the most efficient design [10]. In these icosahedral capsids proteins can be geometrically clustered in two types of morphological units: *pentamers*, which are groups of five proteins aggregated around each vertex of the icosahedron; and *hexamers*, which are clusters of six proteins evenly distributed on the faces and edges of the capsid (see Fig. 1.5).

The starting point in the CK model is a flat hexagonal lattice or equivalently its dual, the triangular one. Using this lattice, there is only a limited number of ways to create a closed shell with icosahedral symmetry. Essentially, one has to replace 12 evenly-distributed hexamers by the 12 pentamers required by Euler's theorem to make a closed surface [26]. The different ways to accomplish that correspond to different *triangulation numbers* (T), that in the CK model serve as

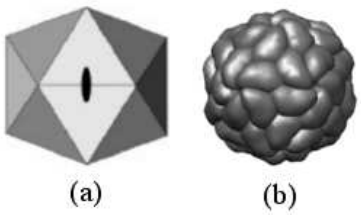


Figure 1.4: The regular icosahedral capsid. (a) Icosahedron centered on a 2-fold axis. (b) Icosahedral viral capsid made of 60 coat proteins. The structure is also centered on a 2-fold axis. It can be seen that a morphological cluster of 5 proteins (pentamer) forms around each of the 12 vertexes of the icosahedron. The capsid corresponds to the satellite tobacco mosaic virus (STMV; PDB 1a34).

a systematic way to classify the structure of quasi-spherical icosahedral viruses (see Figs 1.6 and 1.7).

The T-number is defined by the vector that joins two adjacent pentamers in the lattice, namely,

$$\vec{C}_T = h\vec{a}_1 + k\vec{a}_2 \equiv (h, k) \tag{1.1}$$

where (h, k) are non-negative integers that give the number of steps to connect two nearest pentamers along the principal directions of the hexagonal lattice, i.e., $\vec{a}_1$ and $\vec{a}_2$ (see Fig. 1.6a).

The smallest triangular face is defined by $(1, 0)$ or equivalently $(0, 1)$, and has an area $S_0 = |\vec{a}_1 \times \vec{a}_2|/2$. The triangulation number T is the area of the equilateral triangle defined by $\vec{C}_T$ divided by S_0, or, equivalently, the number of equilateral triangles of area S_0 contained in a face of the resulting icosahedron. Using the elementary properties of the hexagonal lattice listed in Appendix A, one obtains

$$T = h^2 + hk + k^2 = Pf^2. \tag{1.2}$$

Since h and k are non-negative integers, T follows a particular series of "magic" numbers, i.e., $T = 1, 3, 4, 7, 9, 12, 13, \ldots$ Interestingly, T also corresponds to the number of quasi-equivalent locations of the subunits in the shell [10, 34].

Since an elementary $T = 1$ triangle accommodates three proteins and the resulting icosahedron is built by twenty T-faces, the total number of proteins in the capsid is

$$N_{sub} = 60T. \tag{1.3}$$

In addition, every structure has $N_P = 12$ pentamers, accounting for sixty proteins, and the remaining subunits are distributed in $N_H = 10(T - 1)$ hexamers.

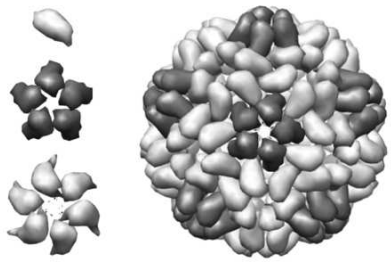


Figure 1.5: From the coat protein to the capsid. In the simplest case, viral capsids are made of multiple copies of the same subunit. The figure illustrates the organization of coat proteins for the brome mosaic virus (BMV). (a) From top to bottom: a single coat subunit, pentamer made of five coat proteins (dark gray), and hexamer made of six (gray). (b) The full capsid of brome mosaic virus (BMV) is made of 180 subunits arranged in a $T = 3$ icosahedral shell (see text). The color coding for capsomers is the same as in (a). The structure was solved by X-ray (PDB 1js9) [32] and has been rendered using Chimera [33].

Therefore the total number of capsomers in the capsid is

$$N = 10\,T + 2. \tag{1.4}$$

This is a very important prediction of the CK model that will be explored in the next chapter.

1.2.2 The P-class

In modern structural virology icosahedral capsids are generally described by the number $T(h, k)$. Nevertheless, Caspar and Klug proposed also a reorganization of the T-structures in terms of *classes*, characterized by a new number P [10], which must not be confused with the pseudo-triangulation number used to label icosahedral capsids made of chemically different proteins. The key point of this reorganization was that shells within the same class have common geometrical properties, and along this thesis we will see that this also has important implications in their physical properties.

To obtain the corresponding class P, one must calculate the greatest common divisor of the pair (h, k), which defines the *multiplicity* $f = gcd(h, k)$. Then one obtains an irreducible pair of indexes, (h_0, k_0), directly related to the original

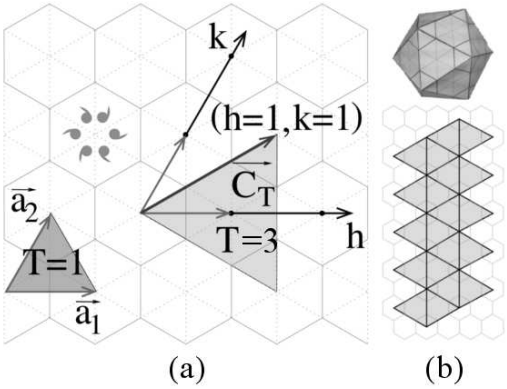


Figure 1.6: (a) Basic elements of the Caspar and Klug construction. The starting point of the CK construction is a hexagonal lattice subdivided in small triangles, which can be associated to single subunits. A particular disposition of asymmetrical proteins is drawn on one hexagon of the network forming a hexamer. The key point of the CK construction is to build an equilateral triangle that will define the face of the final icosahedron. The vertexes of this triangle will be associated to the vertexes of the icosahedral capsid. The different distances and orientations (respect the hexagonal lattice) of the line that joins two vertexes of this face-triangle, given by the vector $\vec{C}_T(h,k)$, determines the possible set of icosahedral structures. The steps along the lattice are expressed in terms of the coordinates h and k related to the hexagonal lattice vectors $\vec{a}_1$ and $\vec{a}_2$. These two vectors alone determine the smallest face-triangle possible that is associated to a regular icosahedron (small triangle shaded). In terms of the hexagonal coordinates it can be expressed by ($h = 1$, $k = 0$) (or vice versa), and its surface defines a $T = 1$ triangle that contains three subunits on average. The big shaded triangle is defined by $\vec{C}_T(h = 1, k = 1)$ and has a triangulation number $T = 3$, i.e., three times the surface of a $T = 1$. (b) The icosahedral T-shell is built by twenty of these T triangles. A flat icosahedral template (bottom) and the corresponding folded structure (top) are shown for the $T = 3$ capsid discussed in (a).

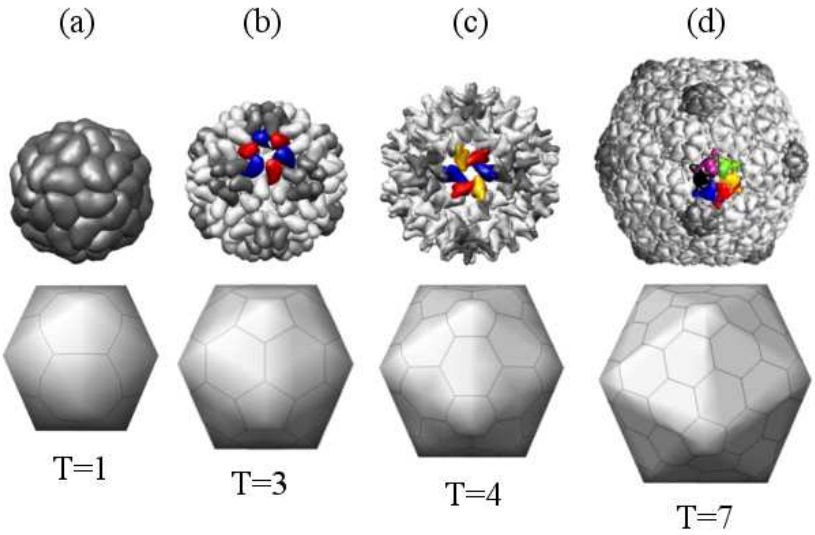


Figure 1.7: Examples of T-number capsids. From left to right the first four icosahedral structures in the CK construction are represented (bottom). On top of each shell we have chosen a high-resolution viral capsid as an example, where pentamers are shown in dark gray, and all hexamers are gray except one. The different colors of this hexamer (plus the subunit on the pentamer) illustrates the T quasi-equivalent positions of the coat protein in the capsid. All capsids have been rendered using Chimera [33], and centered on a 2-fold axis. (a) Satellite tobacco mosaic virus (STMV; PDB 1a34) is made of 60 subunits and adopts a $T = 1$ structure ($h = 1$, $k = 0$). (b) The $T = 3$ structure ($h = k = 1$) of brome mosaic virus (BMV, PDB 1js9) that is build by 180 coat proteins. (c) Hepatitis B virus (HBV, PDB 2g33), a $T = 4$ shell ($h = 2$, $k = 0$) composed of 240 subunits (*in vivo* HBV can also form $T = 3$ structures). (d) Finally, the $T = 7$ shell of Hong Kong 97 (HK97, PDB 1ohg), which is made of 420 coat proteins. Notice that the $T = 7$ shell is skewed and corresponds to the *laevo* construction $h > k$, i.e., $(h = 2\,, k = 1)$. Its specular image will be a *dextro* capsid $T = 7d$ with coordinates $(1, 2)$.

	T			
$P \setminus f$	1	2	3	4
1	1	4	9	16
3	3	12	27	48
7	7	28	63	112
13	13	52	117	208

Table 1.2: T-numbers for the first four P classes (rows) classified by the multiplicity f (columns).

ones by $h = fh_0$ and $k = fk_0$. Hence, by applying this description into the $T(h, k)$ expression, Eq. (1.2), one gets

$$P = h_0^2 + h_0 k_0 + k_0^2 \tag{1.5}$$

The class P is a subset of T ($P = 1, 3, 7, 13, \ldots$) that adopts the value of the smallest triangulation number in the group, i.e., $P = T(h_0, k_0)$ (see Table 1.2). In particular, shells with $(h, 0)$ or $(0, k)$, i.e. $T = 1, 4, 9, \ldots$, belong to the class $P = 1$, and are characterized by the presence of a straight line of hexamers connecting neighboring pentamers (see Fig. 1.8a). Those characterized by (h, h), i.e. $T = 3, 12, 27, \ldots$ belong to the class $P = 3$, and in this case pentamers are connected by a zigzag line of capsomers (see Fig. 1.8b); finally, any other situation, i.e., $h \neq k \neq 0$, leads to classes with $P > 3$ that are characterized by chiral structures. This means that (h, k) generates a skew architecture specular to the shell represented by (k, h). To distinguish these two "isomers" it is customary to use the label l, which means *laevo* or *left-handed*, for $h > k$, and d, meaning *dextro* or *right-handed*, for $h < k$ (see Fig. 1.9).

Interestingly, icosahedral capsids within the same class have self-similar faces with an analogous distribution of capsomers. In other words, the triangular face of the structure $T(h_0, k_0)$ perfectly tiles the rest of T-shells from the same class (see Fig. 1.8). The number of P-triangles necessary to tile the $T(h, k)$-face is given by f^2, see Eq. (1.2). For instance, the face of a $T = 27$ has nine times ($f = 3$) the distribution of proteins of a $T = P = 3$ triangle, Fig. 1.8b.

Remarkably, this reorganization of capsids in terms of classes P will be also useful to study the geometrical properties of prolate capsids.

1.2.3 Degeneration of the T-number

The modern classification of quasi-spherical viruses is made usually in terms of this triangulation number T. But it is important to clarify that the triangulation

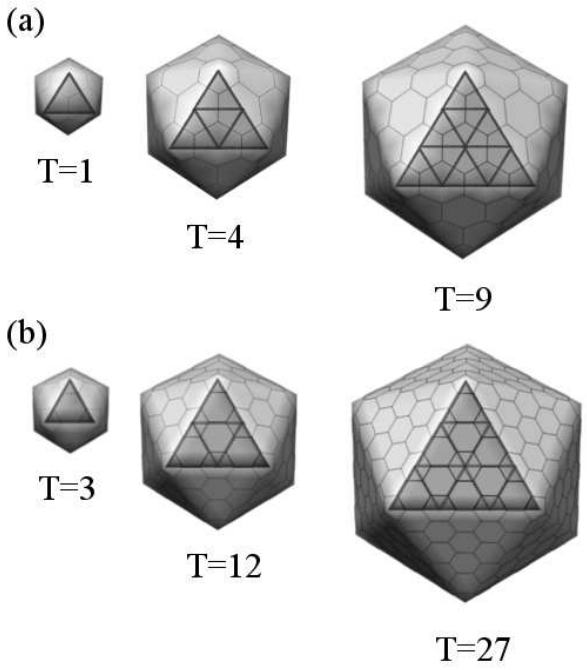


Figure 1.8: Self-similarity of capsids within a class. All capsids that belong to the same class P have an analogous distribution of hexamers. (a) In this case the first three triangular numbers of the class $P = 1$ ($h_0 = 1$, $k_0 = 0$) are shown. From left to right, $T = 1$ ($f = 1$), $T = 4$ ($f = 2$), and $T = 9$ ($f = 3$). In each capsid a face of the icosahedron is shaded. Notice that a triangular face of the smallest shell, i. e., $T_0 = P$, perfectly tiles the rest of capsids of the class using f^2 triangular faces on each case. Thus, $T = 4$ and $T = 9$ embed 4 ($f = 2$) and 9 ($f = 3$) face-triangles of a $T = 1$, respectively. (b) Same as in (a), but now for the first three triangulation numbers of the class $P = 3$ ($h_0 = k_0 = 1$), corresponding to $T = 3$ ($f = 1$), $T = 12$ ($f = 2$), and $T = 27$ ($f = 3$). The structures in (a) and (b) are not at scale.

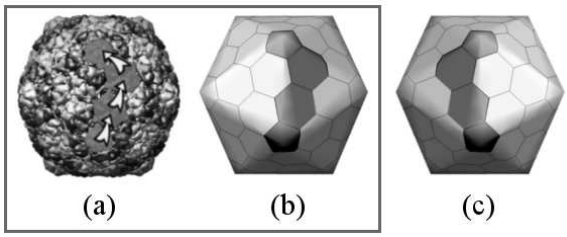


Figure 1.9: Skew icosahedral capsids. All capsids $P > 3$ are skewed and have two enantiomorph structures. If $h > k$ the structure is called *laevo*, whereas if $h < k$ it is called *dextro*. The letters l and d indicate the chirality, respectively. (a) Cryo-em reconstruction of bacteriophage HK97, whose capsid adopts a $T = 7l$. The shaded capsomers and the arrows illustrate the steps along the hexagonal lattice ($h = 2$, $k = 1$) that joint two neighboring pentamers. (b) CK geometrical construction of a $T = 7l$ corresponding to the HK97 capsid shown in (a). (c) The specular *dextro* CK construction of (b), i. e., $T = 7d$ ($h = 1$, $k = 2$).

number does not always provide a unique characterization of a shell, since for some $T \geq 49$ more than one pair (h, k) can share the same T. For instance, $(7, 0)$ and $(5, 3)$ both give $T = 49$. The same applies to the class P. Although it will work for instance in the previous example of $T = 49$ corresponding to $P = 1$ and $P = 49$, there are still cases where different structures have the same P, e.g., $P = 91$ can be obtained by two different pairs (h_0, k_0): $(6, 5)$ and $(9, 1)$.

Obviously, the simplest way to univocally describe a spherical capsid is through the indexes h and k. Although alternative geometrical parameters providing interesting structural information are also possible [35].

1.2.4 Limitations of the CK model

It is appropriate to make some clarifications regarding the CK model that will also apply to the extension for prolate shells that will be developed in the next section.

The CK construction only determines the point group symmetry of the capsid, and is compatible with different clustering of the proteins. Thus coat proteins are not strictly required to form real pentamers or hexamers. For instance, capsids with $20T$ trimers, $30T$ dimers, or $60T$ monomers are completely valid. However, independently on the physical clustering, it is customary to describe viral capsids in terms of capsomers (or morphological units), i.e., pentamers or hexamers.

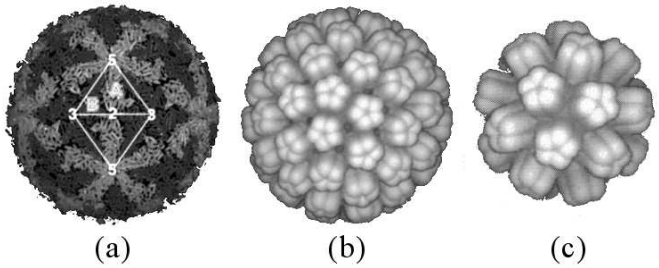


Figure 1.10: Caspar and Klug exceptions. (a) The bluetongue virus (BTV) adopts an icosahedral capsid made of 120 proteins ("$T = 2$"). Considering the dimer AB indicated in the picture (dark and light gray) as the fundamental coat subunit, the structure is equivalent to a $T = 1$ [36]. (b) Native capsid of polyoma virus made of 72 pentamers arranged as in a $T = 7d$ [37]. (c) Snub cube structure made of 24 pentamers adopted also by polyoma virus reconstituted *in vitro* [37].

Furthermore, the fundamental subunit in the CK construction is not necessarily a single protein. For example, the bluetongue virus (BTV) core [36] is an icosahedral shell composed of 120 proteins, which would correspond to a *forbidden* triangulation number, $T = 2$, that apparently violates the CK construction (see Fig. 1.10a). However, if one considers the protein dimers as the fundamental building blocks of the capsid, then the shell becomes a regular structure made of 60 units (in this case dimers) organized as in a $T = 1$ shell.

Finally, not all quasi-spherical viruses strictly comply with the CK model. In their native form, polyoma and papilloma viruses are built only by pentamers arranged as in a $T = 7d$ capsid [37]. In this case 60 pentamers act as *hexons* and 12 as *pentons*[3]. Polyomavirus can also be reconstituted *in vitro* in a quasi-spherical, but non-icosahedral, structure that resembles a snub cube [37], which is completely outside the CK model. In this context, we must point out that both CK's capsids and the exceptions mentioned above have been found to be free energy minima of protein aggregates [13], as we will show in the next chapter. Some of the all-pentamer viral exceptions have also been justified in geometric terms by using a tiling approach [38, 39].

[3] A hexon and a penton are capsomers surrounded by six and five capsomers, respectively.

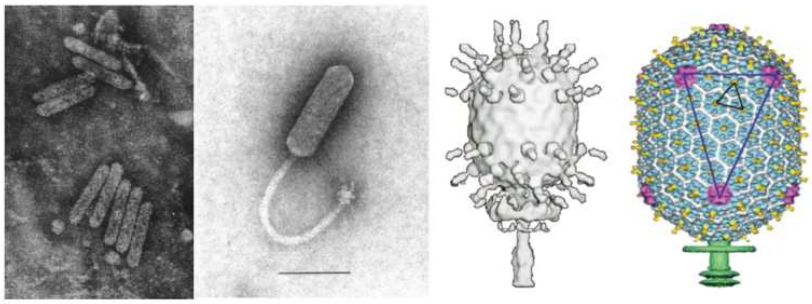


Figure 1.11: Bacilliform viruses. There are many viruses that adopt an elongated shaped closed by hemi-spherical caps at both ends, which have been identified by electron microscopy. Two examples of electron micrographs extracted from the ICTVdB [5] are shown: the plant virus cacao swollen shoot virus (a), and the bacteriophage bacillus virus $\theta 1$ (b). However, only two bacilliform viruses have been reconstituted so far by high resolution imaging techniques (cryo-electron microscopy): bacteriophage $\phi 29$ [18] (c), and bacteriophage T4 [6] (d).

1.3 Bacilliform capsids

A significant number of viral species have a closed elongated capsid whose precise structure is not as well characterized as the spherical shells described above (see Fig. 1.11). Here we will discuss the geometrical principles involved in the construction of bacilliform capsids, and in the next chapter we will justify their occurrence from a physical perspective.

We saw above that most spherical viruses have icosahedral symmetry and follow the Caspar and Klug construction. Therefore a natural guess would be that bacilliform viruses should be made in a similar way by an elongation of an icosahedron. In the mid-sixties, Moody was the first to suggest an adaptation of the CK model for prolate bacteriophages [27]. An important number of phage viruses contain dsDNA and are characterized by a protein tail connected to the capsid [40] (see Fig. 1.11b), which can act as a syringe to drive the injection of the genetic material into the host. Interestingly, Moody analyzed the position of the tail for wild types and mutants for several T-even phages[4], and pointed out that those prolate bacteriophage capsids should be CK shells elongated through

[4]The T-even bacteriophages are any of the $T2$, T4, or $T6$ viruses of Escherichia coli. They share some features, but the T in the name is not related to the icosahedral index T.

the 5-fold axis [17, 27] (see Fig. 1.12a and 1.13).

Contrary to spherical capsids, there are just a few high resolution reconstructions of prolate shells. The only cases that we are aware of, bacteriophages $\phi 29$ and T4, have been reconstructed by cryo-EM studies [6, 18] (see Figs. 1.11c and 1.11d, respectively), and they both follow the model suggested by Moody [17, 27] (see Section 1.3.1).

However, in the seventies, several studies already suggested that some bacilliform plant viruses might deviate from Moody's prescription. These capsids have no tail attached (see Fig. 1.11a), and diffraction and molecular weight experiments pointed to structures with 3-fold rather than 5-fold axial symmetry [41,42]. More specifically, Hull studied some possible designs [16], but he did not describe precisely the geometrical rules to construct them.

The main goal of this section is to generalize CK's and Moody's ideas to describe the construction of bacilliform capsids, based on the elongation of an icosahedron along all its different axes of symmetry, i.e., 5-fold, 3-fold, and 2-fold. We will also derive the general rules that dictate the number of proteins, the radius, length, and chirality of the body. Finally, we exemplify some applications of these ideas by analyzing several viruses in detail.

1.3.1 Geometrical construction of prolate capsids

The simplest way of building a prolate capsid is by elongating an icosahedron along one of its possible axes of symmetry. In such procedure, the 20 triangular faces of the icosahedron have to be distributed among the tubular body and the two caps (see Fig. 1.12). The body must be free of pentamers, except in the boundaries, and each cap must keep 6 pentameric vertexes (otherwise one would obtain a conical rather than a cylindrical structure). These constrains fix the number of triangular faces in the cap for each axial symmetry, as shown in Fig. 1.12, and it is the first property that we need to determine.

In particular, for the 5-fold case the cap is made by five triangular faces (see Fig. 1.12a), whereas for the 3-fold and 2-fold situations caps are built by four faces (see Figs. 1.12b and 1.12c). Thus, if we define $\triangle^{cap}$ as the number of faces in the cap, we have $\triangle^{cap}_{5F} = 5$ and $\triangle^{cap}_{3F} = \triangle^{cap}_{2F} = 4$ for the different symmetries. Remarkably, the 2-fold case has two specular possibilities shown in Figs. 1.12c and 1.12d. Both are valid to define the 2-fold cap, but they have some differences related to the handedness of the structure as will be discussed later.

The number of triangular faces in the body, $\triangle^{body}$, is then obtained by subtracting from the 20 faces of an icosahedron the number of triangles involved in

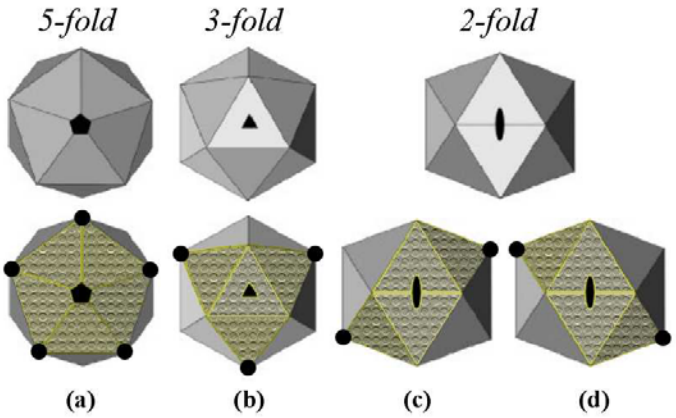


Figure 1.12: The three axes of symmetry of an icosahedron: 5-fold (a), 3-fold (b), and 2-fold (c and d). The patterned triangles emphasize the end-faces that constitute the cap of the elongated structure. The solid dots highlight the vertexes that define the rim of the caps. In (c) and (d) we show that the construction of the 2-fold prolate is intrinsically skewed and has two possibilities.

the caps

$$\triangle^{body} = 20 - 2\triangle^{cap} \tag{1.6}$$

Thus in the 5-fold case the body has ten triangular faces, $\triangle_{5F}^{body} = 10$, and in the 3-fold and 2-fold situations it has twelve, $\triangle_{3F}^{body} = \triangle_{2F}^{body} = 12$. In general, for each prolate not all the triangular body faces will be equal. Fortunately, symmetry arguments allow us to calculate the number of non-equivalent triangles required to construct the body of a prolate capsid for each axial symmetry.

Any prolate capsid constructed as an elongation of an icosahedron has two different types of rotational symmetry. First, the axial symmetry, i.e., 5-fold, 3-fold, or 2-fold, indicating that we have 5, 3, and 2 equivalent views of the particle obtained by rotation around the longitudinal axis, respectively. The second symmetry stems from the fact that the two caps of the structure are equivalent. More precisely, we can always find in the middle of the body a 2-fold symmetry axis perpendicular to the axis of the prolate.

The product of these two general symmetries defines the *symmetry number*

$$s_t = 2s \tag{1.7}$$

that sets the number of body-faces that must be equivalent, i.e., the number of triangular faces that should be the same to respect the symmetries of the prolate.

s	$\triangle^{cap}$	$\triangle^{body}$	$s_t = 2s$	$\triangle^{n-eq}$
5	5	10	10	1
3	4	12	6	2
2	4	12	4	3

Table 1.3: Distribution of the triangular faces for the different icosahedral prolate constructions (see also Fig. 1.12). Rows are associated to the axial symmetry s; $\triangle^{cap}$ is the number of faces in each cap; $\triangle^{body}$ is the total number of triangular faces in the body; s_t is the symmetry number; and $\triangle^{n-eq}$ is the number of non-equivalent triangles present in the body of a prolate.

Now, by dividing the total number of faces in the body by the symmetry number, we obtain the number of non-equivalent groups of body-triangles

$$\triangle^{n-eq} = \frac{\triangle^{body}}{s_t} \tag{1.8}$$

Thus from Table 1.3 is straightforward to see that for the 5-fold case only one triangle is needed to characterize the body (see Fig. 1.13), $\triangle_{5F}^{n-eq} = 1$, whereas for the 3-fold and 2-fold prolates $\triangle_{3F}^{n-eq} = 2$ and $\triangle_{2F}^{n-eq} = 3$, i.e. two and three different triangles are needed, respectively.

All triangles in both caps are equilateral and equal, and are determined by the cap vector $\vec{C}_T$, Eq. (1.1). As in the CK model, this vector defines the triangulation number of the caps $T_{end} \equiv T$, Eq. (1.2), and fixes the radius of the prolate. To describe the triangular faces of the elongated body we need a second vector $\vec{C}_Q$, which connects a pentamer in one cap to the closest one in the opposite cap. This vector is given by

$$\vec{C}_Q = h'\vec{a}'_1 + k'\vec{a}'_2 \equiv (h', k')' \tag{1.9}$$

where $(h', k')'$ are integers from a second pair of hexagonal coordinates, rotated counterclockwise 60^o with respect to the original ones (see Fig. 1.13 and Appendix A). Even though it is not strictly necessary to define a new pair of axes, this representation is particularly convenient, because for $h' = h$ and $k' = k$ we recover in the body the equilateral triangle that defines the face of an icosahedron.

Using these ideas, let's now discuss the construction of elongated capsids for the three icosahedral symmetries.

5-fold prolates

This architecture was already proposed by Moody [17,27]. It is the simplest case because the body is made by ten copies of the same mid-triangle (see Table 1.3 and Fig. 1.13). Indeed, the triangular body-face is defined straightforwardly by $\vec{C}_T$ and $\vec{C}_Q$. Analogously to the quasi-spherical case, the normalized surface of this mid-triangle, $|\vec{C}_T \times \vec{C}_Q|/2S_0$, defines a new triangulation number, this time associated to the body,

$$T_1 = hh' + hk' + kk' \equiv Q_1 f, \tag{1.10}$$

which Moody labeled as T_{mid}. This T_1-number can also be expressed in terms of the multiplicity $f = gcd(h, k)$, where

$$Q_1 = h_0 h' + h_0 k' + k_0 k' \tag{1.11}$$

Since in the 5-fold case there is only one kind of body-triangle, we will characterize the length of the resulting structure by using the number

$$Q_{5F} \equiv Q_1 \tag{1.12}$$

Then any 5-fold prolate capsid will be determined by the triangulation number of the cap T and the Q_{5F} number of the body; or alternatively by its class P, the multiplicity f, and Q_{5F}.

For a given P and f of the cap, the indexes h' and k' control the length of the shell. In particular, for $h' = h$ and $k' = k$, one obtains $T_1 = T_{end}$ and

$$Q^0_{5F} = Pf \tag{1.13}$$

thus recovering the original CK's icosahedral shell. The elongation with respect to the spherical capsid can be then characterized by (see Fig. 1.13a)

$$\Delta Q \equiv Q - Q^0 \tag{1.14}$$

In general, the length of a prolate will be larger than that of the icosahedral structure, i.e., $\Delta Q > 0$. However, it is geometrically possible also to build structures with a shorter length where $T_1 < T_{end}$, i.e., $\Delta Q < 0$ (see Fig. 1.14d). But those oblate capsids have not been found experimentally [17].

The number of proteins in a prolate capsid is determined by the total surface of the structure, which can be obtained by summing up the triangulation numbers of all faces. In this case there are ten end-triangles in the caps, $10\,T_{end}$, and ten

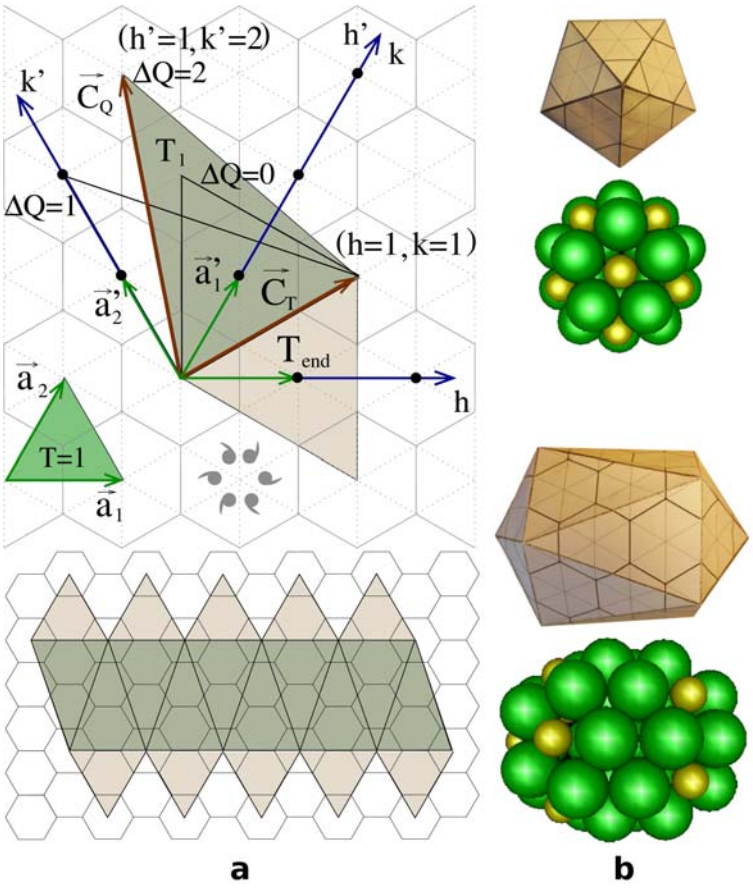


Figure 1.13: (a)(top) Illustration of Moody's geometrical model for 5-fold prolate capsids. (bottom) Complete flat design of a $T_{end} = 3$ and $T_{mid} = Q = 5$ prolate capsid, which corresponds to the shell of bacteriophage ϕ29 [18]. (b) Zenithal (top) and lateral (bottom) views of the folded structure of a $T_{end} = 3$ and $T_{mid} = Q = 5$ prolate capsid. Below each view, there is a "ping-pong" model representation of the same capsid, where hexamers are colored in green and pentamers in gold.

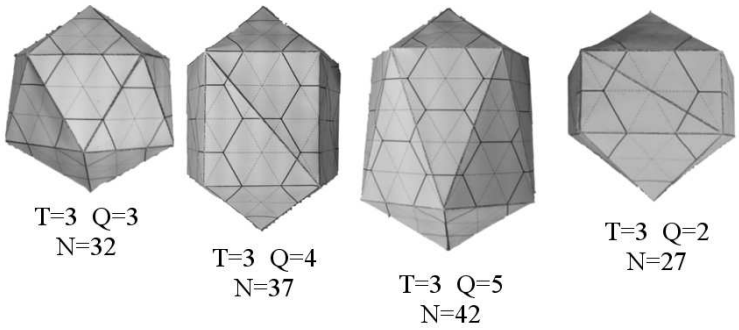


Figure 1.14: Elongation of a $T = 3$ 5-fold prolate. (a) Icosahedral CK construction. (b) First length step given by $\Delta Q = 1$ or $\Delta N = 5$. (c) Second length step with $\Delta Q = 2$ and $\Delta N = 10$. (d) *Shrunk* or oblate prolate given by $\Delta Q = -1$ or $\Delta N = -5$.

mid-triangles in the body, $10\,T_1$. Since three proteins can be accommodated in a $T = 1$ triangle the total number of protein subunits in the capsid is

$$N_{sub}^{5F} = 3(10\,T_{end} + 10\,T_1) = 30f(Pf + Q_{5F}). \tag{1.15}$$

As in icosahedral capsids, the twelve pentamers of a prolate require sixty proteins. Thus the remaining proteins are distributed in $N_H^{5F} = 5(T_{end}+T_1)-10$ hexamers, and the total number of capsomers is

$$N_{5F} = 5(T_{end} + T_1) + 2 = 5f(Pf + Q_{5F}) + 2. \tag{1.16}$$

As discussed above, the value of T_{end} controls the radius of the structure. If we fix it, i.e., if P and f are constant, the different values of $Q_{5F}(h', k')$ in Eq. (1.16) give the possible lengths of the prolate in terms of number of capsomers. Since h' and k' are integers, the number of capsomers and proteins in the body of a prolate can only adopt a discrete set of values. One can prove using Bezout's identity (see Appendix B) that the minimum step possible in Q_{5F} is

$$\Delta Q_{5F}^{min} = 1. \tag{1.17}$$

Thus, unlike T_{end}, Q_{5F} can be any non-negative integer. Combining Eqs. (1.16) and (1.17), the minimum length step of a prolate in terms of capsomer numbers is

$$\Delta N_{5F}^{min} = 5f\Delta Q_{5F}^{min} = 5f \tag{1.18}$$

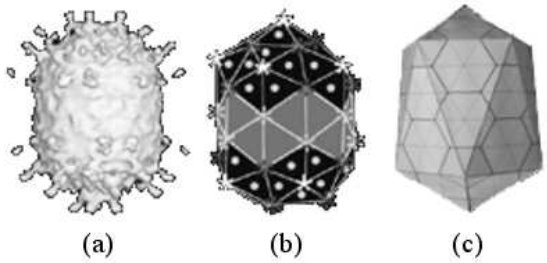


Figure 1.15: Prolate virus bacteriophage $\phi 29$. From left to right: (a) cryo EM reconstruction, (b) triangulation net for a $T = 3$ $Q = 5$ proposed by Tao *et al* [18], and (c) equivalent polyhedral shell construction centered on a 5-fold axial symmetry.

or similarly $\Delta N^{5F}_{sub} = 30f$ in terms of subunits. Hence, prolates based on $f = 1$ caps, e.g., $T_{end} = 1$ or $T_{end} = 3$, have lengths discretized in steps of five capsomers, i.e., 30 proteins. However, those based on $f = 2$, e.g., $T_{end} = 4$ or $T_{end} = 12$ caps, must add multiples of ten capsomers, i.e., 60 proteins, to enlarge the structure.

As mentioned before, the two detailed reconstructions of prolate viruses up to date, corresponding to bacteriophages $\phi 29$ and T4 [6,18], have a 5-fold elongated capsid characterized by $T_{end} = 3$ $Q_{5F} = 5$ and $T_{end} = 13l$ $Q_{5F} = 20$, respectively (see Fig. 1.15 and Section 1.3.3).

3-fold prolates

Alternatively, prolates can also be made by the elongation of an icosahedron along one of its 3-fold axes (see Fig. 1.12b). As discussed above, to build this structure one needs two different types of body-triangles (see Table 1.3). One of the body triangles is determined again by T_1, as in the 5-fold situation. The second triangle can be defined by the body vector $\vec{C}_Q$ and a 120^o counterclockwise rotation of the cap vector (see Fig. 1.16)

$$\vec{C}_T^{120^o} = (-h-k)\vec{a}_1 + h\vec{a}_2 \equiv (-h-k, h). \tag{1.19}$$

This is because these two non-equivalent body-faces and three end-triangles from the cap must share a common vertex at the origin, which defines a pentamer. Therefore, to properly close the structure, we must introduce a 60^o wedge in the plane between the second triangle and the adjacent end-face, Fig. 1.16. The normalized surface of this second triangle is then $|\vec{C}_Q \times \vec{C}_T^{120^o}|/2S_0$, which defines

a second triangulation number for the body

$$T_2 = hh' + kh' + kk' \equiv Q_2 f \tag{1.20}$$

where

$$Q_2 = h_0 h' + k_0 h' + k_0 k'. \tag{1.21}$$

Since the 3-fold prolate has two non-equivalent body-triangles, it is convenient to characterize the structure by a single Q_{3F}-number obtained by summing up their surfaces,

$$Q_{3F} \equiv Q_1 + Q_2 = h_0(2h' + k') + k_0(2k' + h'). \tag{1.22}$$

For the spherical case this reduces to

$$Q^0_{3F} = 2Pf = 2T_{end}/f \tag{1.23}$$

which is twice the value obtained for Q^0_{5F}, Eq. (1.13). This is a direct consequence of having two non-equivalent body faces. Then, an isomeric $T = 3$ structure is equivalent to $T_{end} = 3$ and $Q_{3F} = 6$ in the 3-fold prolate description.

The surface of the capsid determines the total number of subunits as in the 5-fold case. Now we have eight end-triangles with T_{end}, six mid-triangles with T_1, and another six with T_2, which leads to

$$N^{3F}_{sub} = 3(8\,T_{end} + 6\,T_1 + 6\,T_2) = 6f(4Pf + 3Q_{3F}) \tag{1.24}$$

Since the twelve pentamers of a prolate account for sixty proteins, the number of hexamers is given by $N^{3F}_H = 4\,T_{end} + 3(T_1 + T_2) - 10$, and the total number of capsomers reads

$$N_{3F} = 4\,T_{end} + 3(T_1 + T_2) + 2 = f(4Pf + 3Q_{3F}) + 2. \tag{1.25}$$

As before, the value of Q_{3F} controls the length of the shell and can only adopt a discrete set of values. In this case the minimum step possible in Q_{3F} is (as derived in Appendix B)

$$\Delta Q^{min}_{3F} = \begin{cases} 3 & \text{if } |h_0 - k_0| \propto 3 \\ 1 & \text{otherwise} \end{cases} \tag{1.26}$$

Here, $|h_0 - k_0| \propto 3$ indicates that the difference $|h_0 - k_0|$ must either be zero or a multiple of 3. Consequently, the number of capsomers for 3-fold prolates is discretized by steps of

$$\Delta N^{min}_{3F} = 3f\Delta Q^{min}_{3F} = \begin{cases} 9f & \text{if } |h_0 - k_0| \propto 3 \\ 3f & \text{otherwise} \end{cases} \tag{1.27}$$

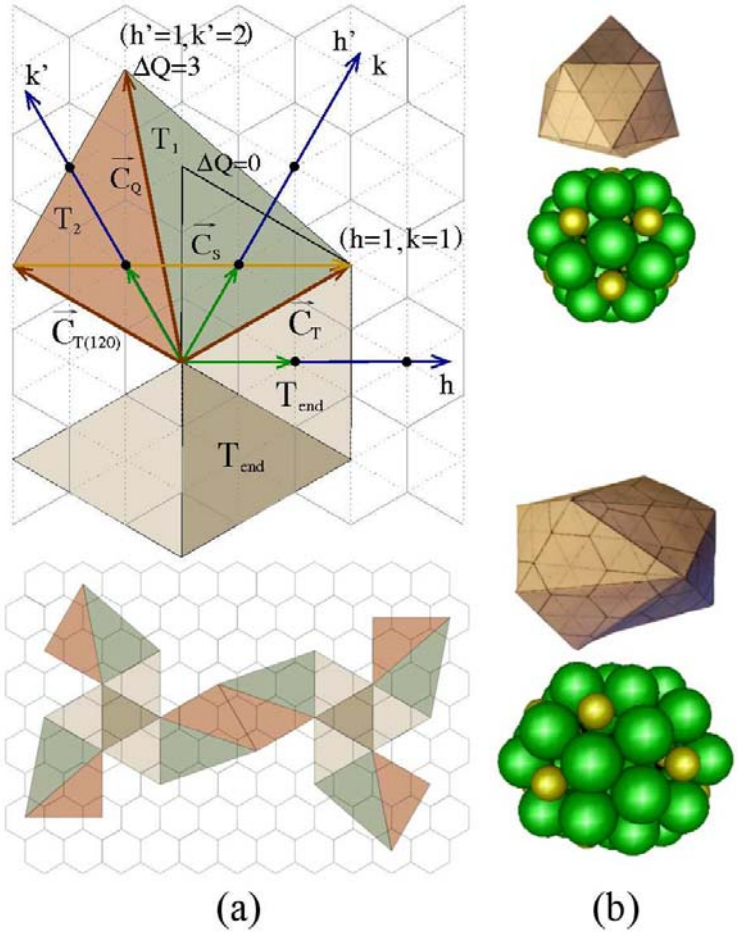


Figure 1.16: (a) (top) Basic elements to build a prolate based on hemispherical icosahedral caps centered on a 3-fold axis. The vector $\vec{C}_s = \vec{C}_h/3$ (in yellow) joins two consecutive pentamers along the rim of the cap. (bottom image) Complete flat design of the prolate with the 8 end-triangles and the 12 body-triangles. (b) Zenithal (top image) and lateral (bottom image) view of the resulting folded structure, along with its ping-pong model representation. The case illustrated in this figure corresponds to a $T_{end} = 3$ $Q_{3F} = 9$.

capsomers.

Accordingly, there are two different situations depending on the value of $|h_0 - k_0|$. In particular, for the class $P = 1$ ($h_0 = 1$,$k_0 = 0$) one gets $\Delta N_{3F} = 3f$, hence the possible lengths of a $T_{end} = 1$ ($f = 1$) capped shell are discretized by steps of at least $\Delta N_{3F}^{min} = 3$ capsomers, i.e., 18 proteins. This *growing law* agrees with the results obtained for AMV [42] and supports its classification as a 3-fold $T_{end} = 1$ bacilliform particle. On the other hand, a $P = 3$ ($h_0 = k_0 = 1$) prolate has $\Delta N_{3F}^{min} = 9f$. Thus, the possible lengths of a $T_{end} = 3$ ($f = 1$) capsid correspond to multiples of $\Delta N_{3F}^{min} = 9$ capsomers. In fact, rice tungro bacilliform virus (RTBV) has been suggested to be a 3-fold $T_{end} = 3$ prolate [43]. We will discuss these cases in detail in section 1.3.3.

2-fold prolates

This case is the most complicated, because the body has in general three non-equivalent mid-triangles (see Table 1.3). As mentioned in Section 1.3.1, there are two possible ways to make a 2-fold prolate. We will focus on the 2-fold *dextro* construction corresponding to Fig. 1.12c, although the arguments are valid for the *laevo* case as well (see Fig. 1.12d). The body of a 2-fold prolate is determined by three non-equivalent mid-triangles, Fig. 1.17. The first body-face is again the T_1-triangle. The second mid-triangle is the T_2-triangle introduce in the 3-fold case. The third mid-triangle is a translation by $\vec{C}_T$ of the first one and has the same triangulation number T_1, but they are not equivalent because is not possible to relate each other applying only the symmetries of a 2-fold prolate.

The area of the three non-equivalent body triangles is $2T_1 + T_2$, so in this case it is convenient to define the following Q-number

$$Q_{2F} \equiv 2Q_1 + Q_2 = h_0(3h' + 2k') + k_0(3k' + h') \tag{1.28}$$

As for the other prolates, the value of Q_{2F} specifies the length of the capsid, which for the spherical structure reduces to

$$Q_{2F}^0 = 3Pf = 3T_{end}/f \tag{1.29}$$

Since the body is determined by three non-equivalent mid-triangles, this value is three times Q_{5F}^0, Eq. (1.13). Thus, a $T = 3$ icosahedral capsid corresponds to $T_{end} = 3$ and $Q_{2F}^0 = 9$.

The total number of proteins of the prolate can be computed as in the 5-fold and 3-fold cases. Now we have eight end-triangles in the caps with T_{end}, and four

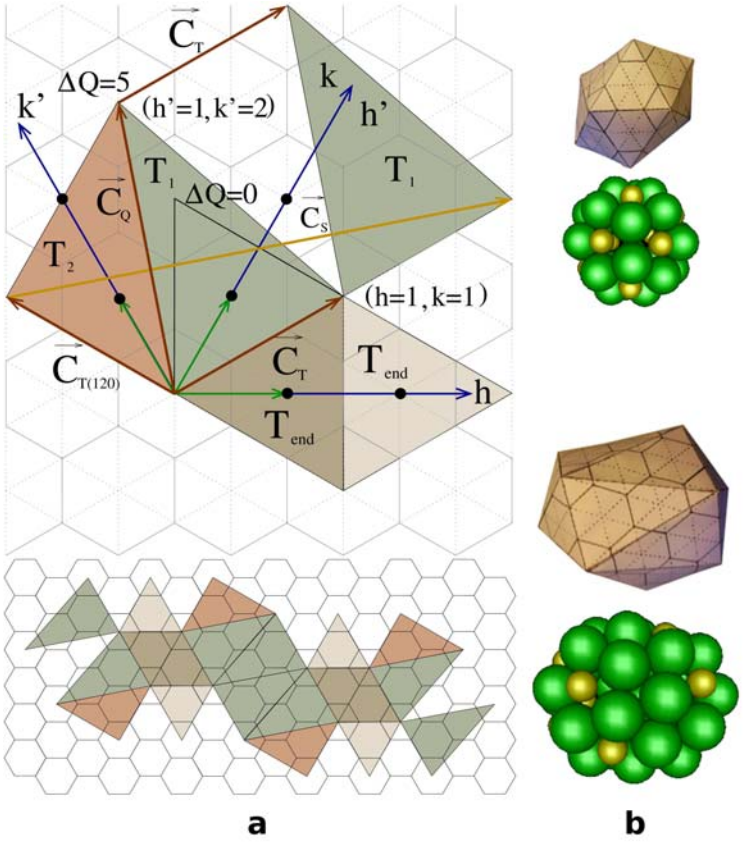


Figure 1.17: (a) (top) Basic elements required to build a prolate capsid with 2-fold axial symmetry. The vector $\vec{C}_s = \vec{C}_h/2$ (in yellow) joins two consecutive pentamers at the rim of the cap. (bottom) Complete flat design of the prolate with the 8 end-triangles and the 12 body-triangles. (b) Zenithal (top image) and lateral (bottom image) view of the resulting folded structure, along with its ping-pong model representation. The case illustrated in this figure corresponds to a $T_{end} = 3$ and $Q_{2F} = 14$ ($N = 42$)

.

of each of the three mid-triangles in the body, where two of them share the same T_1 number. Thus one obtains

$$N_{sub}^{2F} = 3(8T_{end} + 8T_1 + 4T_2) = 12f(2Pf + Q_{2F}) \tag{1.30}$$

Again, the 2-fold prolate has twelve pentamers accounting for 60 proteins, hence the number of hexamers is $N_H^{2F} = 4T_{end} + 2(2T_1 + T_2) - 10$, and the total number of capsomers is given by

$$N_{2F} = 4T_{end} + 2(2T_1 + T_2) + 2 = 2f(2Pf + Q_{2F}) + 2. \tag{1.31}$$

As in the 5-fold and 3-fold constructions, the value of Q_{2F} determines the length of the prolate, which can grow at discretized steps of (see Appendix B)

$$\Delta Q_{2F}^{min} = \begin{cases} 7 & \text{if } |h_0 - 2k_0| \propto 7 \\ 1 & \text{otherwise} \end{cases} \tag{1.32}$$

or, in terms of capsomers,

$$\Delta N_{2F}^{min} = 2f\Delta Q_{2F} = \begin{cases} 14f & \text{if } |h_0 - 2k_0| \propto 7 \\ 2f & \text{otherwise} \end{cases} \tag{1.33}$$

One then has two different cases depending on the value of h_0 and k_0. In particular for the class $P = 1$ ($h_0 = 1$,$k_0 = 0$) or $P = 3$ ($h_0 = k_0 = 1$) the growing law is $\Delta N_{2F}^{min} = 2f$, hence the possible number of capsomers of a $T_{end} = 1$ or a $T_{end} = 3$ capped shell is discretized at intervals of 2 capsomers (see Fig. 1.18). On the other hand, for $P = 7l$ with ($h_0 = 2$, $k_0 = 1$) the minimum step is $\Delta N_{2F}^{min} = 14f$, but for the specular case ($h_0 = 1$, $k_0 = 2$) one obtains $\Delta N_{2F}^{min} = 2f$. Thus the possible lengths of a shell based on a $T_{end} = 7l$ are discretized by steps of 14 capsomers, whereas for a $T = 7d$ the minimum step is of 2 hexamers. Hence, for a 2-fold prolate the two isomers of a chiral cap can have different associated bodies, and the handedness plays an important role. This occurs because the caps are already skewed (see Figs. 1.12c and d). In addition, in the *laevo* 2-fold construction, by symmetry, one would obtain the same results but interchanging h_0 and k_0 and so the handedness of the solutions.

Finally, we are not aware of any prolate virus which is known for sure to have a 2-fold construction. However, it is possible that a particle of AMV [44] might be the case, as discussed in Section 1.3.3.

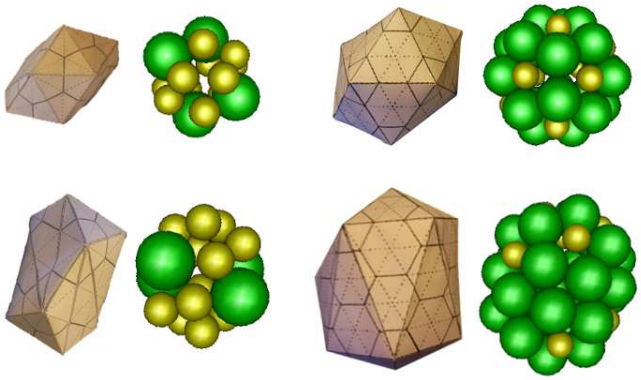


Figure 1.18: Examples of 2-fold elongated structures. (a) Second 2-fold elongation of a $T = 1$ prolate, i.e., $\Delta Q_{2F} = 2$. (b) Fifth elongation of a 2-fold $T = 3$ bacilliform structure, i.e., $\Delta Q_{2F} = 5$.

1.3.2 Tubular body description

The generalized model of elongated capsids introduced above allows us to enumerate all possible icosahedral prolates. In this section, we will describe a procedure to compute, for any icosahedrally-capped shell, the radius, the length, and the position of the capsomers in the body. This geometrical characterization of the resulting capsids was carried out neither in CK nor in Moody's model, but it turns out to be quite helpful for both recognition and design of viral shells.

The tubular body of an elongated virus can be built by rolling up an hexagonal sheet, much in the same way as with carbon nanotubes [45] (see Fig. 1.19). However, for prolate viral capsids only the subset of tubes closed by icosahedral caps with 5-fold, 3-fold, or 2-fold axial symmetry are valid.

This procedure involves an approximation, since one assumes that the surface of the resulting cylinder is the same as that of the flat lattice, which implies that hexamers will be bent and stretched in the tube. However, it can be shown that even for the smallest capsids this turns out to be a very good approximation (see Appendix C).

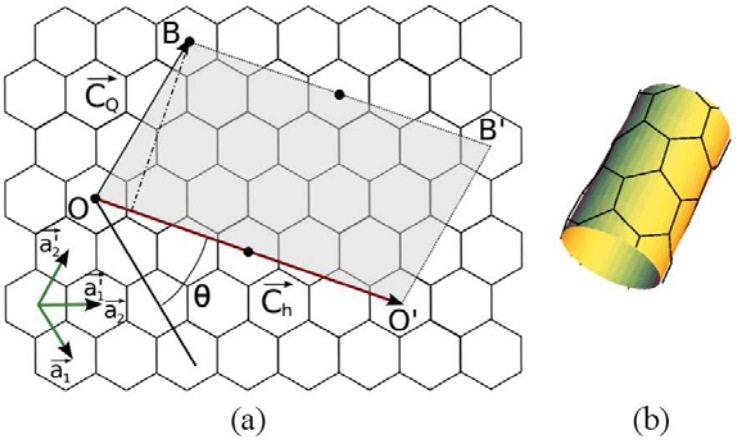


Figure 1.19: (a) The unrolled tubular body of a prolate virus (shaded area) shown on the honeycomb lattice. The solid dots indicate the location of the pentamers at the rim. The chiral angle θ between $\vec{C}_h$ and $\vec{a}_1$ of the hexagonal lattice characterizes the arrangement of hexamers in the body. A body made of rings of hexamers is obtained when $\theta = 0^o$ (i.e., $n = 0$), whereas zigzag tubes correspond to $\theta = 30^o$ (or $m = n$). (b) Tubular body obtained by rolling up the shaded area in the direction of $\vec{C}_h$ so that O meets O' and B meets B'. The example corresponds to the body of a $T_{end} = 1$ $Q_{2F} = 9$ 2-fold prolate, with $h = 1$, $k = 0$, and $h' = 3$,$k' = 0$.

Radius of the tube

In the tubular description the properties of the body can be evaluated from the associated flat hexagonal representation (see Fig. 1.19). Let's define $\vec{C}_h$ as the chiral vector that determines the circumference of the tube

$$\vec{C}_h = m\vec{a_1} + n\vec{a_2} \equiv (m, n) \tag{1.34}$$

This vector belongs to the hexagonal lattice, i.e., m and n are integers, and connects all pentamers along the rim of the cap. Hence, $\vec{C}_h$ must be related to the cap vector $\vec{C}_T$, even though the specific relation will depend on the axial symmetry.

Let's discuss first the case for a 5-fold capsid, which is particularly simple. It is clear from Figs. 1.12a and 1.13 that the rim of the tube is delimited by five vertexes, or pentamers, of the icosahedral cap, and they all lie on a plane perpendicular to the axis of the prolate. In addition, the cap vector $\vec{C}_T$, Eq.

(1.1), connects two adjoining pentamers on this boundary, so by unrolling the body of the capsid it is easy to realize that the circumference of the tube is just made by five times the cap vector, and so

$$\vec{C}_h^{5F} = 5\vec{C}_T = 5f(h_0, k_0) \tag{1.35}$$

In particular, for prolates of the class $P = 1$ one gets $\vec{C}_T^{5F} = (5f, 0)$, and for the class $P = 3$ one has $\vec{C}_T^{5F} = (5f, 5f)$.

For a 3-fold bacilliform capsid, the circumference of the cap is defined by three non-consecutive vertexes (pentamers), which lie again in a section perpendicular to the axis (see Fig. 1.12b). The vector that connects two of these vertexes lying on the rim is plotted in Fig. 1.16 and can be expressed in terms of the cap vector as $\vec{C}_T - \vec{C}_T^{120^o}$. Now, the chiral vector is just obtained by summing up three times this vector

$$\vec{C}_h^{3F} = 3(\vec{C}_T - \vec{C}_T^{120^o}) = 3f(2h_0 + k_0, k_0 - h_0). \tag{1.36}$$

Similarly, in the 2-fold case the circumference of the tube is defined by the two pentamers which are farther apart and lie on a section perpendicular to the axis (see Fig. 1.12c). As one can see in Fig. 1.17, the rim vector that connects the two pentamers in the unrolled body is $2\vec{C}_T - \vec{C}_T^{120^o}$. Hence, the chiral vector is made by two times the rim vector, i.e.,

$$\vec{C}_h^{2F} = 2(2\vec{C}_T - \vec{C}_T^{120^o}) = 2f(3h_0 + k_0, 2k_0 - h_0) \tag{1.37}$$

For symmetry reasons the specular construction defined in Fig. 1.12d leads to the same results but permuting h_0 and k_0.

Since the length of the chiral vector is the circumference of the tubular body, the radius of a prolate is given by the simple formula

$$R = \frac{|\vec{C}_h|}{2\pi} \tag{1.38}$$

Computing now the absolute value of $\vec{C}_h$ for each symmetry, Eqs. (1.35), (1.36), and (1.37) (see also Appendix D), one obtains the general expressions for the radius of all possible prolate viruses based on icosahedral caps

$$R_{5F} = \frac{5}{2\pi}\sqrt{T_{end}}\, a \tag{1.39a}$$

$$R_{3F} = \frac{3}{2\pi}\sqrt{3T_{end}}\, a \tag{1.39b}$$

$$R_{2F} = \frac{2}{2\pi}\sqrt{7T_{end}}\, a \tag{1.39c}$$

where a is the lattice parameter, i.e., the distance between hexamers. All radii scale with the square root of the cap surface, $R \sim \sqrt{T_{end}}$. Interestingly, if we compare the coefficients of the three cases for a given T_{end} we get the general inequality

$$R_{5F} \lesssim R_{3F} \lesssim R_{2F} \tag{1.40}$$

Hence, even though the three radii are similar, the 5-fold prolate is the smallest, followed by the 3-fold and finally the 2-fold structure.

Chiral angle and distribution of hexamers in the body

The distribution of hexamers in the body of a prolate is determined by the chiral angle θ, which is the angle between the chiral vector $\vec{C}_h$ and the first vector $\vec{a_1}$ of the hexagonal lattice, Fig. 1.19. Accordingly

$$cos(\theta) = \frac{\vec{C}_h \cdot \vec{a}_1}{|\vec{C}_h|} = \frac{2m+n}{2\sqrt{m^2+mn+n^2}} \tag{1.41}$$

where, due to the symmetry of the hexagonal lattice, θ is defined between 0^o and 60^o.

In terms of the arrangement of hexamers in the body, there are two specially symmetric situations (see Fig. 1.20). The cases $n = 0$, corresponding to $\theta = 0^o$, or $m = 0$, yielding $\theta = 60^o$, both lead to a tubular body where the hexamers are arranged in rings; whereas $m = n$ corresponds to a chiral angle $\theta = 30^o$ and a distribution of hexamers in zigzag rows[5]. Any other case will lead to a skewed arrangement of body hexamers.

Using the expressions of the chiral body vector for the different symmetries, Eqs. (1.35), (1.36), and (1.37), we can explore which caps are associated to the symmetrical bodies described above (Table 1.4). In particular, rings are obtained for $P = 1$ 5-fold, $P = 3$ 3-fold, and $P = 7l$ 2-fold, whereas zigzag tubes appear for $P = 1$ 3-fold, $P = 3$ 5-fold, and $P = 21d$ 2-fold. Notice that to enlarge a prolate one needs a minimum number of capsomers that have to respect the axial symmetry. For the symmetrical bodies one then has to add a complete ring or half of a zigzag to increase the length of the structure, respectively. Thus, these type of bodies are clear examples of the division in cases observed in the growing laws for ΔN^{min}. Although we can also find skewed examples, for instance, a prolate based on $h_0 = 4$, $k_0 = 1$ ($P = 21$) cap in the 3-fold configuration has a chiral angle $\theta \approx 19^0$, but follows $\Delta N^{min} = 9f$, like for the ring body case R_9 ($P = 3$) commented above.

[5]Note that in carbon nanotubes the nomenclature is different: rings and zigzag layers of hexamers correspond to zigzag and armchair nanotubes, respectively [45].

(h_0, k_0)	P	S	N_0	$\frac{\Delta N}{f}$	$\frac{\pi R}{af}$	$\frac{\Delta L^{min}}{a}$	$\theta(^o)$	Body
		5		5	$\frac{5}{2}$	$\frac{1}{2}\sqrt{3}$	0	R_{5f}
$(1,0)$	1	3	$10f^2+2$	3	$\frac{3}{2}\sqrt{3}$	$\frac{1}{2}$	30	Z_{6f}
		2		2	$\sqrt{7}$	$\frac{1}{2}\sqrt{\frac{3}{7}}$	19.11	skew
		5		5	$\frac{5}{2}\sqrt{3}$	$\frac{1}{2}$	30	Z_{10f}
$(1,1)$	3	3	$30f^2+2$	9	$\frac{9}{2}$	$\frac{1}{2}\sqrt{3}$	0	R_{9f}
		2		2	$\sqrt{21}$	$\frac{1}{2\sqrt{7}}$	10.89	skew
		5		5	$\frac{5}{2}\sqrt{7}$	$\frac{1}{2}\sqrt{\frac{3}{7}}$	19.11	skew
$(2,1)$	7	3	$70f^2+2$	3	$\frac{3}{2}\sqrt{21}$	$\frac{1}{2\sqrt{7}}$	10.89	skew
		2		14	7	$\frac{1}{2}\sqrt{3}$	0	R_{14f}
$(1,2)$		2		2	7	$\frac{1}{14}\sqrt{3}$	21.79	skew
		5		5	$\frac{5}{2}\sqrt{13}$	$\frac{1}{2}\sqrt{\frac{3}{13}}$	13.90	skew
$(3,1)$	13	3	$130f^2+2$	3	$\frac{3}{2}\sqrt{39}$	$\frac{1}{2\sqrt{13}}$	16.10	skew
		2		2	$\sqrt{91}$	$\frac{1}{2}\sqrt{\frac{3}{91}}$	5.21	skew
$(1,3)$		2		2	$\sqrt{91}$	$\frac{1}{2}\sqrt{\frac{3}{91}}$	27.00	skew

Table 1.4: Expressions for the most important properties of an elongated virus particularized for each axial symmetry s (5-fold, 3-fold and 2-fold) and the lowest classes P. N_0 is the number of capsomers in the spherical seed of the prolate; ΔN^{min} is the smallest step in number of capsomers between different lengths of the prolate; R is the radius of the prolate in the tubular approximation; ΔL^{min} is the smallest length step; θ is the chiral angle, which determines the orientation of hexamers in the tubular part of the prolate (the chiral angles of specular skew structures are related by $\theta \to -\theta$, i.e. $60^o - \theta$, and for 2-fold prolates this implies also the transformation of the intrinsic chirality); and the column "Body" classifies the arrangement of hexamers in the body as rings R_i or zigzags Z_i of i hexamers, and skewed structures.

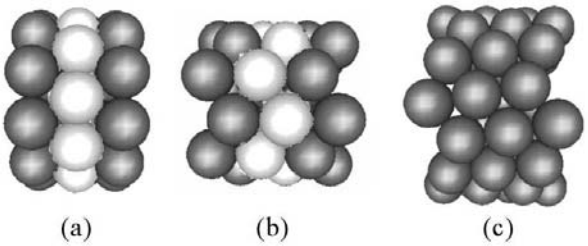


Figure 1.20: Symmetrical and skew tubular bodies. There are two possible non-chiral tubular constructions, which are associated to prolates of different class P depending on the axial symmetry. (a) A tubular ring body is illustrated, where each layer, highlighted in white, is made of 10 capsomers, R_{10}. This tube is associated to prolates capped by $T_{end} = 4$ (5-fold). (b) The zig-zag tube shown is composed of layers made of 10 hexamers, Z_{10}, as the one highlighted. It corresponds to bacilliform capsids based on $T_{end} = 3$ (5-fold). (c) There are infinite types of chiral bodies. Here the case associated to a prolate capped by $T_{end} = 7l$ (5-fold) is shown.

Smallest length step and particle length

The chiral vector controls the diameter of the prolate and the precise arrangement of hexamers in the body. However, the length of the capsid is controlled by the body vector $\vec{C}_Q$, which connects two closest pentamers from different caps (see Fig. 1.19). By combining both vectors it is possible to compute the remaining relevant properties of a bacilliform shell.

In particular, the length of the tubular body will be given by the perpendicular projection of $\vec{C}_Q$ onto $\vec{C}_h$ (see Fig. 1.19), namely,

$$L = \frac{|\vec{C}_h \times \vec{C}_Q|}{|\vec{C}_h|} = \frac{\sqrt{3}}{2}\frac{Q}{|\vec{C}_h|} sf\, a^2 \tag{1.42}$$

where the value of s can be 5, 3, or 2 in consonance with the axial symmetry, and Q and $\vec{C}_h$ have different expressions depending on s, as we have seen in the previous sections. For every symmetry, Q is discretized by ΔQ^{min}, so we get straightforwardly the discretization of the lengths of a prolate

$$\Delta L^{min} = \frac{\sqrt{3}}{2}\frac{\Delta Q^{min}}{|\vec{C}_h|} sf\, a^2 = \frac{\sqrt{3}}{2}\frac{\Delta N^{min}}{2\pi R}\, a^2 \tag{1.43}$$

that can be also be expressed in terms of ΔN^{min}. Now, if we apply the results obtained for each axial symmetry we obtain

$$\Delta L_{5F}^{min} = \frac{\sqrt{3}}{2\sqrt{P}}\, a \tag{1.44a}$$

$$\Delta L_{3F}^{min} = \frac{\Delta Q_{3F}^{min}}{2\sqrt{P}}\, a \tag{1.44b}$$

$$\Delta L_{2F}^{min} = \frac{\sqrt{21}}{14}\frac{\Delta Q_{2F}^{min}}{\sqrt{P}}\, a \tag{1.44c}$$

ΔL^{min} defines also the shortest tube and is directly related to the elongation of the body, Eq. (1.14),

$$\Delta L = \Delta L^{min}\left(Q - Q^0\right) = \Delta L^{min}\Delta Q \tag{1.45}$$

which becomes zero for a spherical capsid, when $Q = Q^0$. In Table 1.5 the values of these properties for each axial symmetry are summarized.

Finally, the aspect ratio of a prolate can also be computed dividing the total length of the particle (including the caps), $L = 2R + \Delta L$, by its width, $2R$, i.e.,

$$ar = \frac{2R + \Delta L}{2R} = 1 + \frac{\Delta L}{2R} \tag{1.46}$$

The length and radius of a capsid depends on the separation between capsomers, a, and thus on the size of the proteins. Contrarily, the aspect ratio is a dimensionless quantity, that could be compared directly, independently of the particular dimensions of a virus.

Finally, the position of all capsomers in the tube can also be evaluated, as shown in Appendix E.

Degeneracy and number of relative orientations

Interestingly, some prolate structures could be built in different ways. Here the term *degenerate* refers to architectures that, for a particular axial symmetry, have the same T_{end} and Q numbers, i.e., radius and length, but differ in the relative orientation of pentamers in both caps. Therefore, we will refer to them as prolate *isomers* (see Fig. 1.21).

The degeneracy arises when there is more than one possible choice of the body vector $\vec{C}_Q$ with the same perpendicular projection onto $\vec{C}_h$ (see Fig. 1.21). This occurs when there is more than one lattice point in the segment $\vec{C}_h/s$ that

Property	General	5-fold	3-fold	2-fold
Q		$h_0(h'+k')+k_0k'$	$h_0(2h'+k')+k_0(2k'+h')$	$h_0(3h'+2k')+k_0(3k'+h')$
N_{sub}		$30f(Pf+Q_{5F})$	$6f(4Pf+3Q_{3F})$	$12f(2Pf+Q_{2F})$
N	$\frac{N_{sub}}{6}+2$	$5f(Pf+Q_{5F})+2$	$f(4Pf+3Q_{3F})+2$	$2f(2Pf+Q_{2F})+2$
ΔQ^{min}	$\frac{gcd(m,n)}{sf}$	1	$\begin{cases}3 & \text{if } \lvert h_0-k_0\rvert \propto 3\\ 1 & \text{the rest}\end{cases}$	$\begin{cases}7 & \text{if } \lvert h_0-2k_0\rvert \propto 7\\ 1 & \text{the rest}\end{cases}$
ΔN^{min}	$sf\Delta Q^{min}$	$5f$	$3f\Delta Q^{min}_{3F}$	$2f\Delta Q^{min}_{2F}$
$\vec{C}_h$	(m,n)	$5f(h_0,k_0)$	$3f(2h_0+k_0,k_0-h_0)$	$2f(3h_0+k_0,2k_0-h_0)$
$cos\theta$	$\frac{2m+n}{2\sqrt{m^2+mn+n^2}}$	$\frac{2h_0+k_0}{2\sqrt{P}}$	$\frac{\sqrt{3}}{2}\frac{h_0+k_0}{\sqrt{P}}$	$\frac{5h_0+4k_0}{2\sqrt{7P}}$
$\frac{R}{a}$	$\frac{\lvert\vec{C}_h\rvert}{2\pi a}$	$\frac{5}{2\pi}\sqrt{T_{end}}$	$\frac{3}{2\pi}\sqrt{3T_{end}}$	$\frac{2}{2\pi}\sqrt{7T_{end}}$
$\frac{\Delta L^{min}}{a}$	$\frac{\sqrt{3}}{2}\frac{\Delta N^{min}}{2\pi R}a$	$\frac{\sqrt{3}}{2\sqrt{P}}$	$\frac{\Delta Q^{min}_{3F}}{2\sqrt{P}}$	$\frac{\sqrt{21}}{14}\frac{\Delta Q^{min}_{2F}}{\sqrt{P}}$
$\frac{\Delta L}{a}$	$\frac{\Delta L^{min}}{a}\Delta Q$	$\frac{\Delta L^{min}_{5F}}{a}(Q_{5F}-Pf)$	$\frac{\Delta L^{min}_{3F}}{a}(Q_{3F}-2Pf)$	$\frac{\Delta L^{min}_{2F}}{a}(Q_{2F}-3Pf)$
D	$f\Delta Q^{min}$	f	$f\Delta Q^{min}_{3F}$	$f\Delta Q^{min}_{2F}$

Table 1.5: General expressions for each axial symmetry of the most important quantities and properties of elongated capsids in terms of the values of (h_0,k_0) and f of the caps. Q characterizes the body; N_{sub} is the total number of subunits; N is the total number of capsomers; ΔQ^{min} and ΔN^{min} are the minimum step in Q and in number of capsomers, respectively, for a given cap; $\vec{C}_h$ is the chiral vector; θ is the chiral angle; R is the radius of the tubular part in units of the distance between hexamers a; ΔL^{min} is the minimum length step; ΔL is the elongation of a particle with respect to the spherical capsid; and D is the number of degenerate structures.

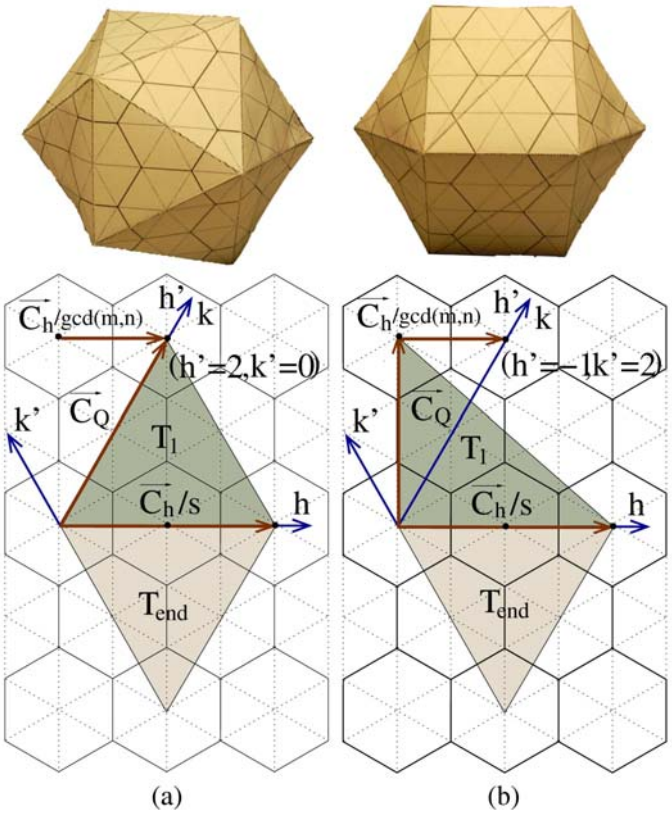


Figure 1.21: The two possible degenerate spherical structures for a $T_{end} = 4$ with 5-fold axial symmetry. The top images show the polyhedral shell and the bottom ones, their construction in a flat hexagonal lattice following the same ideas as in Fig. 1.13. (a) Classical isomeric case that corresponds to the CK icosahedral capsid. (b) Quasi-spherical structure that has the same radius, length, and number of proteins as in (a), but with a different relative position of pentamers between both caps.

joins two consecutive pentamers on the rim, offering then the chance of choosing more than one relative orientation between the caps for a given tubular length. Mathematically, the degeneracy, i.e., the number of different structures with the same T_{end}, axial symmetry, and length, can be calculated as the ratio between $\vec{C}_h/s$ and the smallest lattice vector in the direction of $\vec{C}_h$, namely,

$$D \equiv \frac{|\vec{C}_h|/s}{|\vec{C}_h|/gcd(m,n)} = f\,\Delta Q^{min} \tag{1.47}$$

Then D determines the number of isomers of a prolate.

Accordingly, all elongated structures based on icosahedral caps with $f > 1$ or $\Delta Q^{min} > 1$ are degenerate. In the 5-fold case $\Delta Q^{min}_{5F} = 1$, so structures with $f = 1$, e.g, $T_{end} = 1, 3, 7, 13\ldots$, have always a unique prolate capsid, whereas for $f > 1$, e.g, $T_{end} = 4, 9, 12$, there are f possible isomers with different relative orientations between the pentamers of the two caps. We stress that this occurs even for the spherical case when $T_{end} = T_1$. For instance, a $T_{end} = 4$ ($f = 2$) can adopt two spherical configurations (see Fig. 1.21), but only one has full icosahedral symmetry [46]. In fact, we will see in the next chapter that for $T = 4$ the non-icosahedral structure is feasible and has the same free energy as the icosahedral one.

Interestingly, elongated structures with 3-fold and 2-fold axial symmetry can be degenerate even for $f = 1$, but only for classes P with $\Delta Q^{min} > 1$. For instance, the $T_{end} = 3$ (3-fold) leads to $\Delta N^{min}_{3F} = 9$, or equivalently $\Delta Q^{min} = 3$. Thus the total number of isomers will be $D = 3$, according to Eq. (1.47). Note that the body of this prolate is a ring made of 9 capsomers (R_9) (see Fig. 1.16b), and since the axial symmetry is 3 is completely natural to have 3 degenerated structures for a given length. Again, this holds even for the non-elongated capsids, leading to 3 possible structures for a spherical $T = 3$ capsid (but only one strictly icosahedral) when a hemi-icosahedral cap is rotated around the 3-fold axis. Another example is the elongated architecture based on $T_{end} = 7d$ with 2-fold axial symmetry and the choice of caps of Fig. 1.12c. In this case $\Delta N^{min}_{2F} = 14$ or $\Delta Q^{min} = 7$, so we obtain $D = 7$ possible isomers. We remark that the body in this case corresponds to a ring of 14 capsomers (R_{14}), i.e., seven times the axial symmetry[6].

[6]Contrarily to the $T = 4$ (5-fold) case, the rotations of spherical structures around the 3-fold and 2-fold axis (when it is geometrically possible) present some energetic disadvantages. Except for the icosahedral structure, the rest of isomers usually bring two pentamers of opposite caps very close. This represents an important energetic cost, which will increase the elastic energy of the shell [47], because topological defects on the hexagonal lattice tend to repel each other.

1.3.3 Applications: structural characterization of prolate viruses

In this section, we will study several viruses to illustrate how the insights gained in this work can be useful for characterization purposes.

Bacteriophage T4

Bacteriophage T4 is one of the few prolate viruses whose capsid structure has been determined at high resolution [6] (see Fig. 1.22). This case represents an excellent example to illustrate how to infer the structure of a bacilliform capsid using few *raw* experimental data.

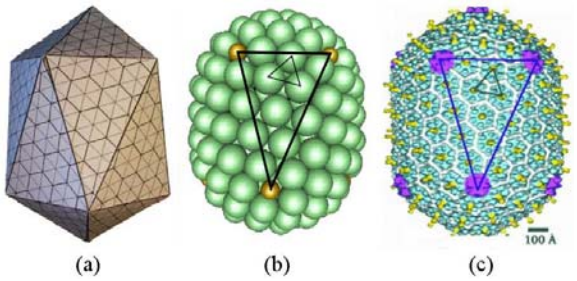


Figure 1.22: Bacteriophage T4. (a) Geometrical construction for a 5-fold prolate $T = 13$ and $Q_{5F} = 20$, which is made of 167 capsomers. (b) The correspondent ping-pong model built using the tubular body approximation combined with hemispherical icosahedral caps. (c) Bacteriophage T4 cryo-EM reconstruction [6].

In particular, we will use the experimental diameter of the capsid, $2R_{exp} = 86 \pm 3\,nm$, and the distance between hexamers in the body, $a^{exp} = 14 \pm 2\,nm$, both obtained from Ref. [6]. Inserting these data in the formulas of the radius for the different symmetries listed in Table 1.5, one obtains $T_{5F} = 14.9 \pm 1.6$, $T_{3F} = 13.8 \pm 1.5$, and $T_{2F} = 13.3 \pm 1.4$ as potential values for the triangulation number of the cap. Since, in principle, $T_{end} = 14$ or 15 are not valid results, the triangulation number should be either $T_{end} = 13$ or 16.

In addition, the shell is composed by 167 capsomers: 155 hexamers made of 930 copies of $gp23^*$, 11 pentamers made of $gp24^*$, and the $gp20$ connector[7], which can be assumed to play the role of an effective pentamer in the model. Taking into account the growing laws for the different T_{end}-caps proposed in Table 1.6, it is evident that only $T_{end} = 13$ with 5-fold axial symmetry leads to a capsid with

[7] *gp* means gene product of the viral genome.

T	N_0	ΔN_{5F}^{min}	ΔN_{3F}^{min}	ΔN_{2F}^{min}
1	12	5	3	2
3	32	5	9	2
13	132	5	3	2
16	162	20	12	8

Table 1.6: Number of capsomers N_0, Eq. (1.4), in an icosahedral capsid based on different T_{end}-caps, and the corresponding growing laws, ΔN^{min}, for each symmetry, Eqs. (1.18), (1.27), and (1.33).

167 capsomers. The number of hexamers involved in the elongation is 35 because the icosahedral shell has 132 capsomers, thus from Eq. (1.18) one gets $\Delta Q_{5F} = 7$. Taking into account that in the spherical case $Q_{5F}^0 = 13$, our analysis suggests that the structure of bacteriophage T4 is a 5-fold prolate with $T_{end} = 13$ and $Q_{5F} = 20$, which is in fact the structure resolved in the cryo-em reconstruction. We can also compute the aspect ratio using Eq. (1.46) obtaining a value of 1.3 in agreement with the experimental value, 1.4 ± 0.2.

Therefore, using three simple inputs, i.e., diameter, distance between hexamers, and total number of proteins, it has been possible to infer the structure of the virus. These data can be obtained from different experimental techniques, e.g., electron micrographs, optical diffraction, or sedimentation, but unfortunately seem not to be available in the literature for most viruses.

Alfalfa Mosaic Virus (AMV)

AMV is a well studied plant virus that adopts different lengths depending on the amount of genetic material encapsidated [41, 42] (see Table 1.7). The number of protein subunits in the *in vitro* reconstituted capsids has been determined from their molecular weights and correspond to $N_{sub} = 60$, 132, 150, 186, and 240 [42]. Only the smallest capsid has been reconstructed by X-ray [48] and is a spherical $T = 1$ composed of twelve pentamers. The elongated particles have a diameter similar to that of the icosahedral one, hence it is natural to assume that they might be based on $T_{end} = 1$ caps. Considering that the body of the prolate particles is formed by hexamers, then the number of proteins can easily be translated into number of capsomers, obtaining the series $N = 12$, 24, 27, 33, and 42. Interestingly, a multiple of at least three hexamers is added in every step, and according to our model, this can only be explained if AMV elongated particles adopt a structure $T_{end} = 1$ centered on a 3-fold symmetry axis (see Tables 1.6 and 1.7, and Fig. 1.23). Optical diffraction studies have revealed

$AMV_{particle}$	n	N_{sub}	N	$L_{exp}(nm)$	Q_{3F}
T_{a-b}	4	132	24	29.9 ± 2.4	6
T_b	5	150	27	34.8 ± 1.6	7
M	7	186	33	43.3 ± 2.0	9
B	10	240	42	56.1 ± 2.2	12

Table 1.7: Different experimental properties for the four particles of AMV: number of proteins in the shell, N_{sub}, which follows the law $60 + (18 \times n)$ [42], number of capsomers (N), length, and the inferred value of Q_{3F}.

that in fact AMV posses three-fold symmetry [41], and a Geodestix model of a similar architecture was suggested in Ref. [16]. However, the precise geometrical properties of this model were never indicated.

The predictions of our model are not only useful to infer the structure, but can also be used to extract other geometrical and structural information. For instance, the architecture proposed has a body made of hexamers arranged in a zigzag pattern Z_6, and the minimum step of $\Delta N_{3F} = 3$ capsomers corresponds to an increment in length of $\Delta L_{3f} = a/2$ (see Table 1.4). Experimentally, it is known that each step of 18 subunits, i.e, 3 hexamers, increases the length by $4.34\,nm$ [42]. Hence the distance between hexamers should be $a \simeq 8.68\,nm$, which is in agreement with diffraction analysis [41], and from that one can estimate, for instance, the typical size of a capsomer or a coat protein.

AMV also makes an alternative elongated particle that contains 120 protein subunits, and does not follow the sequence discussed above [44]. Assuming that the central body is built of hexamers, this number of proteins corresponds to $N = 22$ capsomers. In addition, the particle has again a similar radius suggesting that it is based on $T_{end} = 1$ caps. Hence, in the framework of our model there are two possible structures for this alternative particle (Table 1.6): a $T_{end} = 1$ bacilliform shell centered on a 5-fold axis with $Q_{5F} = 2$; and a $T_{end} = 1$ prolate centered on a 2-fold axis with $Q_{2F} = 5$. However, Ref. [44] suggests that the particles show an oblate shape. In that case the 2-fold situation seems a better candidate since 2-fold structures are quite distorted and could lead to deformed shapes. In any case, experimentally is not clear whether this particle is polymorphic.

Rice Tungro Bacilliform Virus (RTBV)

Rice tungro virus shows spherical and bacilliform shapes *in vivo* of similar diameter, but no structure of the virus has been fully determined yet. As in the case of AMV, let's explore some basic experimental results to predict a plausible

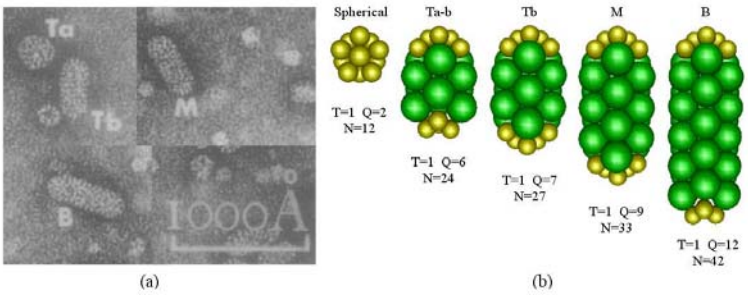


Figure 1.23: Alfalfa mosaic virus (AMV). (a) Electron micrograph showing AMV particles of spherical type, T_a, and some elongated ones, T_b, M, and B. The picture is a composition based on the research of Hull [41] (Fig. 3 therein). (b) Structures predicted by the geometrical model and obtained also as optimal structures in the simulations of the next chapter. The sequence corresponds to the size of AMV found in [42]. We notice that Hull obtained slightly different apparent molecular weights for the particles shown in (a) [41], although the organization of AMV in types of different lengths are qualitatively in agreement.

capsid architecture.

The diameter of the tubular part has been obtained from EM micrographs, and has a value $2R_{exp} = 30 \pm 3\, nm$. In addition, diffraction experiments suggest that the distance between hexamers in the body is $a_{exp} = 10 \pm 2\, nm$, and that capsomers are arranged in rings [43].

In the framework of our geometrical model, for each axial symmetry there is only one possible class P associated to a body made of hexamer rings. More specifically, $P = 1$, $P = 3$, and $P = 7l$ for 5-, 3-, and 2-fold symmetries, respectively (see Table 1.4). One can now use the experimental estimate of a_{exp} to calculate what would be the expected radius of the prolate for each situation. Thus, the values $2R_{5F}^{P=1} = f(16 \pm 2)\, nm$, $2R_{3F}^{P=3} = f(28 \pm 3)\, nm$, and $2R_{2F}^{P=7} = f(44 \pm 9)\, nm$, are obtained in each case. Comparing with the experimental value we can conclude that RTBV is either based on a $T_{end} = 4$ (5-fold) or a $T_{end} = 3$ (3-fold) structure, with a body made of rings of ten, R_{10}, or nine hexamers, R_9, respectively.

The lack of further experimental information does not allow us to discriminate between both possibilities. However, our model can be used to predict what would be the expected geometrical properties of the virus in each case. If RTBV is based on a $T_{end} = 4$ (5-fold), its number of subunits should follow the law $N_{sub}^{5F} = 240 + (n \times 60)$, and its total length should be $L(n) = 2R + n\Delta L$, where $\Delta L = 8.7 \pm 1.7\, nm$

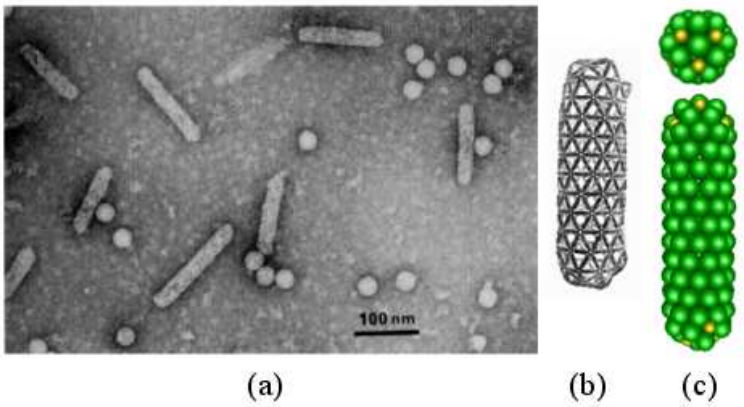


Figure 1.24: Rice tungro bacilliform virus (RTBV). (a) Electron micrograph showing spherical and bacilliform particles of rice tungro virus [43]. (b) Geodestix model suggested by Hull based on the diffraction patterns from the tubes in the electron micrographs (see Fig. 3 [43]). (c) Tubular construction of the predicted bacilliform structure based on our analysis. This corresponds to a 3-fold $T = 3$ $Q_{3F} = 39$ ($N = 131$). On the top we have a view of the cap, and at the bottom we show the body.

is the same for both architectures because they have ring-bodies. Experimentally, the length of the predominant particle is $L_{exp} = 130 \pm 3\, nm$. Therefore, we obtain a value of $n = 11 \pm 3$ for the number of steps, so taking into account that $Q^0_{5F} = 2$, leads to $Q_{5F} = 13 \pm 3$. Thus, the structure would have $N = 152 \pm 30$ capsomers or $N_{sub} = 900 \pm 180$ proteins. Analogously, if RTBV is based on a $T_{end} = 3$ (3-fold) architecture, one would obtain a structure characterized by $N = 131 \pm 30$, $Q_{3F} = 39 \pm 9$, and $N^{3F}_{sub} = 180 + (n \times 54) = 770 \pm 160$ protein subunits. Notice that, just by knowing the total number of proteins or the molecular weight of the capsid, one could know which is the right structure.

Again, based on diffraction experiments, Hull proposed that RTBV might be an elongated particle based on $T_{end} = 3$ (3-fold) [43], which is one of the solutions of our analysis (see Fig. 1.24). However, from the experimental data used above we cannot reject the $T_{end} = 4$ (5-fold) architecture.

Bacteriophage P22 polyheads

Bacteriophage P22 has been intensively investigated, and constitutes one of the most well known models of viral assembly [49, 50]. The coat subunits form an

empty $T = 7$ spherical capsid, when they coassemble with the scaffolding proteins, and maturation of the capsid and packaging of the viral dsDNA are required for P22 to become infective. Interestingly, recent experiments have applied amino acid substitutions in the β-hinge structure of the coat protein, which seems to be a relevant motif controlling the assembly [51]. In particular, two different substitutions at position 170, i.e., $F170L$ and $F170A$, promote the formation of long tubes that show at least one capped end. These structures have been called polyheads and will be the focus of our study in this subsection.

Fortunately, P22 polyheads were long enough to allow high-resolution cryo-EM reconstructions [52] (see Fig. 1.25). The $F170L$ polyheads show two types of tubes smaller than the wild type spherical shell, and with axial symmetries 9-fold (C_9) and 1-fold (C_1, meaning no symmetry), respectively. Moreover, the shape of the hexamers resemble that of the skewed hexons in the P22 procapsid, and the interaction with scaffolding proteins corrects the assembly. On the other hand, $F170A$ polyheads only show one type of tube (see Fig. 1.25c), which is similar in diameter with the P22 shell, and has a 3-fold axial symmetry (C_3). Interestingly, the hexons in this case are isometric like those found in mature P22 virions, and scaffolding proteins do not correct the tubular formation.

Here, our goal is to extract some basic geometric information of the tubes and to infer the possible structure of the polyhead caps. For all cases, we first measure the average hexamer-hexamer distance from the exterior reconstruction (see Fig. 1.25a). To avoid distortions due to the curvature we only look to the hexons closest to the center of the tube. This determines the lattice parameter of the tube a. Then the width of the different tubes is measured, taking into account an intrinsic error due to the protrusions in the surface, and the normalized radius respect the lattice parameter, R/a, is computed. Finally, we measure the chiral angle of the tube, which is related to the angle between a line of hexamers and a perpendicular section of the tube (see Fig. 1.19). All these values are summarized in Table 1.8, and will be enough to predict the structures based on the geometrical model.

The radius R/a is directly related to the triangulation of the cap by Eq. (1.39) (see also Table 1.5). Then for $F170$ (C_9) the possible candidates for the triangulation number are $T_{5F} = 4.5 \pm 0.5$, $T_{3F} = 4.2 \pm 0.5$, and $T_{2F} = 4.1 \pm 0.5$. These results suggest that the structure could be based on a triangulation number $T_{end} = 3$ or 4. In addition, the chiral angle indicates a ring body (see Table 1.4), and the cross section reconstruction clearly shows a ring of nine hexamers (see Fig. 1.25c). Then the only compatible solution corresponds to a $T_{end} = 3$ prolate centered on the 3-fold axis. On the other hand, the radius of the $F170L$ (C_1) is

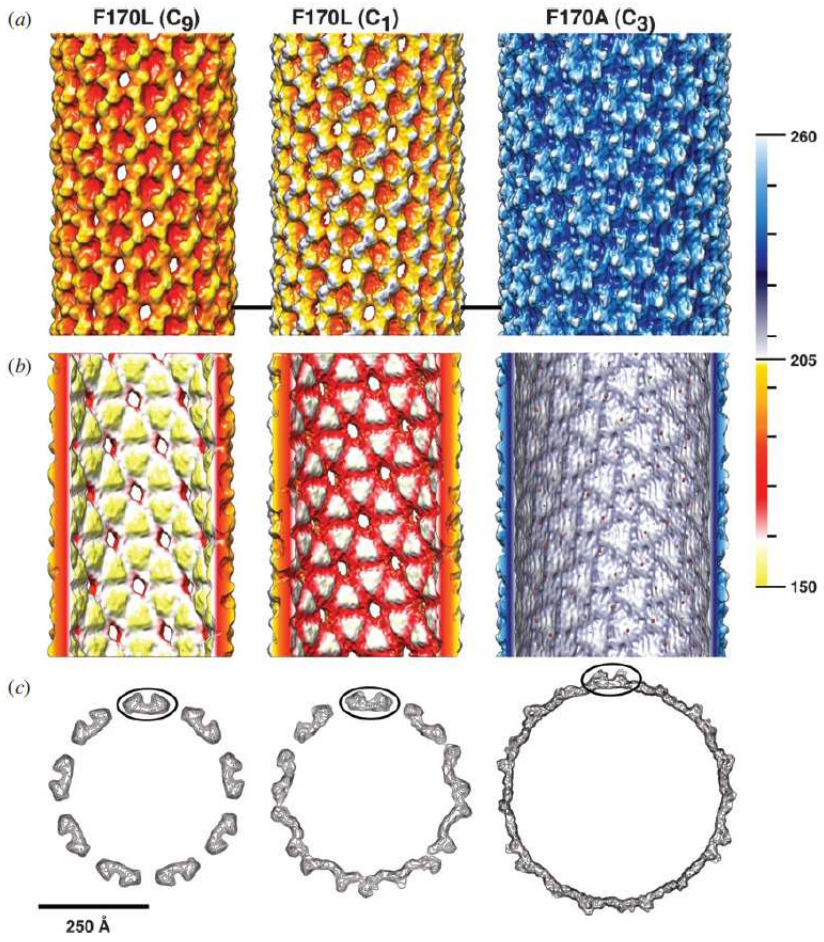


Figure 1.25: Cryo-em reconstructions of bacteriophage P22 polyheads. The three structures were reconstructed from long available tubes of P22. From left to right: $F170L$ (C_9), $F170L$ (C_1), $F170A$ (C_3). The panels correspond to the exterior view of the tube (a), a slice of half tube from the interior (b), and the axial projection of the tubes, where one hexamer is identified in each case to facilitate the interpretation. This last panel clearly unveils the axial symmetries of each tube. Extracted from Ref. [52].

Tube	$a\,[\pm 0.2\,nm]$	R/a	$\theta\,[\pm 1.5^o]$	Prediction
$F170L\ (C_9)$	12.4	1.7 ± 0.2	0	$T_{end} = 3$ (3-fold)
$F170L\ (C_1)$	11.7	1.9 ± 0.3	5.8	–
$F170A$	13.1	2.0 ± 0.2	12.5	$T_{end} = 7$ (3-fold)

Table 1.8: Properties and predictions for P22 polyheads. For each tube the lattice parameter a, the normalized radius R/a, and the chiral angle θ have been measured from the image of Fig. 1.25. The most plausible caps for the polyheads (see text) have been obtained by comparing those values with the predictions for the geometry of prolates.

slightly bigger with intermediate values between $T = 4$ and 7, more specifically $T_{5F} = 5.3 \pm 0.9$, $T_{3F} = 5.7 \pm 0.8$, and $T_{2F} = 5.1 \pm 0.8$. Moreover, the chiral angle does not seem to be compatible with any of the reported cases for these caps in the theory (compare Tables 1.8 and 1.4). Since the characteristics of this tube are similar to the former one, we suggest that this polyhead is a malformation of the $F170$ (C_9), which could be promoted either by an error in the cap formation, or by direct assembly of the tube. Notice that the restricted architectures of the bodies in the geometrical model arise due to the constrain of having an ordered structure of the caps. Interestingly, in the absence of scaffolding, the coat protein of P22 self-assembles *in vitro* into $T = 3$ and $T = 4$ structures, which suggests some connections with the structures adopted by $F170$ polyheads.

Finally, let's analyze the $F170A$ tube. As shown in Table 1.8, the radius is the biggest one; in particular it is associated to either $T_{5F} = 6.3 \pm 0.6$, $T_{3F} = 5.8 \pm 0.6$, and $T_{2F} = 5.6 \pm 0.6$, so the most probable triangulation number of the cap seems to be $T = 7$. In addition, the chiral angle is very close to the prediction corresponding to a 3-fold of this type of cap (see Table 1.4), which is compatible with the C_3 axial symmetry observed experimentally (see Fig. 1.25c). Then the $F170A$ polyhead seems to be a $T_{end} = 7$ (3-fold) bacilliform capsid.

1.4 Discussion

In the previous section we have applied the results of our geometrical model of prolates to characterize specific viruses and make predictions. Depending on the available experimental data, the model can be used in different ways. In any case, only simple algebraic equations, which could be easily used by anyone interested on characterizing a prolate virus, are required. In this context, we emphasize that there is an important number of electron micrographs that identify elongated

viruses, whose structures remain unknown [5, 15, 53]. The combination of these images with basic molecular information of the capsids, e.g., the coat protein size, or the number of coat proteins, would provide a rapid prediction of the structures, which might facilitate the study of these viruses. Unfortunately, in many cases the basic molecular information is not easy to found in the literature or the micrographs available are not of good quality.

The advances in cryo-EM have allowed to reconstruct high-resolution images of asymmetric structures that have been elusive to crystallization and X-ray analysis. In particular, two viruses, and some of their main components, have been studied in detail so far: bacteriophage ϕ29 [18, 19, 54] and bacteriophage T4 [6]. The precise characterization of these structures is fundamental to elucidate the molecular mechanisms involved in the different stages of the virus life cycle. To this end it is very important to achieve high resolution images to identify the precise amino-acids or molecular groups involved in each part of the structure. The quality of the reconstructions can be in general enhanced by imposing the right symmetries of the structure. In this sense, the geometrical theory of prolates described in this chapter provides an excellent tool to suggest the potential local and global symmetries present in a prolate virus.

As we have seen in this chapter, most bacteriophage prolates seem to adopt a 5-fold structure [17]. In terms of the assembly, this type of axial symmetry is probably promoted by the nucleation of the capsid around the connector, which plays the role of an effective pentamer [18]. On the contrary, elongated plant viruses seem more prone to adopt a 3-fold architecture. Interestingly, two polyheads of the mutant P22 phage show 3-fold structures rather than 5-fold ones (see Table 1.8). This could be related to an assembly mechanism based on trimeric interactions proposed recently [52]. The fact that a phage would be able to form a 3-fold prolate, supports one of the key ideas of this thesis that will be formulated in the next chapter, i.e., that the precise structures of viral capsids arises because they are free energy minima.

It is worth to mention that the connector of the tail in bacteriophages usually have a 12-fold symmetry [18, 19], but it is placed in a 5-fold axis of the prolate. This symmetry mismatch has raised some discussion in the literature, see for instance Refs. [19, 55]. One might think that this problem could be avoided by adopting an elongated structure with 3-fold axial symmetry. However, it is worth to stress that the organization of the capsomers (or the coat proteins) at the tips of the caps for a 3-fold prolate changes depending on the T_{end}. Nevertheless, a 5-fold bacilliform architecture is always built with a pentamer on the tip of the cap, regardless the value of T_{end}.

Another surprising result of the geometrical prolate model is that allows the construction of *shrunk* capsids (see Fig. 1.14d), thus obtaining an oblate structure, which in some cases might resemble a lemon-shaped shell. Interestingly there are some bacteriophages [40] and archaea viruses [56] that adopt lemon-shaped capsids, but we have not performed any analysis yet to investigate their structure.

1.5 Conclusions

In this chapter we have presented a geometrical model that establishes the architectural principles controlling the construction of spherical and prolate viruses with icosahedral symmetry. The ideas were based on the classical CK design for quasi-spherical shells [10], and the work of Moody for 5-fold prolates [17, 27]. We have completed the catalog of icosahedral capped prolates by introducing the precise constructions for the 3-fold and 2-fold bacilliform architectures.

Interestingly, there is a finite set of possibilities to construct them, and that leads to discretization rules for the length and number of proteins. These rules are determined by the axial symmetry and the T number of the cap, which also dictate the radius and length of prolates, as well as the arrangement of capsomers in the tubular body.

Remarkably, we have shown that it is possible to construct quasi-spherical capsids that can be conceived as two hemispherical caps rotated around one of their symmetry axes. These isomeric or degenerate shells do not have complete icosahedral symmetry, but are spherical structures that could compete and interfere in the assembly of viral particles. In fact, preliminary results suggest that, in terms of free energy, these structures are in some cases equally stable as the normal icosahedral ones. In general, elongated viruses with $f > 1$ or $\Delta Q^{min} > 1$ can also have more than one structure with the same length but differing on the relative orientation between the caps. This might have some implications in the possible reconstructions of prolate structures by high-resolution techniques.

It is worth to mention that, as we will see in the next chapter using a very simple model of interaction between capsomers, these icosahedral prolate structures are indeed free energy minima, thus justifying their possible occurrence in nature. However, not all of them seem to be equivalent in energy or even energetically optimal, which might be the reason why some structures, specially those based on 2-fold axial symmetry, seem hard to be observed in native viruses.

On the other hand, there are viruses, like polyomavirus, that are able to adopt elongated structures built exclusively by pentamers [57, 58]. Strictly, these

structures do not follow the geometrical model described above, although an alternative tiling theory has been proposed to explain some of the architectures [39, 59]. In the next chapter we will see that all-pentamer prolate capsids closed by icosahedral caps are indeed plausible energy minima, and follow exactly the same selection rules predicted by the geometrical model introduced here.

The results of this work open the door to a simple characterization of elongated viruses using a few parameters, e.g., subunit's size or number, particle dimensions or chirality of the body, which can be obtained from different standard experimental techniques, such as electrophoresis, electron microscopy and electron or X-ray diffraction.

The fact that prolate viruses can adopt different lengths suggests that in principle it should be possible to control them by using the proper assembly conditions and/or using, for instance, different lengths of genetic or nongenetic materials. This possibility would facilitate the design of artificial viral capsids in applications such as nanopatterning or nanotemplating. The structural information provided by the geometrical principles laid out in this work could be potentially very helpful in this task.

Appendices

A Useful properties of a hexagonal lattice

In this appendix, we summarize some properties of the hexagonal lattice that are used in the calculation of the geometry of prolate viruses. The vectors used in this work to characterize the end-caps and the body are described by two different systems of hexagonal coordinates. The first system is characterized by the basis vectors (see Fig. 1.13)

$$\vec{a}_1 = a(1,0) \quad ; \quad \vec{a}_2 = a(0,1) \tag{A-1}$$

where a is the lattice parameter of the hexagonal sheet, i.e., the distance between the centers of two adjacent hexamers. Moreover, since the hexagonal axes are not orthogonal, the elementary products among basis vectors are

$$\vec{a}_1 \cdot \vec{a}_1 = \vec{a}_2 \cdot \vec{a}_2 = a^2 \quad ; \quad \vec{a}_1 \cdot \vec{a}_2 = \frac{a^2}{2} \tag{A-2}$$

$$\vec{a}_1 \times \vec{a}_1 = \vec{a}_2 \times \vec{a}_2 = 0 \quad ; \quad |\vec{a}_1 \times \vec{a}_2| = \frac{\sqrt{3}\,a^2}{2} \tag{A-3}$$

The second set of coordinates, labeled with primes and used to characterize the body vector, are given in terms of a pair of hexagonal axes rotated 60^o counterclockwise respect to the first system, and its basis vectors are

$$\vec{a'}_1 = a(1,0)' \quad ; \quad \vec{a'}_2 = a(0,1)' \tag{A-4}$$

Therefore, $\vec{a'}_1$ coincides with $\vec{a}_2$ of the original basis. The inner and crossed products between $\vec{a'}_1$ and $\vec{a'}_2$ are analogous to those between $\vec{a}_1$ and $\vec{a}_2$, given by Eqs. (A-2) and (A-3). In addition, the inner products between vectors of the two pairs of basis are

$$\vec{a}_1 \cdot \vec{a'}_1 = \frac{a^2}{2} \quad ; \quad \vec{a}_1 \cdot \vec{a'}_2 = -\frac{a^2}{2} \tag{A-5}$$

$$\vec{a}_2 \cdot \vec{a'}_1 = a^2 \quad ; \quad \vec{a}_2 \cdot \vec{a'}_2 = \frac{a^2}{2} \tag{A-6}$$

and their cross products are

$$|\vec{a}_1 \times \vec{a'}_1| = |\vec{a}_1 \times \vec{a'}_2| = \frac{\sqrt{3}\,a^2}{2} \tag{A-7}$$

$$\vec{a}_2 \times \vec{a'}_1 = 0 \quad ; \quad |\vec{a}_2 \times \vec{a'}_2| = \frac{\sqrt{3}\,a^2}{2} \tag{A-8}$$

Using these products it is possible to compute every property reported in this work derived from an operation between vectors.

In general, the area of a triangle defined by two vectors $\vec{a}$ and $\vec{b}$ can be computed as $|\vec{a} \times \vec{b}|/2$. In particular, the area of the elementary triangle, defined by $\vec{a}_1$ and $\vec{a}_2$, and corresponding to $T = 1$, is

$$S_0 = \frac{|\vec{a}_1 \times \vec{a}_2|}{2} = \frac{\sqrt{3}\, a^2}{4} \tag{A-9}$$

B Minimum step ΔQ^{min} and the greatest common divisor $gcd(a, b)$

To calculate the minimum step of ΔQ one needs to introduce, essentially, four properties related to the greatest common divisor of two integer numbers a and b, i.e., $gcd(a, b)$. (i) The Bézout's identity, establishing that if p and q are integers, the smallest positive integer value of $d = ap+bq$ is the $gcd(a, b)$. Moreover, for any integer n: (ii) $gcd(a + nb, b) = gcd(a, b)$, (iii) $gcd(a \times n, b) = gcd(a, b) \times gcd(n, b)$, and (iv) $gcd(a, 0) = |a|$ for $a \neq 0$.

Once the values of (h_0, k_0) and f that characterize the caps are fixed, the length and the area of the body depend only on $Q(h', k')$. Any change in length corresponds to transform the body vector from (h', k') to $(h' + p, k' + q)$. Using Eqs. (1.11), (1.22) and (1.28) that define Q for each symmetry, we have that the change in Q is:

$$\Delta Q = \begin{cases} p(h_0) + q(h_0 + k_0) & \text{5-fold} \\ p(2h_0 + k_0) + q(h_0 + 2k_0) & \text{3-fold} \\ p(3h_0 + k_0) + q(2h_0 + 3k_0) & \text{2-fold} \end{cases} \tag{B-1}$$

Using Bézout's identity, the smallest value of ΔQ, i.e., the smallest step in Q, is then given by the greatest common divisor of the numbers that multiply p and q in each case. We can then use the properties of the gcd mentioned above to derive the growing laws for each symmetry, namely

$$\Delta Q^{min}_{5F} = gcd\,(h_0, h_0 + k_0) = gcd\,(h_0, k_0) = 1 \tag{B-2}$$

$$\Delta Q^{min}_{3F} = gcd\,(2h_0 + k_0, h_0 + 2k_0) = gcd\,(3k_0, h_0 - k_0) = gcd\,(3, h_0 - k_0) \tag{B-3}$$

$$\Delta Q^{min}_{2F} = (3h_0 + k_0, 2h_0 + 3k_0) = gcd\,(7k_0, h_0 - 2k_0) = gcd\,(7, h_0 - 2k_0) \tag{B-4}$$

which correspond to Eqs. (1.17), (1.26) and (1.32).

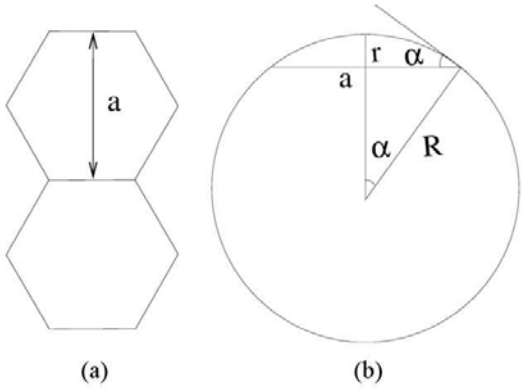


Figure C.1: (a) Contact of two hexamers in a ring layer. (b) Representation of one rigid hexamer in the cross section of a ring body R_n.

C Estimate of the error in the tubular approximation

As an estimation of the error associated to neglecting curvature effects in the tubular approximation, let's calculate the relative discrepancy in the radius for a particularly simple case: a ring body made of n rigid hexamers. Let's first compute the exact radius, R_{ex}, using the fact that each ring layer is composed of n capsomers in contact (Fig. C.1a).

A ring of n hexamers satisfies the equation $2\alpha\, n = 2\pi$, where 2α is the arc defined by a hexamer of the layer (Fig. C.1b). This angle is related to the radius of the tube by $sin\,\alpha = \frac{a}{2R}$. Hence, combining both equations the exact radius of a ring tube is obtained

$$R_{ex} = \frac{a}{2sin\frac{\pi}{n}} \tag{C-1}$$

In our geometrical model, the tubes were constructed by wrapping the hexagonal sheet into a cylinder, with a resulting radius given by Eq. (1.38). For the particular case of a ring body of n capsomers, $|\vec{C_h}| = n\,a$, yielding

$$R = \frac{n}{2\pi}a \tag{C-2}$$

which coincides with the exact radius, Eq. (C-1), in the limit $n \gg 1$. For a given n, the relative error between Eq. (C-2) and the exact radius is

$$\frac{\Delta R}{R_{ex}} = 1 - \frac{n}{\pi}sin\frac{\pi}{n}, \tag{C-3}$$

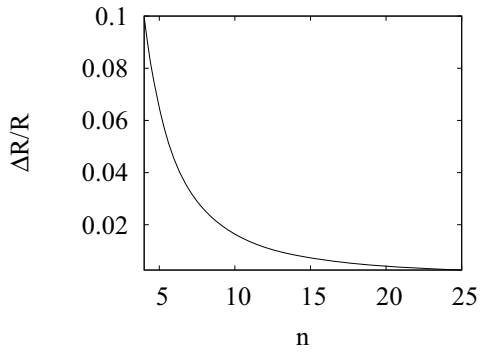


Figure C.2: Relative error of the flat approximation. n is the number of hexamers in the ring layer.

where $\Delta R = R_{ex} - R$. The smallest tube in the prolate model is $T = 1$ 5-fold ($n = 5$), which has a relative error of 6% as shown in Fig C.2. For bigger structures the relative error decays as $1/n^2$. Hence the tubular approximation used for bacilliform viruses is reasonably good even for the smallest particles.

D Value of the $gcd(m,n)$ and $|\vec{C}_h|$ for each axial symmetry

Many of the properties of the tubular part of an elongated virus depend on the value of $gcd(m,n)$. Contrarily to the case of carbon nanotubes, for an icosahedrally capped capsid not all values of (m,n) are valid. The $gcd(m,n)$ can be computed from the values of (m,n) adequate for each symmetry, defined by Eqs.(1.35), (1.36), and (1.37), leading to

$$gcd(m,n)_{5F} = 5f\,gcd(h_0,k_0) = 5f \tag{D-1}$$

$$gcd(m,n)_{3F} = 3f\,gcd(3,h_0-k_0) = \begin{cases} 9f & \text{if } |h_0-k_0| \propto 3 \\ 3f & \text{the rest} \end{cases} \tag{D-2}$$

$$gcd(m,n)_{2F} = 2f\,gcd(7,h_0-2k_0) = \begin{cases} 14f & \text{if } |h_0-2k_0| \propto 7 \\ 2f & \text{the rest} \end{cases} \tag{D-3}$$

We emphasize that the same results obtained for the growing laws, ΔN^{min}, in the prolate construction Eqs. (1.18), (1.27), and (1.33), can be obtained from $gcd(m,n)$.

Another important property is the modulus of the chiral vector $|\vec{C}_h|$. Since the values of (m, n) are restricted by the symmetry, the possible values of $|\vec{C}_h|$ are

$$|\vec{C}_h| = \begin{cases} 5\sqrt{T_{end}}\,a & \text{for a 5-fold} \\ 3\sqrt{3T_{end}}\,a & \text{for a 3-fold} \\ 2\sqrt{7T_{end}}\,a & \text{for a 2-fold} \end{cases} \tag{D-4}$$

E Coordinates of the capsomers in any tubular body

In this block, we use the tubular approximation introduced above to compute the coordinates of the capsomers in the body of any prolate. With this knowledge we have been able to built *ping-pong* models of bacilliform viruses, which have been very useful to illustrate the structures, and to explore further physical properties.

E.1 Unit-cell tubes

In order to determine the position of all capsomers in the tubular body of a prolate virus, it is convenient to introduce the concept of *unit-cell* as the minimum repeating unit of an infinitely long tube. This unit cell is the rectangular portion of the hexagonal lattice defined by the chiral vector, $\vec{C}_h$, and the step vector, $\vec{t}$ (see Fig. E.1),

$$\vec{t} = t_1\vec{a_1} + t_2\vec{a_2} = (t_1, t_2) \tag{E-1}$$

which is the shortest lattice vector in the axial direction, i.e., perpendicular to $\vec{C}_h$.

To calculate the components of the step vector one must take into account three properties: (i) $\vec{t}$ is perpendicular to the chiral vector, i.e., $\vec{C}_h \cdot \vec{t} = 0$; (ii) it belongs to the hexagonal lattice, i.e, t_1 and t_2 must be integers; and (iii) it has the smallest non-zero absolute value; this means that the components t_1 and t_2 do not have a common factor, otherwise one could built a shorter step vector dividing by it, so $gcd(t_1, t_2) = 1$.

From condition (i) an expression relating the components of the step and chiral vectors is obtained

$$m\,t_1 + n\,t_2 = 0 \tag{E-2}$$

One can now give for instance t_2 in terms of the other components, $t_2 = -t_1\,m/n$, and taking into account (ii) and (iii), and using the properties described in the Appendix B, it is possible to obtain an explicit formula for the step vector

$$\vec{t} = \frac{1}{d_R}(-(2n+m), 2m+n) \tag{E-3}$$

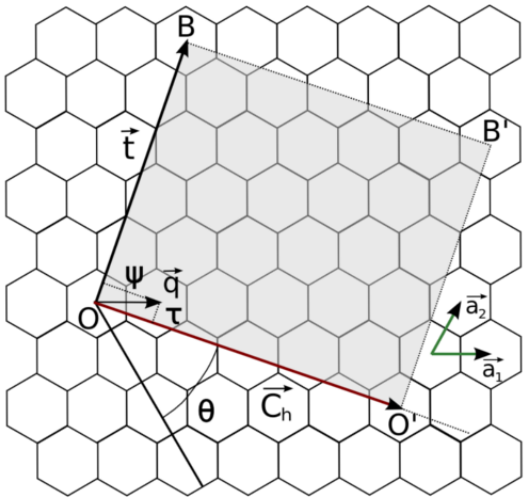


Figure E.1: The unrolled tubular body of a prolate virus (shaded area) shown on the honeycomb lattice. The chiral vector $\vec{C}_h = (m, n)$ defines the rolling direction and the rim of the body where the caps have to be fitted. The fact that icosahedral caps close the tube restricts the possible values of (m, n). The rectangular area delimited by $\vec{C}_h$ and the step vector $\vec{t}$, which is perpendicular to $\vec{C}_h$, defines a unit-cell tube, i.e., the simplest repeating unit of an infinitely long tube. This unit-cell tube has the same relative orientations of pentamers in both caps. However, the shortest tube that can be built is defined by the symmetry vector $\vec{q}$. ψ and τ are the projections of $\vec{q}$ in the directions of $\vec{C}_h$ and $\vec{t}$, respectively, and are related to the pitch and the length of the prolate body.

where d_R is

$$d_R = gcd(2n + m, 2m + n) = gcd(3, m - n)\, gcd(m, n) \tag{E-4}$$

Therefore, for a specific prolate with a chiral vector (m, n), the step vector $\vec{t}$ is completely determined.

The unit cell of the elongated body is defined by $\vec{C}_h$ and $\vec{t}$. Nevertheless one needs still another tool to explicitly compute the position of the hexamers in this body: the symmetry vector $\vec{q}$.

E.2 Symmetry vector

The *symmetry vector*, $\vec{q}$, is defined by the origin of $\vec{C}_h$ and the hexamer in the unit-cell closest to the rim (see Fig. E.1). By construction, $\vec{q}$ is a lattice vector,

and can be expressed in hexagonal coordinates as

$$\vec{q} = q_1\vec{a_1} + q_2\vec{a_2} = (q_1, q_2) \tag{E-5}$$

where, q_1 and q_2 must be integers. Moreover, $\vec{q}$ is the smallest vector in that direction, so it must verify $gcd(q_1, q_2) = 1$.

In the unit-cell there are N_{uc} hexamers, determined by the product $|\vec{C_h} \times \vec{t}|$, i.e., the area of the cell. Then, applying N_{uc} consecutive times the symmetry vector, starting from the origin of $\vec{C_h}$ and taking the end of $\vec{q}$ as the next starting point every time, one gets the positions of all hexamers in the unit-cell. However, the finite body of a prolate is discretized by steps of ΔN^{min}, rather than by complete unit cells, $N_{uc} \geq \Delta N^{min}$. Thus the direct application of $\vec{q}$ is not appropriate to describe the tube of prolates, but this can be circumvent by decomposing $\vec{q}$ in cylindrical components and taking into account the axial symmetry of the structure.

The hexamers in the tubular body of a prolate are in general arranged in an *helical pattern* that can be better described in cylindrical coordinates. Then, it is more convenient to use the parallel ψ and perpendicular τ projections of the symmetry vector with respect to the boundary $\vec{C_h}$ (see Fig. E.1), which are given by

$$\psi = \frac{\vec{q} \cdot \vec{C_h}}{|\vec{C_h}|} \frac{2\pi}{|\vec{C_h}|} \tag{E-6}$$

$$\tau = \frac{\left|\vec{C_h} \times \vec{q}\right|}{|\vec{C_h}|} = \Delta L^{min}. \tag{E-7}$$

The components ϕ and τ correspond to the rotation and longitudinal translation, respectively, that have to be applied to position of a hexamer to obtain the next one in the helical body.

E.3 Hexamers positions

One can then calculate the position of the hexamers in the tubular part of any elongated prolate as a linear combination of the chiral $\vec{C_h}$ and the symmetry vector $\vec{q}$ as

$$i\vec{q} + j\frac{\vec{C_h}}{gcd(m, n)} \tag{E-8}$$

where i, j are integers and $\vec{C_h}/gcd(m, n)$ is the shortest lattice vector in the direction defined by the chiral vector. It is more convenient to reexpress the

position of the hexamers in cylindrical coordinates yielding

$$z = i\tau \quad \text{for } i = 0 \text{ up to } \Delta L/\Delta L^{min} \tag{E-9}$$

$$\phi = i\psi + j\frac{|\vec{C}_h|}{gcd(m,n)}\frac{2\pi}{|\vec{C}_h|} = i\psi + j\frac{2\pi}{gcd(m,n)} \tag{E-10}$$

where ΔL is the length of the tube, ψ and τ are the projections of $\vec{q}$ in the directions parallel and perpendicular to $\vec{C}_h$, and periodic boundary conditions in the azimuthal angle ψ have to be imposed, i.e., for each layer i the index j runs from 1 to $gcd(m,n) = \Delta N^{min}$. In more rigorous mathematical terms, the position of each successive hexamer in the tube can be obtained as a rotation of ψ around the tube axis combined with a translation τ in the axial direction, which reflect the basic space group symmetry operations for a chiral tube [45, 60].

References

[1] S. J. Flint, L. W. Enquist, V. R. Racaniello, and A. M. Skalka, *Principles of virology*. Washington, D.C. ASM Press cop., 2004.

[2] H. Fraenkel-Conrat and R. C. Williams, "Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components," *Proc. Natl. Acad. Sci. USA.*, vol. 41, pp. 690–698, 1955.

[3] A. Klug, "The tobacco mosaic virus particle: structure and assembly," *Philos. Trans. R. Soc. London. [Biol.]*, vol. 354, pp. 531–5, 1999.

[4] W. K. Kegel and P. van der Schoot, "Physical regulation of the self-assembly of tobacco mosaic virus coat protein," *Biophys. J.*, vol. 91, pp. 1501–12, 2006.

[5] ICTVdB, "The Universal Virus Database, version 4." http://www.ncbi.nlm.nih.gov/ICTVdb/ICTVdB/index.htm.

[6] A. Fokine, P. R. Chipman, P. G. Leiman, V. V. Mesyanzhinov, V. B. Rao, and M. G. Rossmann, "Molecular architecture of the prolate head of bacteriophage T4," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, pp. 6003–8, 2004.

[7] B. K. Ganser, S. Li, V. Y. Klishko, J. T. Finch, and W. I. Sundquist, "Assembly and analysis of conical models for the HIV-1 core," *Science*, vol. 283, pp. 80–3, 1999.

[8] F. H. C. Crick and J. D. Watson, "Structure of small viruses," *Nature*, vol. 10, pp. 473–475, 1956.

[9] R. W. Horne and P. Wildy, "Symmetry in virus architecture," *Virology*, vol. 15, pp. 348–73, 1961.

[10] D. L. D. Caspar and A. Klug, "Physical principles in the construction of regular viruses," in *Cold Spring Harbor Symp. Quant. Biol.*, vol. 27, pp. 1–24, Cold Spring Harbor Laboratory Press, 1962.

[11] J. E. Johnson and J. A. Speir, "Quasi-equivalent viruses: a paradigm for protein assemblies," *J. Mol. Biol.*, vol. 269, pp. 665–75, 1997.

[12] T. S. Baker, N. H. Olson, and S. D. Fuller, "Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs," *Microbiol. Mol. Biol. Rev.*, vol. 63, pp. 862–922, 1999.

[13] R. Zandi, D. Reguera, R. F. Bruinsma, W. M. Gelbart, and J. Rudnick, "Origin of icosahedral symmetry in viruses," *Proc. Natl. Acad. Sci. USA.*, vol. 101, pp. 15556–15560, 2004.

[14] D. H. Bamford, J. M. Grimes, and D. I. Stuart, "What does structure tell us about virus evolution?," *Curr. Opin. Struc. Biol.*, vol. 15, pp. 655–63, 2005.

[15] C. A. Tidona and G. Darai, eds., *The Springer index of viruses.* Springer -Verlag Berlin Heidelberg, 2002.

[16] R. Hull, "The structure of tubular viruses," *Adv. Virus. Res.*, vol. 20, pp. 1–32, 1976.

[17] M. F. Moody, "Geometry of phage head construction," *J. Mol. Biol.*, vol. 293, pp. 401–33, 1999.

[18] Y. Tao, N. H. Olson, W. Xu, D. L. Anderson, M. G. Rossmann, and T. S. Baker, "Assembly of a tailed bacterial virus and its genome release studied in three dimensions," *Cell*, vol. 95, pp. 431–7, 1998.

[19] A. Guasch, J. Pous, B. Ibarra, F. X. Gomis-Ruth, J. M. Valpuesta, N. Sousa, J. L. Carrascosa, and M. Coll, "Detailed architecture of a DNA translocating machine: the high-resolution structure of the bacteriophage phi29 connector particle," *J. Mol. Biol.*, vol. 315, pp. 663–76, 2002.

[20] J. Tang, R. S. Sinkovits, and T. S. Baker, "Three-dimensional asymmetric reconstruction of tailed bacteriophage," *Methods Enzymol.*, vol. 482, pp. 185–210, 2010.

[21] J. B. Heymann, C. Butan, D. C. Winkler, R. C. Craven, and A. C. Steven, "Irregular and semi-regular polyhedral models for Rous sarcoma virus cores," *Comput. Math. Methods. Med.*, vol. 9, pp. 197–210, 2008.

[22] C. Butan, D. C. Winkler, J. B. Heymann, R. C. Craven, and A. C. Steven, "RSV capsid polymorphism correlates with polymerization efficiency and envelope glycoprotein content: implications that nucleation controls morphogenesis," *J. Mol. Biol.*, vol. 376, pp. 1168–1181, 2008.

[23] B. K. Ganser-Pornillos, M. Yeager, and W. I. Sundquist, "The structural biology of HIV assembly," *Curr. Opin. Struc. Biol.*, vol. 18, pp. 203–17, 2008.

[24] O. Pornillos, B. K. Ganser-Pornillos, and M. Yeager, "Atomic-level modelling of the HIV capsid," *Nature*, vol. 469, pp. 424–427, 2011.

[25] D. Wales, H. McKay, and E. Altschuler, "Defect motifs for spherical topologies," *Phys. Rev. B*, vol. 79, p. 224115, 2009.

[26] H. S. M. Coxeter, *Introduction to geometry.* Wiley, 2nd ed., 1989.

[27] M. F. Moody, "The shape of the T-even bacteriophage head," *Virology*, vol. 26, pp. 567–76, 1965.

[28] J. D. Bernal, I. Fankuchen, and D. P. Riley, "Structure of the crystals of tomato bushy stunt virus preparations," *Nature*, vol. 142, p. 1075, 1938.

[29] C. H. Carlisle and K. Dornberger, "Some x-ray measurements on single crystals of tomato bushy-stunt virus," *Acta Crystallogr.*, vol. 1, pp. 194–196, 1948.

[30] J. D. Bernal and C. H. Carlisle, "Unit cell measurements of wet and dry crystalline turnip yellow mosaic virus," *Nature*, vol. 162, pp. 139–140, 1948.

[31] D. L. D. Caspar, "Structure of bushy stunt virus," *Nature*, vol. 177, pp. 475–476, 1956.

[32] R. W. Lucas, S. B. Larson, and A. McPherson, "The crystallographic structure of brome mosaic virus," *J. Mol. Biol.*, vol. 317, p. 95, 2002.

[33] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. G. E. C. M. EC, and T. E. Ferrin, "UCSF Chimera–a visualization system for exploratory research and analysis," *J. Comput. Chem.*, vol. 25, pp. 1605–1612, 2004.

[34] R. Kerner, "Classification and evolutionary trends of icosahedral viral capsids," *Comput. Math. Methods. Med.*, vol. 9, pp. 175–181, 2008.

[35] R. Mannige and C. L. Brooks, "Periodic table of virus capsids: implications for natural selection and design," *PloS one*, vol. 5, p. e9423, 2010.

[36] J. M. Grimes, J. N. Burroughs, P. Gouet, J. M. Diprose, R. Malby, S. Ziéntara, P. P. Mertens, and D. I. Stuart, "The atomic structure of the bluetongue virus core," *Nature*, vol. 395, pp. 470–8, 1998.

[37] D. M. Salunke, D. L. D. Caspar, and R. L. Garcea, "Polymorphism in the assembly of polyomavirus capsid protein VP1," *Biophys. J.*, vol. 56, pp. 887–900, 1989.

[38] R. Twarock, "A tiling approach to virus capsid assembly explaining a structural puzzle in virology," *J. Theor. Bio.*, vol. 226, pp. 477–82, 2004.

[39] R. Twarock, "Mathematical models for tubular structures in the family of papovaviridae," *Bull. Math. Biol.*, vol. 67, pp. 973–987, 2005.

[40] H. Ackermann, "Bacteriophage observations and evolution," *Research in Microbiology*, vol. 154, pp. 245–251, 2003.

[41] R. Hull, G. J. Hills, and R. Markham, "Studies on alfalfa mosaic virus. II. The structure of the virus components," *Virology*, vol. 37, pp. 416–28, 1969.

[42] R. A. Heijtink, C. J. Houwing, and E. M. Jaspars, "Molecular weights of particles and RNAs of alfalfa mosaic virus. Number of subunits in protein capsids," *Biochemistry*, vol. 16, pp. 4684–93, 1977.

[43] R. Hull, "Molecular biology of rice tungro viruses," *Annu. Rev. Phytopathol.*, vol. 34, pp. 275–97, 1996.

[44] S. Cusack, G. T. Oostergetel, R. C. J. Krijgsman, and J. E. Mellema, "Structure of the top$_{a-t}$ component of alfalfa mosaic virus (a non-icosahedral virion)," *J. Mol. Biol.*, vol. 171, pp. 139–155, 1983.

[45] R. Saito, M. S. Dresselhaus, and G. Dresselhaus, *Physical properties of carbon nanotubes*. Imperial College Press, 1998.

[46] E. L. Altschuler and A. Pérez-Garrido, "Do all spherical viruses have icosahedral symmetry?," *arXiv*, vol. 0902.3566v, pp. 1–4, 2009.

[47] L. D. Landau and I. M. Lifshitz., *Theory of elasticity*. Pergamon, London, 1986.

[48] A. Kumar, V. S. Reddy, V. Yusibov, P. R. Chipman, Y. Hata, I. Fita, K. Fukuyama, M. G. Rossmann, L. S. Loesch-Fries, T. S. Baker, and J. E. Johnson, "The structure of alfalfa mosaic virus capsid protein assembled as a $T = 1$ icosahedral particle at 4.0-Å resolution," *J. Virol.*, vol. 71, pp. 7911–7916, 1997.

[49] P. E. Prevelige, D. Thomas, and J. King, "Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells," *Biophys. J.*, vol. 64, pp. 824–35, 1993.

[50] K. N. Parent, R. Khayat, L. H. Tu, M. M. Suhanovsky, J. R. Cortines, C. M. Teschke, J. E. Johnson, and T. S. Baker, "P22 coat protein structures reveal a novel mechanism for capsid maturation: stability without auxiliary proteins or chemical crosslinks," *Structure*, vol. 18, pp. 390–401, 2010.

[51] M. M. Suhanovsky, K. N. Parent, S. E. Dunn, T. S. Baker, and C. M. Teschke, "Determinants of bacteriophage P22 polyhead formation: the role of coat protein flexibility in conformational switching," *Mol. Microbiol.*, vol. 77, pp. 1568–82, 2010.

[52] K. N. Parent, R. S. Sinkovits, M. M. Suhanovsky, C. M. Teschke, E. H. Egelman, and T. S. Baker, "Cryo-reconstructions of P22 polyheads suggest that phage assembly is nucleated by trimeric interactions among coat proteins," *Phys. Biol.*, vol. 7, p. 045004, 2010.

[53] M. J. Adams and J. F. Antoniw, "DPVweb: a comprehensive database of plant and fungal virus genes and genomes," *Nucleic. Acids. Res.*, vol. 34, pp. D382–D385, 2006.

[54] Y. Xiang and M. G. Rossmann, "Structure of bacteriophage ϕ29 head fibers has a supercoiled triple repeating helix-turn-helix motif," *Proc. Natl. Acad. Sci. USA.*, vol. 2011, pp. 1–5, 2011.

[55] M. C. Morais, Y. Tao, N. H. Olson, S. Grimes, P. J. Jardine, D. L. Anderson, T. S. Baker, and M. G. Rossmann, "Cryoelectron-microscopy image reconstruction of symmetry mismatches in bacteriophage ϕ29," *J. Struct. Biol.*, vol. 136, pp. 190–200, 2001.

[56] D. Prangishvili, P. Forterre, and R. A. Garrett, "Viruses of the Archaea: a unifying view," *Nat. Rev. Microbiol.*, vol. 4, pp. 837–48, 2006.

[57] N. A. Kiselev and A. Klug, "The structure of viruses of papilloma-polyoma type (V. tubular variants built of pentamers)," *J. Mol. Biol.*, vol. 40, pp. 155–171, 1969.

[58] T. Baker, Caspar, D. L. D., and W. T. Murakami, "Polyoma virus 'hexamer' tubes consist of paired pentamers," *Nature*, vol. 303, pp. 446–448, 1983.

[59] T. Keef, A. Taormina, and R. Twarock, "Classification of capped tubular viral particles in the family of Papovaviridae," *J. Phys.: Cond. Matt.*, vol. 18, pp. S375–S387, 2006.

[60] M. Dresselhaus, "Physics of carbon nanotubes," *Carbon*, vol. 33, pp. 883–891, 1995.

Chapter 2

Optimal Structures for Spherical and Elongated Viral Capsids

2.1 Introduction

As we have seen in the introduction of the thesis and in the previous chapter, viral capsids generally have well-defined and precise structures. Indeed, they adopt a common set of architectures (namely, spherical, rod-like, bacilliform, or conical) that can be described and classified by the same construction principles. This is truly remarkable, specially taking into account that these viruses infect different hosts and are made out of coat proteins, whose amino-acid sequences, sizes or conformations can be very diverse [1–6].

More importantly, in many cases the wild type structure of a viral capsid can be reconstituted *in vitro* in a self-assembly process from a solution of the coat proteins alone [7–18], in combination with the scaffolding proteins (when those are required) [19–22], or in the presence of the viral genome [23, 24], producing infectious particles in the latter case. This spontaneity observed in the formation of viral capsids suggests that the process of assembly should be driven by a fundamental thermodynamic principle: the free energy minimization [25–29].

Furthermore, several *in vitro* experiments have shown that viral coat proteins are able to assemble into well-defined structures different from the native capsid at different physicochemical conditions [7, 8, 11, 13, 30–32]. In particular, in the mid 1970s Adolph and Butler already showed the self-assembly polymorphism of CCMV coat proteins at different pH and salt concentrations (see Fig. 2.1), a

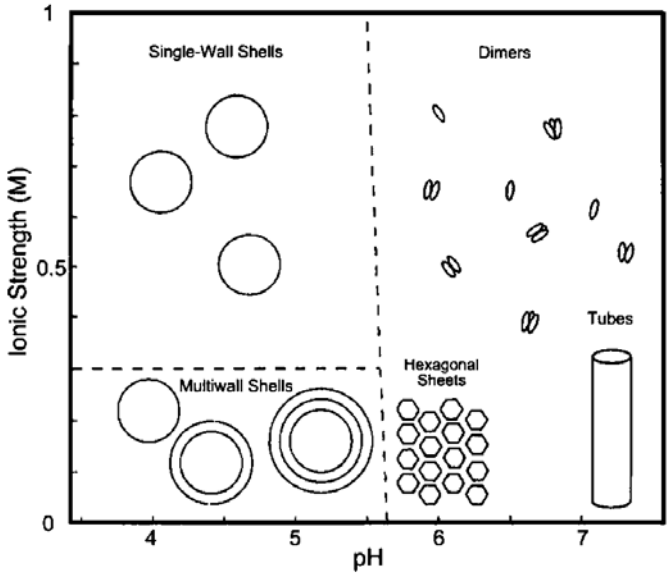


Figure 2.1: Self-assembly capsid protein diagram of CCMV adapted from Adolph and Butler [8]. The axes show the range of pH and ionic strength conditions that were used for the different assembly experiments. The pictures in the diagram illustrate the dominant species at each condition.

seminal experiment that was recently reproduced and extended in Ref. [30]. The existence of common structures among wild type viral capsids, and the intrinsic ability of their coat proteins to self-assemble into different shapes suggest that the formation of these structures should be controlled by some generic mechanisms of interaction.

Following these ideas there has been an increasing effort to understand the structure of viruses from a physical standpoint. The case of quasi-spherical viruses have been the most investigated as we will review in the first section of the chapter. In particular, it is worth to mention the work of Zandi *et al.* [33], which has recently unveiled the underlying reasons behind the emergence of icosahedral symmetry in viruses using a simple physical model. On the other hand, prolate viruses have been less explored, and the physical reasons justifying their possible structures remain unknown. The main goal of this chapter is precisely to adapt the investigation performed in Ref. [33] to the case of bacilliform viral capsids, and to study whether the icosahedral prolate structures predicted by the geometrical model in Chapter 1 are all feasible and can be justified from a

physical point of view.

The present chapter is organized as follows: in Section 2.2 we rederive the results of Ref. [33] for spherical capsids, using a slightly modified physical model that will be the basis of our extension to study the elongated capsids in Section 2.3. Finally, Section 2.4 summarizes the more relevant findings of our study, making a special emphasis on the connection with the geometrical model of Chapter 1, and on the prediction of the structure of several bacilliform viruses.

2.2 Spherical capsids

As mentioned in the previous chapter, most spherical viruses self-assemble building a capsid with icosahedral symmetry. The fact that many viruses can be reconstituted *in vitro* from their molecular components in a spontaneously self-assembly process points out that the emergence of icosahedral symmetry, and the rules derived from the CK construction, might be generic features of the free energy minima of aggregates of viral capsid proteins.

The best way to justify this hypothesis would be to systematically explore the equilibrium structures of viral capsids by simulating the coat proteins and their interactions for typical aggregation conditions. However the evaluation of the capsid free energy by realistic all-atom or quasi-all-atom simulations is computationally very expensive, and only the smallest viruses at very short time scales (at the of pico to nanoseconds) can be studied [34].

An alternative would be to consider "coarse-grained" models that would allow to efficiently tackle the problem. Indeed, the initial effort to design simple models to study the emergence of the well-defined structure of capsids was based on attempts to connect viral icosahedral symmetry with the mathematical problem of obtaining the closest packing of N equal disks (representing the capsomers) on the surface of a sphere [35–37]. Interestingly, this approach was intimately related to one problem posed much earlier by Thomson [38] that focused on the minimum energy configuration of N electrons on a spherical surface. That subsequently raised the question of determining the arrangement of N overlapping disks, with a minimum size, to completely cover the surface of a sphere [39]. Following these ideas, a more recent model derived a self-assembly phase diagram by analyzing the energy minima structures made of identical capsomers with a given cohesion energy and spontaneous curvature [25]. Unfortunately, all these models only deal with identical morphological units and led to capsid structures with symmetries lower than icosahedral.

On the other hand, when icosahedral symmetry is imposed in the modeling of

viral structures, it is possible to recover the quasi-equivalent organization present in the CK construction [40]. In fact, recent investigations have even allowed the possibility to infer the configuration of the internal genetic material or the organization of multishell icosahedral capsids for certain viruses [41–43], by affine extensions of icosahedral symmetry. Interestingly, it has been possible to recover also the coat protein organization of some icosahedral all-pentamer viruses, which deviate from the standard CK classification, using principles of tiling theory [44] or pentagon packing models [45]. Other approaches based on the assumption of icosahedral symmetry have focused instead on the pathways and kinetics of the capsid formation process [46,47], studying, for instance, mechanisms that enhance or preclude the proper formation of viral capsids [48] (see also Chapters 6 and 7). Finally, elasticity arguments demonstrated that, for large shells, the pentameric disclinations in a capsid destabilize the spherical shape, and leads to a faceting that produces a more polyhedral-like structure (through a phenomenon named "buckling transition"), in consonance with the observations of big quasi-spherical viral capsids (see also Chapter 4) [49, 50].

However, the origin of icosahedral symmetry in viruses, the validity conditions for the CK construction, and the physical principles underlying the quasi-equivalence principle were not firmly established until the work of Zandi *et al.* in Ref. [33], which was able to successfully explain the emergence of CK icosahedral shells and other type of structures observed in *in vitro* experiments using a very simple minimal model. That model, properly adapted and extended, is the one that we have used to analyzed prolate capsids, and it is introduced in the following section.

2.2.1 Minimal model of spherical capsids

As mentioned in the introduction, capsid proteins of distinct viruses can differ in terms of the sequence, size and conformations of their amino acids. However, in their final structures they are invariably clustered into morphological units, e.g., pentamers and/or hexamers, which are very similar in all viruses [51]. Thus, while the interaction between proteins can be very complex and species-specific, capsomers are expected to interact through a more generic and isotropic potential.

The model that we will use then focuses on the interactions at the level of capsomers, rather than individual proteins, and it is an adaptation on the previous work of Zandi *et al.* [33]. The minimal model presented in Ref. [33] was designed to capture the essential ingredients of the interaction between capsomers: a short-range repulsion, associated to the subunit conformational rigidity, that prevents protein overlapping; and a longer-range attraction representing the driving force

for capsomer aggregation. There are many possible potentials that qualitatively fulfill these requirements. For sake of simplicity, a Lennard-Jones-like potential was chosen, a well-known interaction model in the condensed matter field. Thus, the effective capsomer-capsomer interaction potential is described by

$$V(r) = \varepsilon_0 \left[\left(\frac{\sigma}{r}\right)^{12} - 2\left(\frac{\sigma}{r}\right)^{6} \right] \tag{2.1}$$

where r is the separation between capsomer centers, σ is their optimal distance, and ε_0 is the binding energy between them (see Fig. 2.2a). The capsomer-capsomer binding energy ε_0 was taken to be $15\,kT$ (where k is the Boltzmann constant and T is the absolute temperature), a typical value reported from atomistic calculations of subunit binding energies [52].

As we have discussed in great detail in the previous chapter, an essential feature of the organization of coat subunits in the final viral capsids is the presence of two morphological structures: pentamers and hexamers, which are made of five and six proteins, respectively. In general, these two capsomers may have different *internal energies* arising from differences in folding, stretching, and contact interactions between the coat subunits. We will represent this free energy difference by ΔE. Accordingly, the Boltzmann factor $e^{-\Delta E/kT}$ would account for the relative probability of a free capsomer to be a pentamer rather than an hexamer. It is very important to emphasize that we are not claiming that, during the assembly of a capsid, hexamers and pentamers are necessarily present in solution and able to physically switch between each other. The mechanisms and the pathways to create these two structural units in the final shell can differ very much depending on the virus. For instance, the coat subunits of bacteriophage HK97 can actually aggregate into hexamers and pentamers before starting the assembly, and ending up forming a $T = 7$ shell [12]. But in the vast majority of cases, such as in CCMV, the building blocks (in that case stable dimers) present in solution are not necessarily the morphological units found in the assembled capsid (a $T = 3$ shell in the case of CCMV [8, 53]). Our analysis here focuses on the final equilibrium structure adopted by viral capsids, not on the assembly pathways, and the Boltzmann factor introduced above is a convenient tool to properly explore the free energy landscape of the proposed thermodynamic model for viral shells.

In the seminal paper of Zandi *et al.* [33], an isotropic interaction model was used, where capsomers were considered as disks of two different sizes. For these disks the equilibrium distance σ was related to the distance of the disks centers at contact, which depended on the curvature of the capsid. Unfortunately, this became an inconvenient when we tried to extend the equilibrium model to study elongated shells, due to the ambiguity in the definition of the equilibrium distance

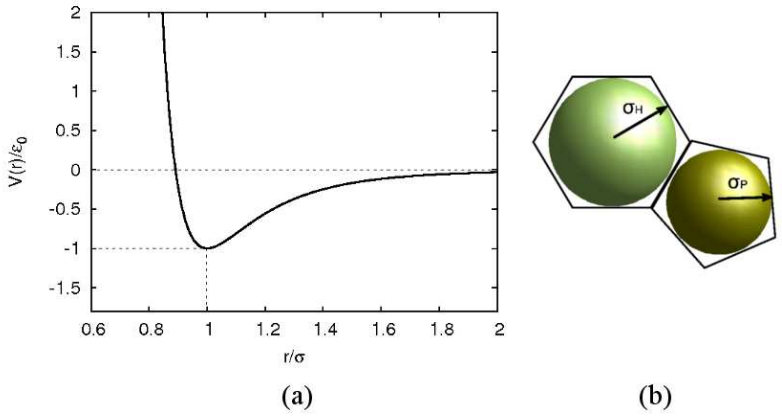


Figure 2.2: Coarse-grained capsomer model. (a) Effective capsomer-capsomer interaction. The Lennard-Jones potential, Eq. (2.1), is plotted as a function of the distance between the center of two capsomers r, where σ is the equilibrium distance, and ε_0 is the binding energy. (b) Hexameric and pentameric capsomers are modeled as spheres. Their radii σ_h and σ_p are related by that of the circumscribed hexagon and pentagon, having the same edge.

at the boundary where the spherical cap and the cylindrical tube meet (see Section 2.3). Thus, we decided to remodel the capsomers as spheres, which is a less realistic shape compared to the actual capsomers, but keeps the isotropic nature of the potential and overcame the problems for bacilliform capsids. In this situation the equilibrium distance is determined exclusively by the distance between the centers of two capsomers in contact.

The size of the model pentamers and hexamers was chosen in such a way that they can produce a good covering of a spherical capsid. In particular, the ratio between their effective radii was determined by inscribing their equatorial circles, respectively, into a pentagon or a hexagon of the same edge, to respect the fact that pentamers and hexamers are typically made of the same protein[1] (see Fig. 2.2b). Therefore, the ratio between the radius of a hexamer, σ_h, and that of a pentamer, σ_p, is given by

$$r_{ph} = \frac{\sigma_p}{\sigma_h} = \frac{\tan(\pi/6)}{\tan(\pi/5)} \approx 0.8 \tag{2.2}$$

[1]In reality, some viruses are made by different coat proteins, but they usually have a similar size.

Since we use a pair potential, the equilibrium distance will depend on which type of capsomers are interacting. In particular, we have three different possibilities, i.e., hexamer-hexamer, pentamer-pentamer, and hexamer-pentamer, and since the equilibrium distance corresponds to the contact distance for each case, we have

$$\sigma_{hh} = 2\,\sigma_h = 2\,\sigma_0 \tag{2.3a}$$
$$\sigma_{pp} = 2\,\sigma_p = 2\,r_{ph}\,\sigma_0 \tag{2.3b}$$
$$\sigma_{ph} = \sigma_p + \sigma_h = (1 + r_{ph})\,\sigma_0 \tag{2.3c}$$

respectively. Here σ_0 is the hexamer radius that defines our length scale, which is straightforwardly related to the lattice parameter a of the geometrical model by (see Chapter 1)

$$a = 2\,\sigma_0 \tag{2.4}$$

This connection will be specially useful to compare the predictions of both the geometrical and equilibrium simulation models.

Finally, the binding energy was assumed to be the same ($\varepsilon_0 = 15\,kT$) in all cases. All simulation results will be reported in reduced units, i.e., in units of ε_0 for the energy and σ_0 for lengths.

2.2.2 Monte Carlo simulations

After formulating the model and the interaction between units, the next step was to minimize the free energy of the system to obtain the equilibrium structures of these model shells. To that end, we resorted to a simple, standard and powerful tool in statistical mechanics: Monte Carlo simulation.

Since we were interested in the structures of spherical capsids, a key simplification in our model was to add the constrain that capsomers are forced to be on the surface of a sphere of radius R. This is equivalent to have an extremely rigid bending energy between capsomers and to assume that coat proteins have associated a certain preferred or spontaneous curvature in their interaction (or alternatively that it is introduced by a scaffold), which imposes the radius of the spherical shell.

Given the strength of the binding energies and the fact that we are only interested on the study of equilibrium structures of viral capsids, which usually are highly symmetrical, the kinetic energy of capsomers is neglected, and we focus on the internal energy of capsids as a result of the capsomer-capsomer

interactions. Thus, the energy of the system is characterized by the Hamiltonian

$$H(N,R) \approx \sum_{i=1}^{N} \sum_{j<i} V(r_{ij}) + E_i^{int} \tag{2.5}$$

where N is the total number of capsomers, r_{ij} is the distance between capsomer i and j, and E_i^{int} is the internal capsomer energy given by

$$E_{int} = \begin{cases} 0 & \text{hexamer } (h) \\ \Delta E & \text{pentamer } (p) \end{cases} \tag{2.6}$$

To minimize this energy and to obtain the equilibrium capsid structures, a simple Monte Carlo (MC) algorithm was implemented. This consisted essentially on two different types of "motions": a small displacement of a morphological unit on the surface of the template R; and an attempt to switch the internal state of the capsomer, i.e., between hexamer and pentamer. After every movement the new energy was computed, $H'(N,R)$, and compared with the old one, $H(N,R)$,

$$\Delta H = H' - H \tag{2.7}$$

The standard metropolis MC algorithm accepts this new configuration with a probability

$$P_{acc} = \begin{cases} 1 & \text{if } \Delta H < 0 \\ e^{-\Delta H/kT} & \text{if } \Delta H > 0 \end{cases} \tag{2.8}$$

Thus, if the new energy is lower, the movement is always accepted, and if it is higher the acceptance follows a Boltzmann probability. Applying this procedure repeatedly ensures the minimization of the energy [54].

Fig. 2.3 illustrates the main steps of our MC exploration. The initial condition of each simulation is that N capsomers, all of them initially hexamers, are placed randomly on the surface of a sphere with radius R. Then, the MC algorithm is applied to equilibrate the structure, and the same procedure is repeated for different radii, which leads to an energy minima at $R_e(N)$ associated to a specific optimal structure. It is worth to notice that for large N, the free energy landscape gets more complex, with a number of metastable configurations increasing exponentially [55]. Thus, to avoid possible trapping on local minima, each simulation $N(R)$ was repeated using different initial conditions, and the average and optimal (the absolute minimum of the energy in all different runs)

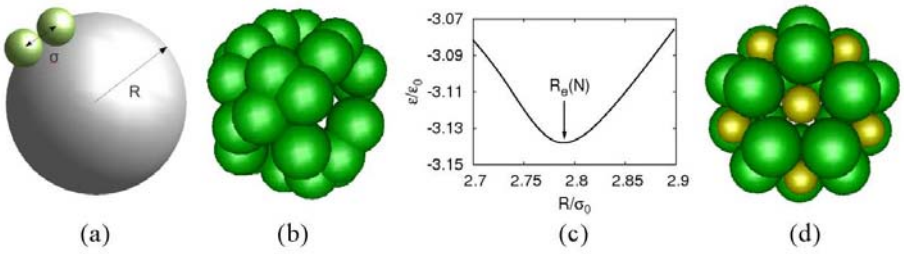


Figure 2.3: Monte Carlo simulations. (a) Spherical surface of radius R where the movement of the capsomers is constrained. (b) Initial random configuration of N hexamers on the template R. (c) The number of capsomers N is fixed during the simulation, and for each radius R we minimize the energy of the system using the Monte Carlo algorithm described in the text. The figure represents the energy per capsomer for a $N = 32$ shell as a function of the radius of the capsid, with the energy minimum located at R_e. In this particular case we allowed the coexistence of both hexamers and pentamers, which corresponds to the $\Delta E = 0$ limit in the text. The optimal structure obtained at R_e is an icosahedral $T = 3$ capsid (d).

configurations were analyzed[2]. Note that during the simulations the number of pentamers in each structure is never imposed.

We performed two different types of simulations: one set with $\Delta E = 0$, corresponding to the case where hexamers and pentamers have the same energy; and a second set of simulations with $\Delta E \gg 0$, to mimic the case where only one type of capsomer is present. In the one-capsomer exploration, each single simulation was composed of 25000 MC cycles, which were usually divided in 15000 equilibration cycles and 10000 production cycles to get statistical properties. We repeated the simulation from 10 to 20 times. For the hexamer-pentamer scenario, we performed also simulations of 25000 MC cycles, but they were decomposed in three parts 5000 equilibration cycles allowing only displacement movements of capsomers, 10000 equilibration steps including also the pentamer-hexamer switching, and 10000 production cycles, allowing both types of movements. The ratio of displacement versus switching moves was chosen as $2N : 1$, which was found to be efficient to find the energy minima. Finally, for each N, we repeated the simulations at different radii R at intervals of $\Delta R_{step} = 0.01\sigma_0$, as illustrated in Fig. 2.3c. The energy versus R typically shows a nearly parabolic behavior. The

[2]We analyzed structures up to $N \sim 100$. However for larger values of N, non standard techniques as parallel tempering or simulated annealing would be necessary to efficiently avoid trapping in metastable energy minima [54].

value of the energy and radius corresponding to the global minimum, define the optimal radius $R_e(N)$ and the optimal energy E_e for a given N.

2.2.3 Two different morphological units

Most spherical viruses have pentamers and hexamers as morphological units in the final capsid, although those units are not necessarily stable in solution. Therefore, let's first analyze the model in the limit of $\Delta E = 0$, where the pentamers and the hexamers have the same energy.

The total energy of a capsid scales with the number of capsomers, $E_t(N) \sim N$. Thus, for bigger capsids the energy decreases linearly and that complicates the comparison between structures with different N. Therefore, we will analyze the structures in terms of their energy per capsomer $\varepsilon(N) = E_t/N$, which is an intensive property that can be interpreted as a sort of equilibrium chemical potential, $\mu_e(N) \approx \varepsilon(N)$, in our model. Following the analogy, we can interpret the different spherical structures obtained for different N as different phases of the system, and the local minima in the $\varepsilon(N)$ landscape will be associated to the more stable equilibrium phases for the viral capsids.

We have minimized the free energy applying the Monte Carlo method described above for N in the range $N \sim 10 - 100$. Fig. 2.4 plots the energy per capsomer $\varepsilon(N)$ in units of ε_0 versus the number N of capsomers. One can observe that on average, $\varepsilon(N)$ decreases as N increases, which is related to the accumulation of hexons, i.e., capsomers with six neighbors. For large N, the energy per capsomer tends to a value slightly larger than that of a flat hexagonal network[3] $\varepsilon_{hex} \approx -3.38\,\varepsilon_0$, due to the energetic penalty of the pentameric defects required to close the shell.

Except for some small capsids, we find that all optimal structures for any N are exclusively composed by 12 pentamers and $N - 12$ hexamers. This seems a natural result, because the six neighbor coordination of hexons maximizes the local energy, and 12 pentameric disclinations are the minimum required to close the structure, in agreement with Euler's theorem. In addition, the most pronounced minima in $\varepsilon(N)$ are seen at $N = 12$, 32, 42, and 72 (see Fig. 2.4), and the capsid structures associated with them are shown in Fig. 2.5. All four minima have icosahedral symmetry and match with the first elements in the CK series, $T = 1$, $T = 3$, $T = 4$, and $T = 7$, respectively. Therefore, we find that the icosahedral

[3]For a binding interaction ε_0 the energy of a capsomer that only interacts with its six nearest-neighbors at the equilibrium distance will be $-3\varepsilon_0$. The influence of the non-nearest-neighbors reduces the total energy, even though the energy of the nearest-neighbors slightly increases due to compression.

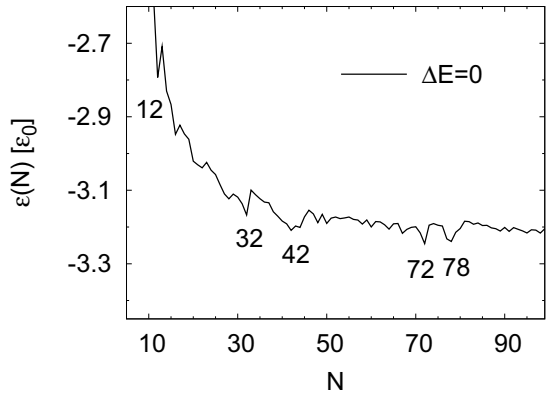


Figure 2.4: Equilibrium landscape for spherical capsids (two morphological units). Energy per capsomer for the optimal configuration of each N capsid. The case corresponding to hexamers and pentamers having the same internal energy, $\Delta E = 0$, is plotted.

symmetry and the T-number organization, commonly observed in viruses, arises naturally from the free energy minimization of a very generic interaction. This is the same result obtained in Ref. [33], despite the differences in the model, which strongly supports the idea that these structures appear in a generic way irrespective of the interaction details.

In our exploration up to $N \sim 100$ we found two surprises: first, there is a pronounced local minima at $N = 78$ that shows a tetrahedral symmetrical structure, rather than icosahedral, although we are not aware of any virus or aberrant particle adopting this architecture. Second, $N = 92$ is not a local minimum, and in fact the energy of a disordered structure was slightly lower than the CK-construction $T = 9$ itself. The absence of $T = 9$ is related to the accumulated stress of this particular architecture due to the spherical template constraint, as it will be discussed in detail in Chapter 4. Remarkably, for N large, the differences in the energy per capsomer start to be relatively small and the landscape of metastable structures becomes more complicated. In fact, many large viruses assemble with the help of scaffolding proteins that might impose the right curvature and select the proper structure. Thus, we did not perform a systematic study for $N > 100$, but we were able to obtain two more CK structures

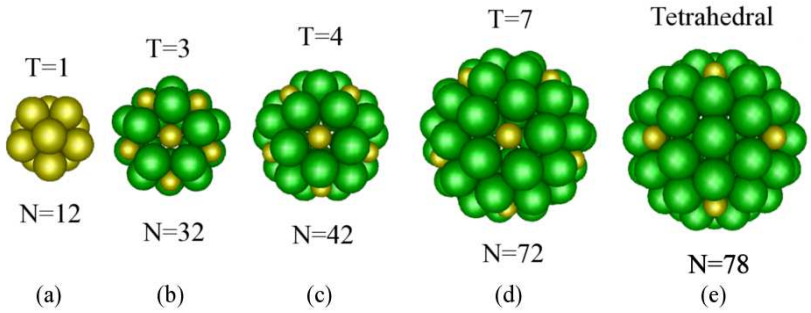


Figure 2.5: Gallery of optimal structures for capsids with two morphological units ($\Delta E = 0$).

as energy minima for $N = 122$ ($T = 12$) and $N = 132$ ($T = 13$)[4].

As we have discussed in the previous chapter, the number of quasi-equivalent positions for the coat units is proportional to the T-number of shell. This suggests that bigger T architectures should be more challenging to build. However, the four equilibrium configurations obtained in the studied range of N, developed spontaneously the icosahedral symmetry, despite the complexity of the structures. For instance, we notice that the $T = 7$ is built by a chiral repeat motif. In particular, in Fig. 2.5d we show the right-handed $T = 7d$. Surprisingly, for the $N = 42$ local minima we find two well defined structures that have the same energy (see Fig. 2.6). One corresponds to the strict icosahedral construction $T = 4$, and the other is the 5-fold isomer predicted by our CK generalized geometrical model (see Chapter 1). We point out that the 3-fold and 2-fold isomers are geometrically feasible for $T = 4$, but in those cases, pentamers become too close in the central body and the energy increases too much, so they are not energetically favorable.

2.2.4 One type of morphological unit

There are some viruses assembled using just one type of morphological unit, typically pentamers [11, 17, 56]. To study this situation in the framework of our model we now focus on the limit where the difference of internal energies between hexamers and pentamers is relatively large, $|\Delta E/\varepsilon_0| \gg 1$. In equilibrium this will

[4]We performed larger MC simulations with 40×25000 MC cycles with ΔR_{step} ranging from 0.0005 up to 0.001. The $T = 12$ and $T = 13$ shells were the local minima for $N = 122$ and $N = 132$, respectively. The optimal values were $R_{opt} = 5.7255$ with $\varepsilon(122) = -3.2337\varepsilon_0$ for $T = 12$, and $R_{opt} = 5.9229$ with $\varepsilon(132) = -3.2345\varepsilon_0$ for $T = 13$.

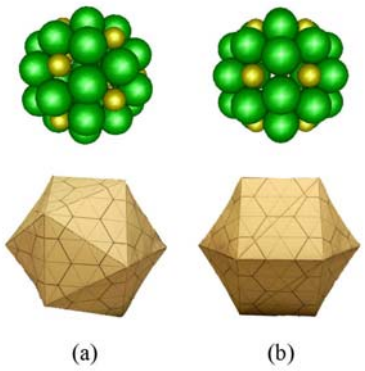


Figure 2.6: The two equilibrium structures with the same energy per capsomer, obtained for $N = 42$. (a) The icosahedral CK construction $T = 4$, and (b) its 5-fold isomer. The folded figures at the bottom have been built following the generalized geometrical model discussed in Chapter 1.

lead to capsids where all units are exclusively hexamers ($\Delta E > 0$), or pentamers ($\Delta E < 0$), depending on the sign of ΔE.

Again, MC simulations were performed to study the range of capsids from $N = 10$ up to 99. The energy landscape $\varepsilon(N)$ for the optimal structures of this capsids made of one type of capsomer corresponds to the gray curve in Fig. 2.7. Comparing the present results with the pentamer-hexamer scenario, one observes a dramatic reorganization of the equilibrium spectrum. In particular, the optimal energy of all structures is higher, and the new landscape $\varepsilon(N)$ is more complex, with many local minima competing specially below $N < 50$. These differences are a direct consequence of the impossibility of combining capsomers of different size.

More specifically, the pronounced minima in the landscape $\varepsilon(N)$ include those at $N = 12$, 24, 32, 44, and 48 (see Fig. 2.7). Interestingly, this sequence of *magic numbers* coincides with maxima of the packing density of N hard disks on a spherical surface [36]. Pentons made using only one type of capsomer are energetically more costly than the pentamer-hexamer pentons of the previous scenario (see Figs. 2.8b and 2.8c). Thus the suppression of capsomer switching clearly has a destabilizing effect on icosahedral symmetry. In this scenario the only surviving T-structures are $T = 1$ ($N = 12$) and $T^\star = 3$ ($N = 32$) (see Fig. 2.9a and c). We use the index $T^\star$ to label icosahedral capsids with $T > 1$ made by only one type of capsomer. These capsids are not strictly included

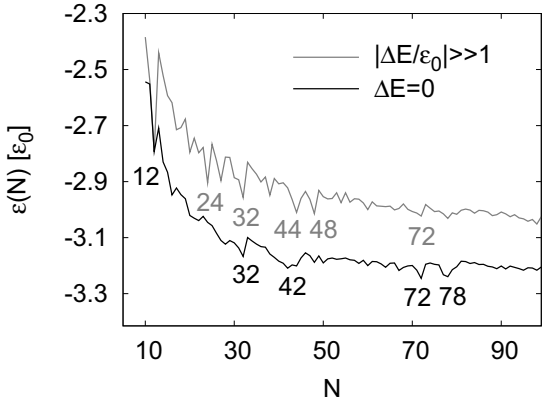


Figure 2.7: Equilibrium landscape for spherical capsids. Energy per capsomer for the optimal configuration of each N capsid. The cases corresponding to one morphological unit, $\Delta E \gg \varepsilon_0$, and for two types of capsomers (hexamers and pentamers), $\Delta E = 0$, are plotted.

in the CK-construction based on coat proteins, but capsomers occupy the same position in both cases and the basic rule $N = 10\,T + 2$ still applies. However, the penalty associated to pentons gives rise to alternative type of defects (see Fig. 2.8d). Interestingly, the $N = 24$ and 48 minima have octahedral symmetry (see Table 1.2 in Chapter 1), characterized by the presence of holes with local 4-fold axial symmetry (see Fig. 2.9e). In particular, the $N = 24$ minimum has the symmetry of a chiral octahedral Archimedean solid known as the *snub-cube*, and the $N = 48$ has an equivalent organization of capsomers, duplicating the number of capsomers, and is called *snub cuboctahedron*. Finally, the $N = 44$ structure has cubic symmetry (*truncated rhombic dodecahedron*), once again with holes with 4-fold axial symmetry (see Fig. 2.9d).

Interestingly, at $N = 72$ we find a local minima that is not very pronounced but it is worth to comment. The structure resembles a $T^\star = 7$, but with some defects (see Fig. 2.9f). The introduction of a small compression ($R < R_e(72)$), favors the presence of icosahedral symmetry, because the stress can be optimally absorbed in the 5-fold sites. Surprisingly, at $N = 42$ we do not have any trace of the intermediate CK-structure $T^\star = 4$. The reason is related also to the absence of $T = 9$ in the pentamer-hexamer scenario. Both cases, $T = 4$ and $T = 9$, belong to the class $P = 1$, which for spherical shells accumulates much more stress in the 5-fold sites in comparison with rest of classes, as we will see in Chapter 4.

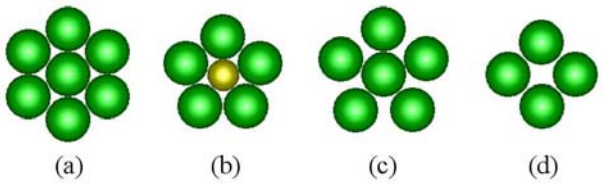


Figure 2.8: Common local configuration of capsomers. (a) Hexon of hexamers (equivalent for all-pentamers). (b) Penton combining one pentamer and 5 hexamers. (c) Penton made of hexamers (equivalent for all-pentamers). (d) Hole with 4-fold local symmetry.

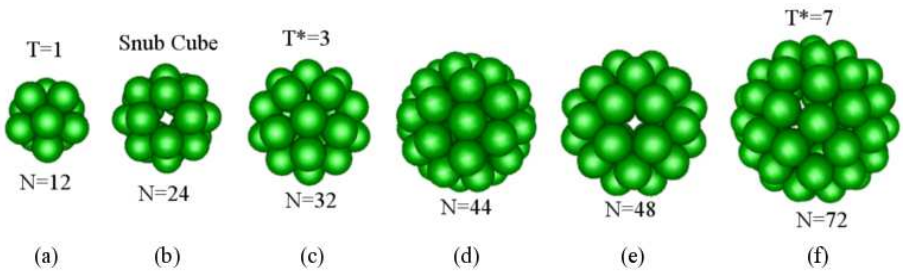


Figure 2.9: Gallery of optimal structures for capsids with one morphological unit ($\Delta E \gg \varepsilon_0$).

The icosahedral structure $T = 1$ is equivalent for the two scenarios explored, and is adopted by many viruses [1]. Interestingly, there are some viruses that have an all-pentamer capsid with $T^\star = 7$, for instance polyoma virus [57], human papilloma virus [1], and simian virus SV40 [58]. Moreover, in vitro experiments performed by Salunke *et al.* in Ref. [11] on the self-assembly of polyoma virus at different conditions, have shown the possibility of getting different empty all-pentamer structures with the same building blocks. More specifically, they obtained the wild type $N = 72$ with $T^\star = 7$, the common $N = 12$ with $T = 1$, and also a $N = 24$ shell having the symmetry of a left-handed snub-cube[5] (see Fig. 2.10). All these structures turn out to be local free energy minima in our model.

Therefore, the results obtained in this section seem consistent with the *in vivo* and *in vitro* experiments commented above. In addition, it has been shown that

[5]More recently, Plevka *et al.* obtained octahedral snub cube particles in the crystals of a bacteriophage MS2 coat protein mutant, rather than the $T = 3$ capsids adopted in the wild type form [59].

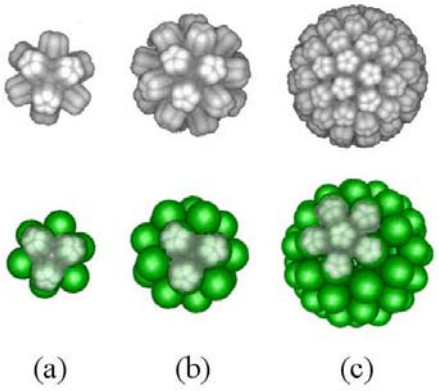


Figure 2.10: Polymorphism of the polyoma capsid. The structures above show the three X-ray reconstructed capsids for the *in vitro* self-assembled polyoma empty capsids [11]. Below there are the correspondent equilibrium structures obtained as local minima in Fig. 2.7. In particular $N = 12$ ($T = 1$) (a), $N = 24$ (snub cube) (b), and $N = 72$ ($T^\star = 7$) (c). We have superimposed a fragment of the original capsid in the capsid models to facilitate the comparison.

the coexistence of two different types of morphological units, i.e., pentamers and hexamers, is not absolutely required to obtain icosahedral capsids with $T > 1$. Nevertheless, in the one-capsomer scenario the formation of pentons is more costly, which has a clear destabilizing effect on the icosahedral symmetry. This favors the emergence of other symmetrical structures and dramatically complicates the spectrum of optimal structures (see Fig. 2.7).

2.2.5 Discussion

So far we have investigated the catalog of optimal equilibrium structures of viral capsids. However, the selection of a specific structure N, or equivalently the selection of one energy minima in Fig. 2.7, involves some extra mechanisms in addition to the basic interaction proposed in our model. Depending on the viral species those mechanisms could be related, for example, to the existence of a preferred "spontaneous curvature" among the coat subunits that can depend on the assembly conditions [11], to the presence of scaffolding proteins that impose a particular radius [20]; or by the size of the enclosed genome [60]. Then, these effects restrict the size of the capsid to a particular range of radii in the neighborhood of one of the local energy minima, for instance, in the range $N = 70 \pm 5$ that

contains the icosahedral solution $T = 7$. But capsids of native virus tend to be perfectly monodisperse, so another intriguing question is how this monodisperisty is achieved, or, referring to the previous example, why the $T = 7$ structure with $N = 72$ is preferentially selected compared to its neighbors $N = 71$ or $N = 73$. Looking at Fig. 2.4, one can see that the energy per capsomer of the $N = 72$ structure differs by only $\approx 0.05\varepsilon_0$ from the chemical potential associated to its neighbors $N = 71$ and $N = 73$. This could seem a small difference, but one should not forget that the total energy difference is 72 times larger, i.e., around $\approx 54\,kT$. Since at equilibrium, the probability of finding a particular structure follows a Boltzmann distribution, the relative abundance of the $N = 71$ and $N = 73$ capsids will be completely negligible, and the minimum of $\varepsilon(N)$ at $N = 72$ will ensure the production of nearly monodisperse $T = 7$ capsids.

Therefore, under this thermodynamic framework it is easy to justify why the coat proteins of some spherical viruses are able to produce alternative icosahedral or highly symmetrical structures when assembled at different conditions [11, 30, 53], or when the regions of the coat proteins involved in the preferred curvature are changed by point mutations [61, 62].

There are two important remarks that have to be made at this point. The first one is that the present model only aims to describe what are the optimal final structures for spherical viruses. The pathways and the precise process of capsid formation lies outside the scope of this model, and will be analyzed in more detail in the third part of this thesis. The second point is that we are not claiming that the Lennard-Jones like potential used in the model is the actual interaction between capsomers in real viruses. Nevertheless, it captures the crucial elements required for a successful capsid self-assembly, i.e, the attraction needed for the aggregation and the excluded volume repulsion associated to the steric effects between capsomers. Remarkably, the exploration of other similar potentials [33], also with different ranges of attraction yield the same results [63]. The truly remarkable conclusion is that the emergence of icosahedral structures of viral capsids seems to be a general consequence of free energy minimization, irrespective of the particular details of the interaction. Indeed, the typical interactions associated to the attraction between proteins, like hydrophobic or hydrogen bonds, are thought to be of a relatively short range. A part from the possible cooperative hydrophobic effects promoted by the solvent during assembly, electrostatic interactions play also a relevant role in the stabilization of capsids [64]. Interestingly, it has been shown that the combination of a short range hydrophobic interaction with an emergent longer range electrostatic repulsion in the aggregation state is an efficient mechanism to stabilize the formation of hollow shells [65, 66]. In any

case, this issue and the influence of more realistic details, like the shape of coat subunits and their intrinsic anisotropic interactions, are out of the scope of this thesis, and will be addressed in future investigations.

2.3 Prolate capsids

Now that the efficiency of our simple model in exploring the landscape of optimal spherical capsids has been shown, we turn into the less studied scenario of prolate viral structures.

The study of elongated structures is more challenging than that of spherical structures from either an experimental and a theoretical point of view. However, the exploration of coarse-grained models have paved the way to get new insights in the physical origin of such structures. Indeed, a phenomenological model was recently developed by Chen and Glotzer [67] to study the optimal packing of identical hard spheres with square-well interactions on a prolate spheroid surface with convexity constraints [67]. They were able to show that the elongated optimal structures composed of 15, 17, 18, and 42 identical units, resemble the hypothetical structures of aberrant FHV [68] and of native ϕ29 [69], although these viruses have two types of morphological units rather than one. Further, Nguyen and Brooks [70] studied the polymorphism observed in the self-assembly of viral capsids, using a simplified model for capsomers. Besides spherical T-number capsids, they obtained some elongated structures including that of the maize streak virus (MSV) [71] as well as some prolates with structures similar to those of FHV and ϕ29, although combining hexamers and more than 12 pentamers. More recently, Fejer *et al.* obtained some elongated shells as energy minima for a generic anisotropic interaction between capsomers of a unique size [72].

However, the range of possible prolate structures, the validity conditions for the geometrical construction introduced in Chapter 1, and the physical principles justifying the appearance of prolate architectures were not fully known. Thus, the goal of this section is to explore what are the energetically optimal structures for prolate viruses. To that end, we have adapted and extended the minimal model of interactions introduced in the previous section based on Ref. [33].

2.3.1 Minimal model of prolate capsids

To study the equilibrium structures for bacilliform capsids, we have adapted the model introduced for the spherical case. The main difference consists on the geometrical template that constraints the movement of the capsomers. Capsomers

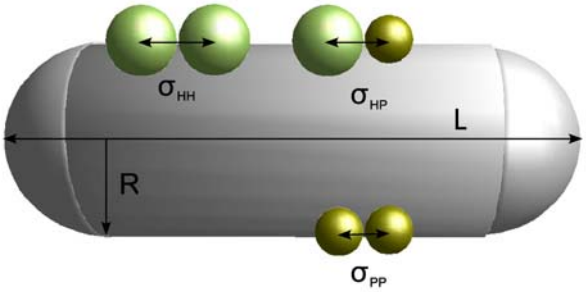


Figure 2.11: Spherocylindrical template. The surface is determined by the radius R and length L. All possible capsomer-capsomer interactions are shown, indicating in each case the equilibrium distance.

are now forced to be on the surface of a spherocylinder made of two hemi-spherical caps of radius R and a cylindrical body of radius R and length L (see Fig. 2.11). As in the spherical study, this surface restriction will allow a quick exploration of the possible equilibrium structures, but presents limitations at big radii, where optimal structures are expected to be more faceted (see Chapter 4). Notice that in the spherical case, for each N it was only necessary to explore the radius to obtain the equilibrium structure. Here, instead one has to deal also with the length. In particular, for a fixed R one can always find an optimal length L that properly closes a structure of N capsomers. Thus, in this case, each N will have associated a complex energy landscape $\varepsilon_N(R)$ with several energy minima at different radii. This is clearly a more complicated scenario, and it required a bigger effort in the analysis of the possible structures.

The algorithm programmed to study prolate viral capsids works as follows. N hexamers are placed at random positions on the surface of a spherocylinder of body length L and radius R. In order to find the free energy minimum structure, for each N at a fixed value of R, three different types of Monte Carlo moves are attempted: (i) a move to change the position of the capsomer on the surface of the spherocylinder; (ii) a move to change the capsomer size, i.e., switching between the hexamer state and the pentamer state; and (iii) a move to adjust the length L of the cylindrical body[6]. We emphasize again that this Monte Carlo protocol can explore the energy minima, but it is intended to describe neither the actual kinetics nor the pathways involved in the physical process of

[6]To refine the results we also applied a similar Monte Carlo simulation where the length was changed manually.

assembly. Different relative ratios between the different types of moves were tested in order to facilitate the reaching of the optimal structure. Typically, the combination that worked best was 200:5:1 (position:size:length). The system was then allowed to equilibrate (typically for 25000 to 125000 MC steps, and from 10 to 40 repetitions), and both the average and the absolute minimum values of the free energy after equilibration were recorded.

2.3.2 Two different morphological units

Here, we analyze the equilibrium structures for capsids where hexamers and pentamers have the same internal energy $\Delta E = 0$. More specifically, we explored the energy landscape for bacilliform capsids with a total number of capsomers between $12 < N < 100$, and for each N we obtained the optimal values of the length, $l_N(R)$, and free energy per capsomer, $\varepsilon_N(R)$, as a function of the radius R.

The solid curve of Fig. 2.12 shows $N = 82$ as a particular, but representative, example of the spectrum $\varepsilon_N(R)$. In general, all equilibrium structures showed a cylindrical body made of hexamers distributed in a regular hexagonal pattern, and an accumulation of six pentons in each cap, usually made of pentamers (see Figs. 2.13). As in the spherical case, this is a consequence of the Euler's theorem. In the absence of bending penalties associated to the curvature, capsids with small radii and long cylindrical bodies are energetically favorable in the equilibrium spectrum. This is a consequence of the fact that the hexagonal network present in the cylindrical body allows local optimal configurations, with six nearest-neighbors located at an equilibrium distance from each other. Instead, the presence of pentons in the spherical caps leads to higher energies on average, because reduces the coordination number, and usually has associated a significant amount of stress on the capsid [73] (see Chapter 4). However, it is evident from Fig. 2.12 that the energy landscape $\varepsilon_N(R)$ is not smooth, and shows some pronounced local minima at some specific radii. In general, these minima are related to well defined architectures that optimize the free energy, and are a consequence of the discrete nature of capsids and the interactions between the coat units. We notice though that the final equilibrium structure of a capsid does not necessarily corresponds to the global energy minimum structure for a given N, but rather corresponds to one of the different *local minima*, which end up being selected once the bending energy, spontaneous curvature, size of the encapsidated genome [60, 74] or even scaffolding mechanisms [20] are taken into account, as in the spherical case.

Interestingly, the energy landscape of $N = 82$ (see Fig. 2.12) shows four

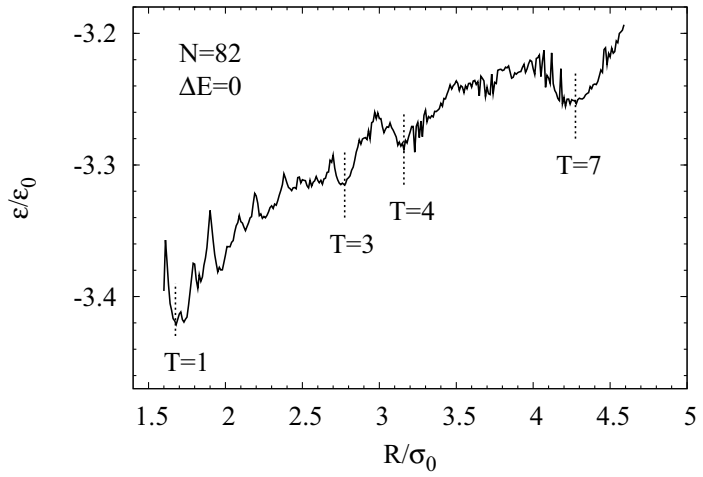


Figure 2.12: Spectrum of the free energy per capsomer $\varepsilon_N(R)$ versus the radius R for a fixed number of capsomers $N = 82$. The curve corresponds to the results for the simulation with two types of capsomers ($\Delta E = 0$). The vertical dashed lines indicate the position of icosahedral prolates T_{end}.

pronounced local minima at radii similar to the CK capsids. A close inspection of the structures shows that they are in fact T_{end} prolates centered on a 5-fold axis (see Fig. 2.13), in agreement with the geometrical bacilliform construction (see Chapter 1). When the spectra for different N are compared, most of the pronounced minima are not always present, but appear periodically at similar radii.

These results suggest the existence of a discrete set of special "magic" values for the radius of the structures, which are particularly optimal from the energetic point of view.

Table 2.1 lists the values of the radii that correspond to the deepest local free-energy minima that appear consistently for different values of N. A careful analysis of the corresponding optimal structures at these special values of the radii indicate that twelve pentamers have always been distributed symmetrically on the caps. Note that the energetically favorable structures correspond to having T-numbered caps with cylindrical bodies made of a rolled-up hexagonal sheet of hexamers. In particular, well-defined local free-energy minima were obtained for the radii that correspond to hemi-spherical $T_{end} = 1$, 3, 4 and 7 caps centered

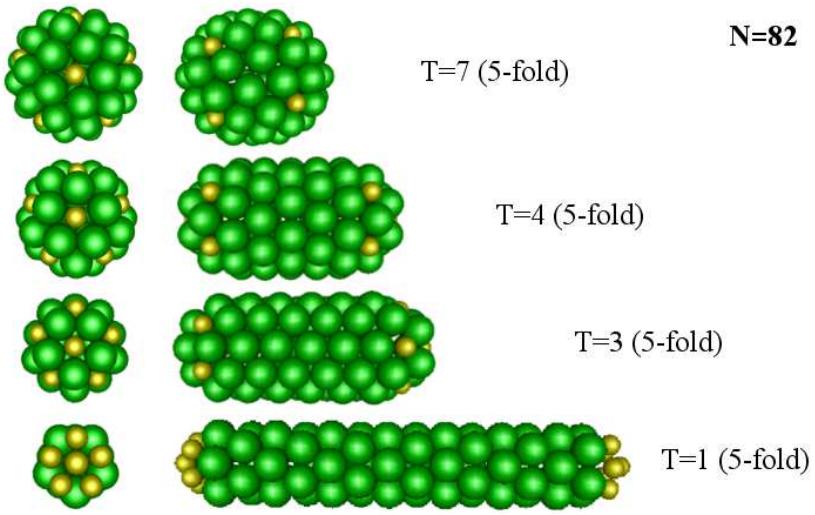


Figure 2.13: Optimal prolates for $N = 82$. We show the pronounced local minima present in the energy landscape $\varepsilon_{82}(R)$ (see Fig. 2.12). They correspond to T_{end} (5-fold) bacilliform capsids.

on a 5-fold axis, coinciding to the lowest triangulation numbers in the CK classification. We also find well-defined minima for radii corresponding to $T_{end} = 1$ and $T_{end} = 3$ with caps centered on a 3-fold axis. In our simulations, prolates with icosahedral caps centered on a 2-fold axis were also observed but only for $T_{end} = 1$. These structures are unusually distorted, and apparently unfavorable for larger caps. Interestingly, for a fixed T_{end} the average radii of the different axial symmetries are slightly different, in agreement with the prolate geometrical theory that predicted $R_{5F} < R_{3F} < R_{2F}$ (see Chapter 1).

Fig. 2.14 shows a gallery with one example for each type of icosahedral capped prolate obtained in our simulations. We observe that all bodies are made of hexamers arranged either in rings (R_n) or zigzag (Z_n) layers of n capsomers, with the exception of $T_{end} = 1$ (2-fold) and $T_{end} = 7$ (5-fold) that have skewed bodies (see also Table 2.1). These are precisely the body architectures that we expected from the prolate geometrical model.

Plots of the minimized energy per capsomer versus the number of capsomers, at a fixed radius, reveal a well-defined sequence of energy minima associated with a specific cap configuration. Fig. 2.15 illustrates, as an example, a plot of $\varepsilon(N)$ versus N for a fixed radius, R, that corresponds to a $T_{end} = 7$ cap centered on a 5-fold axis. The smallest optimal prolates associated with these energy minima

$< R >$	Cap	Body	N^0_{seq}	ΔN	$< \Delta L >$
1.668	$T_{end} = 1$ (5-f)	R_5	12	5	1.68
1.673	$T_{end} = 1$ (3-f)	Z_6	12	3	0.97
1.674	$T_{end} = 1$ (2-f)	skew	12	2	0.61
2.761	$T_{end} = 3$ (5-f)	Z_{10}	32	5	1.00
2.875	$T_{end} = 3$ (3-f)	R_9	32	9	1.72
3.205	$T_{end} = 4$ (5-f)	R_{10}	42	10	1.71
4.243	$T_{end} = 7$ (5-f)	skew	72	5	0.70

Table 2.1: Properties of the optimal sequences corresponding to the case where hexamers and pentamers have the same internal energy ($\Delta E = 0$). $< R >$ is the average radius of the sequence; *Cap* indicates the structure (the T number) of the hemispherical caps, as well as the axial symmetry for the icosahedral ones; *Body* describes the pattern of hexamers in the cylindrical part; N^0_{seq} is the number of capsomers of the spherical seed of the sequence; ΔN is the number of hexamers added to the body for consecutive particles in a sequence, and $< \Delta L >$ is the length step obtained from the simulation.

are also shown in the figure. It is remarkable that the difference in the number of capsomers between these structures is fixed and equal to $\Delta N = 5$, which is in agreement with the predictions of Chapter 1 for the construction of prolate viruses with $T_{end} = 7$ caps centered on 5-fold symmetry [5, 77]. Our results confirm the different rules predicted in the previous chapter for the prolates that are closed by T-numbered caps centered on the 3-fold or 2-fold axes, listed also in Table 2.1.

Remarkably, the results of our simulations indicate that not all icosahedrally-capped prolates that can be geometrically constructed seem to be energetically optimal. In particular, we do not obtain 3-fold prolates for $T_{end} > 3$, and 2-fold architectures seem only thermodynamically relevant for $T_{end} = 1$. To understand this phenomenon it is useful to compare the energies of prolates with the same T_{end}-number for the different axial symmetries. Fig. 2.16 plots the sequences corresponding to $T_{end} = 1$ and $T_{end} = 3$. Surprisingly, for $T_{end} = 1$ the 3-fold structure has a slightly lower energy than the 5-fold one, and the energy of the 2-fold particles is not that high, and some of them are in fact locally optimal structures. For $T_{end} = 3$, the 3-fold structures still appear as locally optimal architectures for particular numbers of capsomers, despite having in general higher energies than those centered on the 5-fold axis. In contrast, for $T_{end} = 3$ the 2-fold symmetry does not appear as an optimal structure in our simulations, and

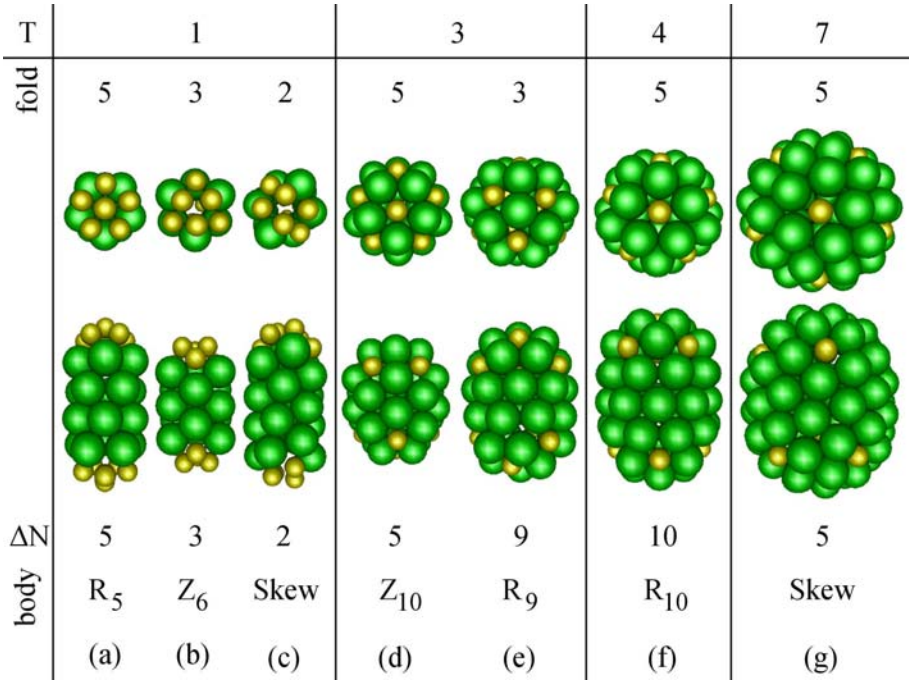


Figure 2.14: Gallery of optimal prolate capsids based on icosahedral caps for $\Delta E = 0$, where hexamers are colored in green, and pentamers are colored in gold. The top images show a zenithal view of the cap, whereas the lower images show a lateral view of a representative example of the complete bacilliform. For each cap, the arrangement of hexamers in the body as well as the minimum step in the number of capsomers ΔN are also indicated. The structures have been rendered using gOpenMol [75, 76].

for $T_{end} > 3$ only the 5-fold symmetry appears to be an optimal architecture. Moreover, it turns out that for $T_{end} > 1$ the 5-fold based structure is consistently the one having the smallest energy, followed by that of the 3-fold and 2-fold (see Fig. 2.16). The divergence in energy for the different prolates seems related to the different distribution of pentamers in the caps. The 5-fold case is the most efficient absorbing the stress induced by the spherical curvature, while the 2-fold case is clearly the worst configuration.

Finally, we have analyzed other less pronounced free-energy minima. They also show precise radii of construction and growing rules ΔN. These non-icosahedral capped structures usually have highly symmetrical tubular bodies

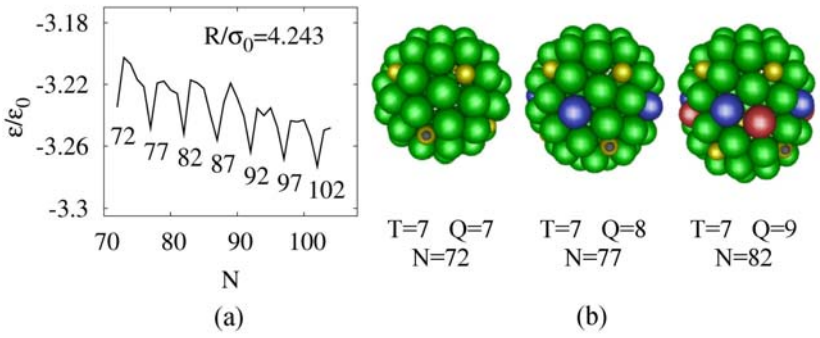


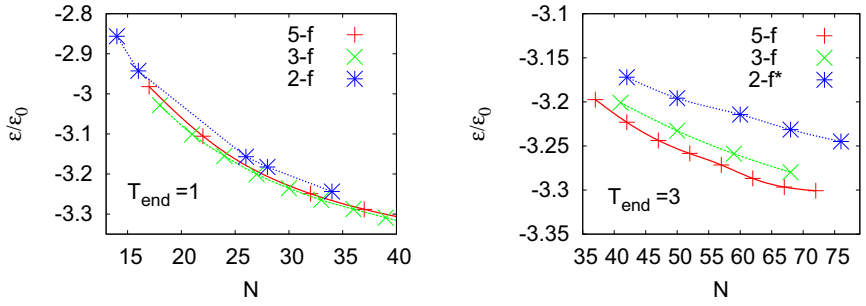
Figure 2.15: Sequence of energy minima at fixed radius. (a) Free energy per capsomer $\varepsilon_R(N)$ versus number of capsomers N at fixed radius R, corresponding to the case $T_{end} = 7$ with two different morphological units. (b) Images of the structures corresponding to the first three minima of the sequence: a spherical $T = 7$ capsid, and two prolate capsids with $T_{end} = 7$, and $Q_{5F} = 8$ and $Q_{5F} = 9$, respectively. At each step hexamers introduced in the body are colored in blue ($Q_{5F} = 8$) and red ($Q_{5F} = 9$). Moreover, the highlighted pentamer illustrates the rotation of the inferior cap.

made of rings or zigzag layers of hexamers. Interestingly, the nonicosahedral capsid of such prolates seem to be associated with the shallow minima in the spherical study, enhanced by the well-formed bodies. We remark that both the icosahedral spheres and prolates produced deeper energy minima, which supports the dominance of icosahedral symmetry among viral capsids. However, these non-icosahedral structures should not be discarded as possible competitors in some circumstances.

In the block 2.3.4 we discuss in detail some experimental findings under the perspective of these results.

2.3.3 One type of morphological unit

Some viruses, such as those of the polyomaviridae family or the maize strike virus (MSV), are made from only one type of capsomer (pentamers in the cited cases), and can assemble into elongated particles [71,78,79]. To study the set of possible optimal structures for these capsids one can again increase the internal energy difference between pentamers and hexamers, $|\Delta E/\varepsilon_0| \gg 1$. In the spherical case (see Section 2.2.1 and Ref. [33]), this led to a dramatic reorganization of the



(a) (b)

Figure 2.16: (a) Sequence of local free energy minima for $T_{end} = 1$ bacilliform particles centered on each of the possible axes of symmetry (the lines are just a guide for the eye). (b) Same plot for $T_{end} = 3$, where the 2-fold structures have been computed based on our geometrical model.

optimal structure spectrum, where non-icosahedral capsids strongly competed with icosahedral ones.

That is also the case for the spherocylindrical model. As an example, the gray line in Fig. 2.17 shows the results of the optimal energy per capsomer as a function of the radius $\varepsilon_N(R)$, for $N = 82$ and $|\Delta E/\varepsilon_0| \gg 1$. The energy is lower for small radii prolates (long bodies), and some well defined energy minima are present.

As in the two capsomer case, an exhaustive analysis was performed for different values of N. Prolates based on caps $T^*_{end} = 1,\ 3$, and 7 centered on both 5-fold and 3-fold symmetries were found[7]. However, now the 2-fold structure is not even present for $T_{end} = 1$, and the 3-fold construction competes also for $T^\star > 3$. Interestingly, here the cylindrical body has stabilized the $T^\star_{end} = 7$ cap bacilliforms, contrarily to the spherical case where a slightly distorted $T^\star = 7$ was obtained at $N = 72$. Nevertheless, as in the spherical shell with only one type of capsomer, no $T^*_{end} = 4$ structure was found. This is in agreement with the fact that spherical caps within the $P = 1$ class accumulate more stress than other classes (as will be shown in Chapter 4), a feature that is accentuated for big shells in general, and specially for capsids made of only one capsomer type. The main properties of these icosahedral prolates are summarized in Table 2.2 and for each case an example is shown in Fig. 2.18.

[7]Recall that the triangulation number T^* indicates that the distribution of capsomers is equivalent to the classical T, but with only one type of capsomer.

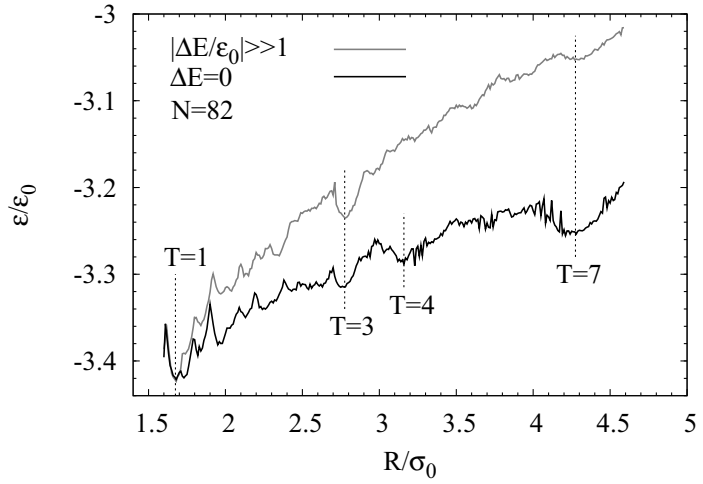


Figure 2.17: Spectrum of the free energy per capsomer $\varepsilon_N(R)$ versus the radius R for a fixed number of capsomers $N = 82$. The solid line corresponds to the results for the simulation with two types of capsomers ($\Delta E = 0$), and the gray line is for one type of capsomer ($|\Delta E/\varepsilon_0| \gg 1$).

Despite the fact that these structures are made by only one kind of capsomer their capsomers occupy again the same positions expected from the generalized prolate geometrical model. Moreover, the same selection rules ΔN, and the rest of geometrical properties for icosahedral capped structures are recovered (see Tables 2.1 and 2.2).

As in the spherical case, non-icosahedral capped prolates become particularly relevant for the one-capsomer scenario. In particular, a bacilliform structure based on a snub cube centered on a 3-fold axis was obtained periodically at steps $\Delta N = 3$ (see Figs. 2.9 and 2.18). For bigger radii the bacilliform shape favors also the formation of stable structures that seem based on the spherical capsid $N = 44$ with $\Delta N = 6$ (see Fig. 2.9), and 6-fold axial symmetry (which was a local symmetry embedded in a 2-fold axis in the spherical structure). This prolate architecture is stable enough to suppress the presence of a possible alternative architecture based on the spherical capsid $N = 48$. The geometrical properties of all these bacilliform shells are summarized in Table 2.2, and the structures are illustrated in Fig. 2.18.

$< R >$	Cap	Body	N^0_{seq}	ΔN	$< \Delta L >$
1.703	$T^*_{end} = 1$ (5-f)	R_5	12	5	1.72
1.724	$T^*_{end} = 1$ (3-f)	Z_6	12	3	1.01
2.593	Snub Cube	skew	24	3	0.69
2.789	$T^*_{end} = 3$ (5-f)	Z_{10}	32	5	1.01
2.912	$T^*_{end} = 3$ (3-f)	R_9	32	9	1.75
3.344	Cubic	Z_{12}	44	6	1.03
4.235	$T^*_{end} = 7$ (5-f)	skew	72	5	0.65
4.244	$T^*_{end} = 7$ (3-f)	skew	72	3	0.39

Table 2.2: Results for the optimal sequences of spherocylinders with one kind of capsomers ($|\Delta E/\varepsilon_0| \gg 1$). The columns have the same meaning as in Table 2.1.

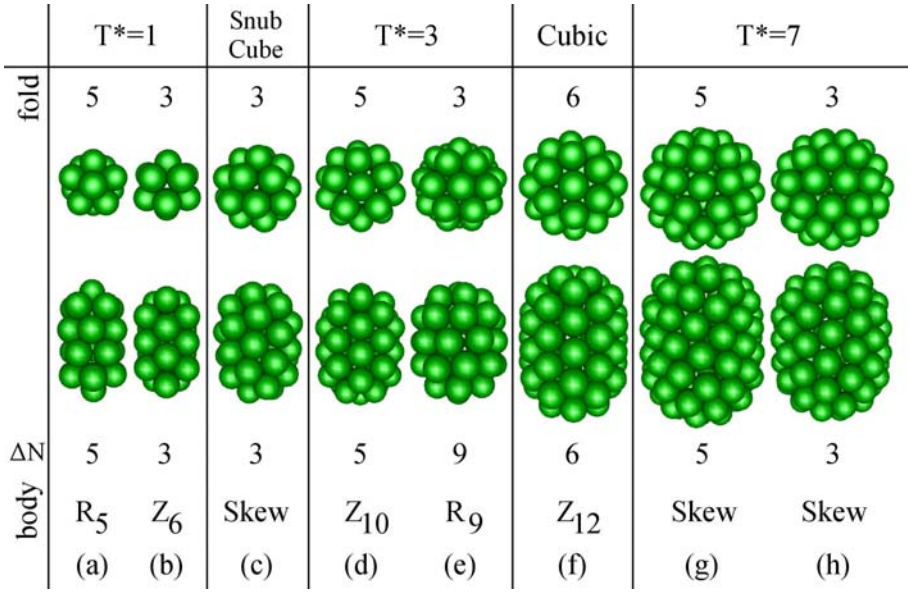


Figure 2.18: Gallery of optimal prolate capsids for $|\Delta E|/k_BT \gg 1$, i.e, for one morphological unit, where capsomers are all pentamers. The top images show a zenithal view of the cap and the bottom images a lateral view of the complete bacilliform. For each cap, the arrangement of hexamers in the body as well as the minimum step in the number of capsomers ΔN are also indicated. The structures have been rendered using gOpenMol [75, 76].

2.3.4 Equilibrium map of prolates

The results of our simple physical model derived above complement the geometrical model of Chapter 1, and show that some prolates, which are geometrically feasible, are not probable structures from an energetic point of view. Moreover, it introduces some alternative non-icosahedral structures, that were not proposed before, but, that could compete with the icosahedral elongated ones. Combining these two models, we summarize in Fig. 2.19 the map of possible optimal structures for bacilliform capsids up to $T = 7$.

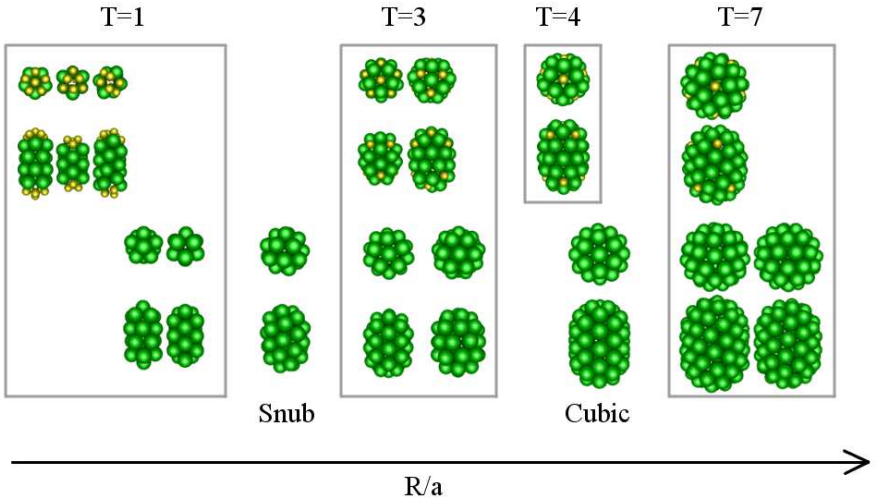


Figure 2.19: Equilibrium map of prolates. We combine the most relevant results for one and two capsomer types bacilliform capsids, and organize the structures in terms of the reduced radius, i.e, R/a, as in the geometrical model.

In the previous chapter the case studies for bacilliform capsids composed of two type of capsomers were practically exhausted. In the context of these optimal structures obtained from the physical model, let's now discuss again the structures of some bacilliform capsids made of two types of capsomers, as well as some all-pentamer elongated viruses.

Prolates with two types of capsomers

Our thermodynamic study confirms that the 5-fold icosahedral bacilliform construction suggested by Moody [77] for bacteriophages is, in general, the most

stable elongated architecture, and follows precisely the geometrical rules derived in Chapter 1. In particular, we recover the structure of $\phi 29$ [69], which is a $T_{end} = 3$ (5-fold) $Q_{5F} = 5$ (Fig. 2.14d). The convenience of the 5-fold geometry was discussed for tailed bacteriophages, but this construction should be also efficient for other types of viruses.

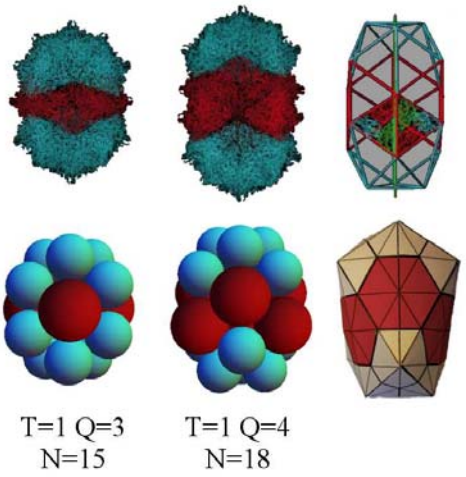


Figure 2.20: Aberrant capsids of flock house virus (FHV). The top row shows two elongated structures proposed for FHV mutants based on the X-ray reconstruction of the wild type $T = 1$ [68]. The associated optimal structures of both FHV mutants obtained in the simulations. The two capsids correspond to $T = 1$ (3-fold) with $Q_{3F} = 3$ and 4, respectively. On the right a polyhedral construction of the bigger mutant is shown.

Interestingly, 3-fold prolates are also optimal structures, at least for small T_{end}. We emphasize that for $T_{end} = 1$ the 3-fold structure has even a lower energy than the 5-fold construction (see Fig. 2.16a), which might justify why several elongated plant viruses, like AMV [80], or some aberrant structures of flock house virus FHV [68] (see Fig. 2.20) seem to adopt a capsid based on a 3-fold rather than the typical 5-fold construction of bacteriophages [5]. For $T_{end} = 3$, 3-fold bacilliform structures are still equilibrium structures, and it has been suggested that some plant viruses, for instance RTBV [81] and badnavirus genus [82], have this particular structure. Finally, 2-fold particles are at least local optimal structures for $T_{end} = 1$, which supports the possibility of a 2-fold

structure of the AMV[8].

Elongated Papovaviridae particles

In the infection process of Papovaviridae viruses, which form all-pentamer capsids, hollow tubular particles, usually with rounded ends, have been observed and characterized experimentally. In particular, Kiselev and Klug observed mostly narrow tubes in human wart virus, wide tubes in polyoma virus, and both kind of tubes in rabbit papilloma virus [83]. The narrow tubes have a diameter about 300 Å, and are made of paired pentamers placed in a very distorted hexagonal network. On the other hand, the diameter of the wide tubes is about 450 to 550 Å , and pentamers in the body show a clear hexavalent-like configuration, although they are still paired [78, 79] (Fig. 2.21). Due to the hexagonal configuration of the tube, we focus on the wide prolate.

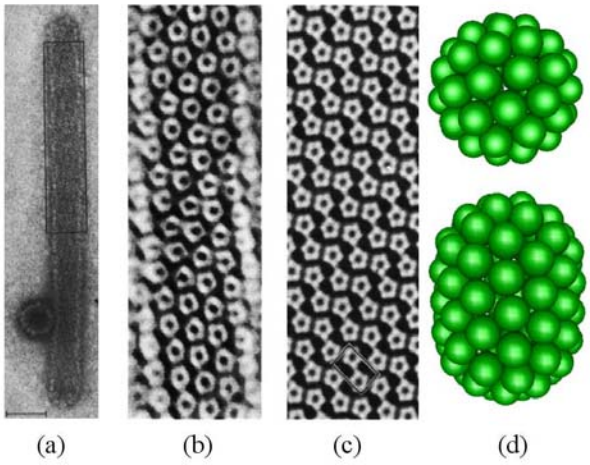


Figure 2.21: (a) Electron micrograph of the most frequent class of "hexamer" tube, which is capped at both ends and appears adjacent to an icosahedral $T^\star = 7$ capsid (extracted from Ref. [78]). The diffraction was performed on the boxed area, which was flattened. (b) Filtered image obtained from the diffraction pattern of (a). (c) Computer planar model from regular pentamers arranged in a p2 lattice. (d) Architecture proposed based on our analysis: icosahedral prolate capped by a $T^\star_{end} = 7$ centered on a 3-fold axis.

[8]See Section 1.3.3 in Chapter 1 for the detailed geometrical discussion about AMV and RTBV.

The distance between the morphological units is about 90 to 100 Å. From this one obtains an approximate radius $R/\sigma_0 \approx 4.5 - 5.5$, which is similar to the $T^{\star}_{end} = 7$ bacilliform capsids of the physical model (see Table 2.2). This is not surprising since the size of these tubes is similar to the wild type spherical shell, which is precisely a $T^{\star} = 7$. In addition, the tube is skewed with a chiral angle 47.73^o, which is very close to the theoretical prediction 49.11^o for the 3-fold architecture for that T_{end} (see Chapter 1). Thus, Papovaviridae wide tubes seem to be $T^{\star}_{end} = 7$ prolates centered on the 3-fold axis (see Fig. 2.18h). We notice that a study based on a tiling theory also pointed to a structure based on a $T^{\star}_{end} = 7$, although the axial symmetry was not confirmed [84, 85].

Geminivirus

The structure of geminate viruses, and the effect of genome length on the size of their capsids, is quite remarkable. For example, both maize streak virus (MSV) and African cassava mosaic virus (ACMV) [86] adopt a spherical $T = 1$ shell, but form geminate capsids when the full length genome is present [71]. A geminate capsid is composed of 22 pentamers and its structure resembles a $T^{*}_{end} = 1$ (5-f), as shown in Fig. 2.22.

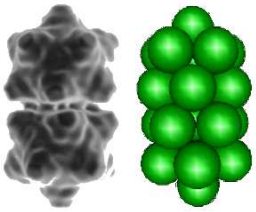


Figure 2.22: X-ray reconstruction of MSV [71] and the corresponding $N = 22$ prolate capsid based on a $T_{end} = 1$ (5-fold) and $Q_{5F} = 3$.

Nevertheless, we remark that the two halves of MSV seem slightly separated and rotated compared to our $N = 22$ particle. This distortion might be caused by the encapsidated ssDNA.

2.4 Conclusions

In summary we find that the minimization of the free energy of a very simple physical model, which captures the essential ingredients of viral assembly, reproduces the structures of spherical and prolate viruses observed experimentally,

both *in vivo* and *in vitro*. For spherical shells, the work of Zandi *et al.* in Ref. [33] already showed that the origin of icosahedral symmetry was intimately related to the free energy minimization of a generic interaction. Here, we have extended these ideas to the case of elongated structures. This study has been particularly challenging, since the inclusion of a new degree of freedom, the length of the bacilliform body, dramatically increased the effort to explore the map of optimal equilibrium solutions. In addition, contrarily to the CK construction of spherical viruses, there was not a complete catalog of icosahedral capped prolates to compare our results. This motivated the development of the extended geometrical model of elongated capsids discussed in Chapter 1.

More specifically, our study reveals that the icosahedrally-capped prolates centered on 5-fold axial symmetry proposed by Moody [77], and adopted by some bacteriophages, like $T4$ and $\phi 29$, are indeed free energy minima structures. Moreover, we also find that bacilliform viruses with icosahedral caps centered on 3-fold and, exceptionally, on 2-fold axis, are also optimal structures and compete with the 5-fold architecture, at least, for small T caps. Interestingly, we have found in the literature several examples of elongated viral capsids, especially among plant viruses, that seem to adopt a $T = 1$ (3-fold) configuration, like AMV and aberrant FHV, or a $T = 3$ (3-fold) prolate structure, like that of RTBV.

Remarkably, the icosahedral prolate structures obtained in the simulations follow precisely the selection rules for the length, the structure of the body, as well as for the total number of capsomers N that were derived in the geometrical prolate model in Chapter 1.

The physical model also allowed the exploration of capsids made of only one type of capsomer, in correspondence with all-pentamer viruses. In these cases icosahedral-capped structures were also energy minima, except for $T = 4$, and followed again the rules of the geometrical model. The 2-fold configuration was not an optimal structure this time, but the 3-fold architecture competed with the 5-fold one even for the largest T prolates explored. A notable example are the elongated capsids of polyomavirus observed *in vivo*, which based on our analysis seem to be a $T^{\star} = 7$ (3-fold) architecture. In addition, in the all-pentamer scenario some non-icosahedral capped structures were particularly relevant and competed energetically with the T-number capped prolates.

Thus, the combination of the physical model and the geometrical construction brings an extended catalog to classify prolates capsids that should be helpful for the experimental determination of structures. In particular, our results shed some light on the structure of several elongated viruses, such as $\phi 29$, AMV, aberrant

FHV, RTBV, polyoma tubular particles, and geminate viruses, many of which have not yet been fully characterized.

Finally, in a more general perspective, the structural insights provided by the results of our work open the door to the design and control of the structure and dimensions of viral-like particles, which can be extremely useful as nanotemplates and customized containers in nanotechnological applications.

References

[1] T. S. Baker, N. H. Olson, and S. D. Fuller, "Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs," *Microb. and Molec. Bio. Rev.*, vol. 63, pp. 862–922, 1999.

[2] ICTVdB, "The Universal Virus Database, version 4." http://www.ncbi.nlm.nih.gov/ICTVdb/ICTVdB/index.htm.

[3] C. A. Tidona and G. Darai, eds., *The Springer index of viruses.* Springer -Verlag Berlin Heidelberg, 2002.

[4] D. H. Bamford, J. M. Grimes, and D. I. Stuart, "What does structure tell us about virus evolution?," *Curr. Opin. Struc. Biol.*, vol. 15, pp. 655–63, 2005.

[5] M. F. Moody, "Geometry of phage head construction," *J. Mol. Biol.*, vol. 293, pp. 401–33, 1999.

[6] R. Hull, "The structure of tubular viruses," *Adv. Virus. Res.*, vol. 20, pp. 1–32, 1976.

[7] J. B. Bancroft, C. E. Bracker, and G. W. Wagner, "Structures derived from cowpea chlorotic mottle and brome mosaic virus protein," *Virology*, vol. 38, pp. 324–335, 1969.

[8] K. W. Adolph and P. J. G. Butler, "Studies on the assembly of a spherical plant virus (I. States of aggregation of the isolated protein)," *J. Mol. Biol.*, vol. 88, pp. 327–338, 1974.

[9] M. Cuillel, C. Berthet-Colominas, B. Krop, A. Tardieu, P. Vachette, and B. Jacrot, "Self-assembly of brome mosaic virus capsids. Kinetic study using neutron and X-ray solution scattering," *J. Mol. Biol.*, vol. 164, pp. 645–50, 1983.

[10] C. Berthet-Colominas, M. Cuillel, M. H. J. Koch, P. Vachette, and B. Jacrot, "Kinetic study of the self-assembly of brome mosaic virus capsid," *Eur. Biophys. J.*, vol. 15, pp. 159–168, 1987.

[11] D. M. Salunke, D. L. D. Caspar, and R. L. Garcea, "Polymorphism in the assembly of polyomavirus capsid protein VP1," *Biophys. J.*, vol. 56, pp. 887–900, 1989.

[12] Z. Xie and R. W. Hendrix, "Assembly in vitro of bacteriophage HK97 proheads," *J. Mol. Biol.*, vol. 253, pp. 74–85, 1995.

[13] B. K. Ganser, S. Li, V. Y. Klishko, J. T. Finch, and W. I. Sundquist, "Assembly and analysis of conical models for the HIV-1 core," *Science*, vol. 283, pp. 80–3, 1999.

[14] A. Zlotnick, R. Aldrich, J. M. Johnson, P. Ceres, and M. J. Young, "Mechanism of capsid assembly for an icosahedral plant virus," *Virology*, vol. 456, pp. 450–456, 2000.

[15] P. Ceres and A. Zlotnick, "Weak protein-protein interactions are sufficient to drive assembly of hepatitis B virus capsids," *Biochemistry*, vol. 41, pp. 11525–11531, 2002.

[16] S. Mukherjee, M. V. Thorsteinsson, L. B. Johnston, P. A. DePhillips, and A. Zlotnick, "A quantitative description of in vitro assembly of human papillomavirus 16 virus-like particles," *J. Mol. Biol.*, vol. 381, pp. 229–37, 2008.

[17] G. L. Casini, D. Graham, D. Heine, R. L. Garcea, and D. T. Wu, "In vitro papillomavirus capsid assembly analyzed by light scattering," *Virology*, vol. 325, pp. 320–7, 2004.

[18] C. Chen, C. C. Kao, and B. Dragnea, "Self-assembly of brome mosaic virus capsids: insights from shorter time-scale experiments," *J. Phys. Chem. A*, vol. 112, pp. 9405–12, 2008.

[19] P. E. Prevelige, D. Thomas, and J. King, "Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells," *Biophys. J.*, vol. 64, pp. 824–35, 1993.

[20] P. A. Thuman-Commike, B. Greene, J. A. Malinski, J. King, and W. Chiu, "Role of the scaffolding protein in p22 procapsid size determination suggested by T=4 and T=7 procapsid structures," *Biophys. J.*, vol. 74, pp. 559–568, 1998.

[21] K. N. Parent, A. Zlotnick, and C. M. Teschke, "Quantitative analysis of multicomponent spherical virus assembly: scaffolding protein contributes to the global stability of phage P22 procapsids," *J. Mol. Biol.*, vol. 359, pp. 1097–106, 2006.

[22] D. E. Kainov, S. J. Butcher, D. H. Bamford, and R. Tuma, "Conserved intermediates on the assembly pathway of double-stranded RNA bacteriophages," *J. Mol. Biol.*, vol. 328, pp. 791–804, 2003.

[23] H. Fraenkel-Conrat and R. C. Williams, "Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components," *Proc. Natl. Acad. Sci. USA.*, vol. 41, pp. 690–698, 1955.

[24] J. Bancroft and E. Hiebert, "Formation of an infectious nucleoprotein from protein and nucleic acid isolated from a small spherical virus," *Virology*, vol. 32, pp. 354–356, 1967.

[25] R. F. Bruinsma, W. M. Gelbart, D. Reguera, J. Rudnick, and R. Zandi, "Viral self-assembly as a thermodynamic process," *Phys. Rev. Lett.*, vol. 90, p. 248101, 2003.

[26] A. McPherson, "Micelle formation and crystallization as paradigms for virus assembly," *BioEssays*, vol. 27, pp. 447–58, 2005.

[27] T. Keef, A. Taormina, and R. Twarock, "Assembly models for Papovaviridae based on tiling theory," *Phys. Biol.*, vol. 2, pp. 175–88, 2005.

[28] A. Zlotnick, "Viruses and the physics of soft condensed matter," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, p. 15549, 2004.

[29] S. Katen and A. Zlotnick, *The thermodynamics of virus capsid assembly*, vol. 455, ch. 14, pp. 395–417. Elsevier Inc., 1 ed., 2009.

[30] 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*, 2009.

[31] J. Lepault, I. Petitpas, I. Erk, J. Navaza, D. Bigot, M. Dona, P. Vachette, J. Cohen, and F. A. Rey, "Structural polymorphism of the major capsid protein of rotavirus," *EMBO J.*, vol. 20, pp. 1498–507, 2001.

[32] S. Kanesashi, K. Ishizu, M. Kawano, S. Han, S. Tomita, H. Watanabe, K. Kataoka, and H. Handa, "Simian virus 40 VP1 capsid protein forms polymorphic assemblies in vitro," *J. Gen. Virol.*, vol. 84, pp. 1899–1905, 2003.

[33] R. Zandi, D. Reguera, R. F. Bruinsma, W. M. Gelbart, and J. Rudnick, "Origin of icosahedral symmetry in viruses," *Proc. Natl. Acad. Sci. USA.*, vol. 101, pp. 15556–15560, 2004.

[34] P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson, and K. Schulten, "Molecular dynamics simulations of the complete satellite tobacco mosaic virus," *Structure*, vol. 14, pp. 437–49, 2006.

[35] M. Goldberg, "Viruses and a mathematical problem," *J. Mol. Biol.*, vol. 24, pp. 337–338, 1967.

[36] B. W. Clare and D. L. Kepert, "The closest packing of equal circles on a sphere," *Proc. R. Soc. London Ser. A*, vol. 405, pp. 329–344, 1986.

[37] D. A. Kottwitz, "The densest packing of equal circles on a sphere," *Acta Cryst. A*, vol. 47, pp. 158–165, 1991.

[38] J. J. Thomson, "On the structure of the atom: an investigation of the stability and periods of oscillation of a number of corpuscles arranged at equal intervals around the circumference of a circle; with application of the results to the theory of atomic structure," *Philos. Mag.*, vol. 7, pp. 237–265, 1904.

[39] T. Tarnai, "The observed form of coated vesicles and a mathematical covering problem," *J. Mol. Biol.*, vol. 218, pp. 485–488, 1991.

[40] C. J. Marzec and L. A. Day, "Pattern formation in icosahedral virus capsids: the papova viruses and Nudaurelia capensis beta virus," *Biophys. J.*, vol. 65, pp. 2559–77, 1993.

[41] T. Keef, R. Twarock, and K. M. Elsawy, "Blueprints for viral capsids in the family of Polyomaviridae," *J. Theor. Biol.*, vol. 253, pp. 808–816, 2008.

[42] T. Keef and R. Twarock, "New insights into viral architecture via affine extended symmetry groups," *Comput. Math. Methods. Med.*, vol. 9, pp. 221–229, 2008.

[43] T. Keef and R. Twarock, "Affine extensions of the icosahedral group with applications to the three-dimensional organisation of simple viruses," *J. Math. Biol.*, vol. 59, pp. 287–313, 2009.

[44] R. Twarock, "A tiling approach to virus capsid assembly explaining a structural puzzle in virology," *J. Theor. Biol.*, vol. 226, pp. 477–82, 2004.

[45] T. Tarnai, Z. Gaspar, and L. Szalai, "Pentagon packing models for all-pentamer virus structures," *Biophys. J.*, vol. 69, pp. 612–618, 1995.

[46] B. Berger, P. W. Shor, L. Tucker-Kellogg, and J. King, "Local rule-based theory of virus shell assembly," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 91, pp. 7732–6, 1994.

[47] A. Zlotnick, "To build a virus capsid: An equilibrium model of the self assembly of polyhedral protein complexes," *J. Mol. Biol.*, vol. 241, pp. 59–67, 1994.

[48] H. D. Nguyen, V. S. Reddy, and C. L. Brooks, "Deciphering the kinetic mechanism of spontaneous self-assembly of icosahedral capsids," *Nano Lett.*, vol. 7, pp. 338–344, 2007.

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

[50] M. Widom, J. Lidmar, and D. Nelson, "Soft modes near the buckling transition of icosahedral shells," *Phys. Rev. E*, vol. 76, p. 031911, 2007.

[51] D. L. D. Caspar and A. Klug, "Physical principles in the construction of regular viruses," in *Cold Spring Harbor Symp. Quant. Biol.*, vol. 27, pp. 1–24, Cold Spring Harbor Laboratory Press, 1962.

[52] V. S. Reddy, H. A. Giesing, R. T. Morton, A. Kumar, C. B. Post, C. L. Brooks, and J. E. Johnson, "Energetics of quasiequivalence: computational analysis of protein-protein interactions in icosahedral viruses," *Biophys. J.*, vol. 74, pp. 546–558, 1998.

[53] J. B. Bancroft, G. J. Hills, and R. Markham, "A study of the self-assembly process in a small spherical virus formation of organized structures from protein subunits in vitro," *Virology*, vol. 31, pp. 354–379, 1967.

[54] D. Frenkel and B. Smit, *Understanding molecular simulations*. Academic Press, 2002.

[55] D. Wales and T. Bogdan, "Potential energy and free energy landscapes," *J. Phys. Chem.. B*, vol. 110, pp. 20765–76, 2006.

[56] S. Mukherjee, S. Kler, A. Oppenheim, and A. Zlotnick, "Uncatalyzed assembly of spherical particles from SV40 VP1 pentamers and linear dsDNA incorporates both low and high cooperativity elements," *Virology*, vol. 397, pp. 199–204, 2010.

[57] I. Rayment, T. S. Baker, D. L. D. Caspar, and W. T. Murakami, "Polyoma virus capsid structure at 22.5 A resolution," *Nature*, vol. 295, pp. 110–115, 1982.

[58] R. C. Liddington, Y. Yan, J. Moulai, R. Sahli, T. L. Benjamin, and S. C. Harrison, "Structure of simian virus 40 at 3.8 A resolution," *Biophys. J.*, vol. 354, pp. 278–284, 1991.

[59] P. Plevka, K. Tars, and L. Liljas, "Crystal packing of a bacteriophage MS2 coat protein mutant corresponds to octahedral particles," *Prot. Sci.*, vol. 17, pp. 1731–1739, 2008.

[60] Y. Hu, R. Zandi, A. Anavitarte, C. M. Knobler, and W. M. Gelbart, "Packaging of a polymer by a viral capsid: the interplay between polymer length and capsid size," *Biophys. J.*, vol. 94, pp. 1428–36, 2008.

[61] D. Ferreira, R. Hernandez, M. Horton, and D. T. Brown, "Morphological variants of Sindbis virus produced by a mutation in the capsid protein," *Virology*, vol. 307, pp. 54–66, 2003.

[62] I. Saugar, D. Luque, A. Oña, J. F. Rodríguez, J. L. Carrascosa, B. L. Trus, and J. R. Castón, "Structural polymorphism of the major capsid protein of a double-stranded RNA virus: an amphipathic alpha helix as a molecular switch," *Structure*, vol. 13, pp. 1007–1017, 2005.

[63] S. N. Fejer, T. R. James, J. Hernandez-Rojas, and D. J. Wales, "Energy landscapes for shells assembled from pentagonal and hexagonal pyramids," *Phys. Chem. Chem. Phys.*, vol. 11, pp. 2098–2104, 2009.

[64] W. K. Kegel and P. van der Schoot, "Competing hydrophobic and screened-coulomb interactions in hepatitis B virus capsid assembly," *Biophys. J.*, vol. 86, pp. 3905–13, 2004.

[65] A. Verhoeff, M. Kistler, A. Bhatt, J. Pigga, J. Groenewold, M. Klokkenburg, S. Veen, S. Roy, T. Liu, and W. Kegel, "Charge regulation as a stabilization mechanism for shell-like assemblies of polyoxometalates," *Phys. Rev. Lett.*, vol. 99, p. 066104, 2007.

[66] E. Mani, E. Sanz, P. G. Bolhuis, and W. K. Kegel, "Stabilization of nanoparticle shells by competing interactions," *J. Phys. Chem. C*, vol. 114, pp. 7780–7786, 2010.

[67] T. Chen and S. C. Glotzer, "Simulation studies of a phenomenological model for elongated virus capsid formation," *Phys. Rev. E*, vol. 75, p. 051504, 2007.

[68] X. F. Dong, P. Natarajan, M. Tihova, J. E. Johnson, and A. Schneemann, "Particle polymorphism caused by deletion of a peptide molecular switch in a quasiequivalent icosahedral virus," *J. Virol.*, vol. 72, pp. 6024–33, 1998.

[69] Y. Tao, N. H. Olson, W. Xu, D. L. Anderson, M. G. Rossmann, and T. S. Baker, "Assembly of a tailed bacterial virus and its genome release studied in three dimensions," *Cell*, vol. 95, pp. 431–437, 1998.

[70] H. D. Nguyen and C. L. Brooks, "Generalized structural polymorphism in self-assembled viral particles," *Nano Lett.*, vol. 8, pp. 4574–4581, 2008.

[71] W. Zhang, N. H. Olson, T. S. Baker, L. Faulkner, M. Agbandje-McKenna, M. I. Boulton, J. W. Davies, and R. McKenna, "Structure of the maize streak virus geminate particle," *Virology*, vol. 279, pp. 471–7, 2001.

[72] S. N. Fejer, D. Chakrabarti, and D. J. Wales, "Emergent complexity from simple anisotropic building blocks: shells, tubes, and spirals," *ACS nano*, vol. 4, pp. 219–228, 2010.

[73] R. Zandi and D. Reguera, "Mechanical properties of viral capsids," *Phys. Rev. E.*, vol. 72, p. 021917, 2005.

[74] R. Zandi and P. van der Schoot, "Size regulation of ssRNA viruses," *Biophys. J.*, vol. 96, pp. 9–20, 2009.

[75] L. Laaksonen, "A graphics program for the analysis and display of molecular dynamics trajectories," *J. Mol. Graph.*, vol. 10, pp. 33–34, 1992.

[76] D. L. Bergman, L. Laaksonen, and A. Laaksonen, "Visualization of solvation structures in liquid mixtures," *J. Mol. Graph. Model.*, vol. 15, pp. 301–306, 1997.

[77] M. F. Moody, "The shape of the T-even bacteriophage head," *Virology*, vol. 26, pp. 567–76, 1965.

[78] T. S. Baker, D. L. D. Caspar, and W. T. Murakami, "Polyoma virus 'hexamer' tubes consist of paired pentamers," *Nature*, vol. 303, pp. 446–448, 1983.

[79] T. S. Baker and D. L. D. Caspar, "Computer image modeling of pentamer packing in polyoma virus 'hexamer' tubes," *Ultramicroscopy*, vol. 13, pp. 137–152, 1984.

[80] R. Hull, G. J. Hills, and R. Markham, "Studies on alfalfa mosaic virus. II. The structure of the virus components," *Virology*, vol. 37, pp. 416–28, 1969.

[81] R. Hull, "Molecular biology of rice tungro viruses," *Annu. Rev. Phytopathol.*, vol. 34, pp. 275–97, 1996.

[82] R. Hull, B. E. Lockhart, D. V. R. Reddy, and J. E. Schoelz, "Genus badnavirus," in *Virus Taxonomy: Seventh Report of the International Committee on Taxonomy of Viruses* (M. H. V. van Regenmortel *et al.*, ed.), pp. 342–347, San Diego: Academic Press, 2000.

[83] N. A. Kiselev and A. Klug, "The structure of viruses of papilloma-polyoma type (V. tubular variants built of pentamers)," *J. Mol. Biol.*, vol. 40, pp. 155–171, 1969.

[84] R. Twarock, "Mathematical models for tubular structures in the family of Papovaviridae," *Bull. Math. Biol.*, vol. 67, pp. 973–987, 2005.

[85] T. Keef, A. Taormina, and R. Twarock, "Classification of capped tubular viral particles in the family of Papovaviridae," *J. Phys.: Cond. Matt.*, vol. 18, pp. S375–S387, 2006.

[86] T. Frischmuth, M. Ringel, and C. Kocher, "The size of encapsidated single-stranded DNA determines the multiplicity of african cassava mosaic virus particles," *J. Gen. Virol.*, vol. 82, pp. 673–676, 2001.

Part II

Mechanical Properties of Viral Capsids

Chapter 3

Theoretical Insights in the Elasticity of Viral Shells

3.1 Introduction

The main role of viral capsids is to protect the genetic material of the virus and to ensure its safe release into a new host. In particular, during the extracellular period between the released of the virion, from an infected cell, until a new host is found, the viral capsid must face a variety of physicochemical conditions [1], that can be even extreme in terms of temperature, pH, radiation or dehydration [2,3]. In addition some results suggest that virions are also subjected to substantial mechanical stress due, for instance, to osmotic shocks related to a sudden change of salt concentration in the environment, or to the packaging of the viral genome at high densities [4–7]. In fact, double stranded DNA (dsDNA) viruses have to withstand up to tens of atmospheres arising from the confined genetic material [7, 8]. Moreover, after the self-assembly of the capsid and before becoming infective, many viruses suffer a maturation process that properly tunes the mechanical properties of the viral shell. This maturation often changes the shape of the capsid through a process named "buckling transition" [1, 9–17]. In this process an initial spherical capsid with icosahedral symmetry undergoes a transformation into a polyhedral-like shell with more flat triangular faces, adopting usually the shape of an icosahedron.

Therefore, the mechanical properties of viral capsids play a very important role in the viability and functionality of a virus. The study of these properties has motivated an intense research in the last years and will be the main topic of this second part of the thesis.

Since different viruses have common designs and can be subjected to similar types of mechanical stresses, this suggests that some of the properties that confer stability to the capsids should be generic, and intimately related, for instance, to the shape or structure of the capsid, or the strength of the protein binding, rather than to the specific amino-acid sequence of the coat protein.

From a physical standpoint, viruses can be regarded essentially as nanoscopic thin shells. Thus, there has been a renewed interest in applying traditional concepts from elasticity theory [18] to understand some of the mechanical properties of viral capsids [19–32]. These studies have been specially promoted by the application of some experimental techniques customarily used in physics that have been able to extract precise mechanical properties. For instance, optical tweezers have allowed to measure the force of the packing motor of bacteriophages during the storage of dsDNA [33], and atomic force microscope (AFM) has been able to probe the mechanical resistance of single capsids [34–52].

Spherical capsids have been the main subject of these studies, and the elasticity models have been built usually based on two different approaches. In one case, capsids are described as continuum thin shells characterized by few elastic moduli [21–28]. These studies have provided interesting insights of several phenomena from buckling [21] to conformational transitions [24]. The other elasticity approach is based on the discretization of capsids in elastic networks. These models go from quasi-atomistic to more coarse-grained descriptions, and have facilitated the implementation of nanoindentation simulations (many times in correspondence with specific AFM experiments) [37–43, 47, 49–51], and also the study of viral capsids fluctuations [19, 20, 29–31], which might play an important role triggering the swelling of capsids for gene delivery [19], or the buckling transition during virus maturation [29]. These two strategies provide powerful techniques to characterize the mechanical properties of viral capsids. Nevertheless, the connection between the macroscopic magnitudes that characterize the structures, with the morphological constituents and their intermolecular interactions has not received so much attention [53–55].

In this chapter we will present some simple ideas to link the global measurements with the local properties of capsids, which should establish a basis to understand the microscopic origin of their elastic behavior. This description is an intermediate approach between the continuum elasticity and elastic network models.

3.2 Elastic properties of quasi-spherical capsids using a simple surface-covering model

In this section, we will use a simple surface-covering model to perform a qualitative comparison between the geometrical and elastic properties of spherical and icosahedral shells. The influence of the precise structure of the virus on these properties, and specially on the occurrence of a buckling transition will be analyzed in more detail in the next chapter.

3.2.1 The size of equilibrium structures

As we have previously discussed the assembly of capsids is driven by attractive interactions among coat proteins, which eventually lead to a specific final structure. To protect efficiently the genetic material, capsomers have to be close packed in the final shell, and this packing is limited by the steric effects of coat units, rather than by the attractive forces. This means that the size and shape of the structural subunits determine the amount of proteins that fit in a particular shell with a given shape and size, irrespective of the details of the cohesive forces, and of whether those are specific (e.g., disulfide bonds or covalent cross-linking) or non-specific (e.g., hydrophobic and electrostatic). Therefore, this paves the way to derive very general laws relating the number of coat proteins to the size of the resulting capsid for any given shape.

In particular, there is a simple way to estimate the optimal radius of a shell based on a surface-covering assumption. The idea is to assign an effective area to each capsomer and then to compute the radius of the structure whose area can accommodate all the capsomers, assuming full coverage. This procedure is particularized for spherical and for polyhedral capsids in the following subsections (see Fig. 3.1), although the same concepts can be applied to other type of structures.

Spherical capsids

Let's first derive a general expression for the radius of spherical capsids with icosahedral symmetry using this simple full-covering assumption (see Fig. 3.1). As we have seen in the first block of this thesis, T-shells are composed of $10(T-1)$ hexamers plus 12 pentamers. Thus, the total surface is given by $S_t = 10(T-1)A_H + 12A_P$, where A_H and A_P are the effective surface contribution of an hexamer and a pentamer, respectively. Since the surface of a spherical shell is $S_{sph} = 4\pi R^2$, by equating both areas we obtain the radius in terms of the

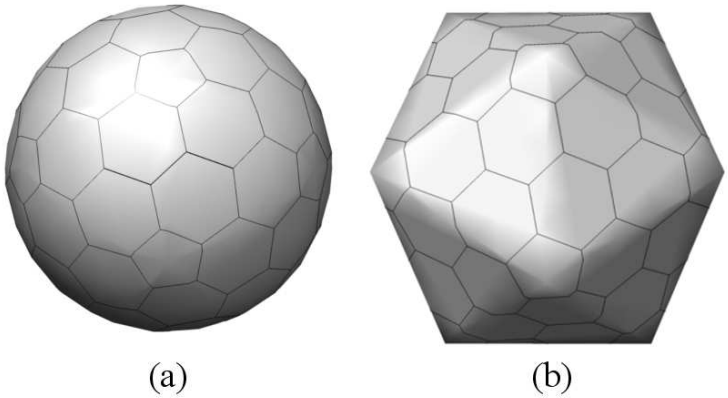


Figure 3.1: Spherical and icosahedral shapes. Two structures with icosahedral symmetry are shown: a spherical capsid (a) and an icosahedron (b). Both cases have a $T = 7l$ configuration and are made of 72 capsomers. Notice that the hexamers and pentamers are not exactly the same in the two morphologies. In particular, pentamers are almost flat in the spherical shape, but adopt instead a pyramidal shape in the icosahedron.

T-number

$$R_{th}^{sph}(T) = \sqrt{\frac{1}{4\pi}\left[10(T-1)A_H + 12A_P\right]} \tag{3.1}$$

This is a general expression that could be also expressed in terms of N just inverting the CK relation $T(N)$, derived in Chapter 1, Eq. (1.4).

Let's now particularize it to the simple physical model for capsomers used in the simulations of Chapter 2. In this context, it is important to stress that the effective area of the capsomers has to be chosen to cover the surface of the shell, rather than just using the physical projection of the actual spherical capsomer on the surface. Accordingly, and as shown in Fig. 3.2, an hexamer is represented by an hexagon of area

$$A_H = 2\sqrt{3}\,\sigma_0^2 \tag{3.2}$$

which circumscribes a capsomer of effective radius σ_0; similarly, a pentamer is represented by a pentagon with the same side as the hexagon, and an area

$$A_P = \frac{5}{3}\cot\left(\frac{\pi}{5}\right)\sigma_0^2 \tag{3.3}$$

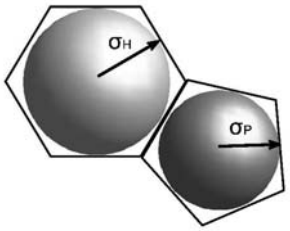


Figure 3.2: Effective area of a hexamer and a pentamer in a spherical capsid. The physical model introduced in Chapter 2 used spherical capsomers of different size for hexamers ($\sigma_H = \sigma_0$) and pentamers (σ_P), which were inscribed in a hexagon and pentagon of the same edge, respectively. The areas of these polygons correspond to the effective surface of the capsomers in the spherical capsid.

Then, the expression for the radius, Eq. (3.1), takes the form

$$R_{th}^{sph}(T) = \sqrt{\frac{5\sqrt{3}}{\pi}\left(T + \frac{1}{\sqrt{3}}\cot\left(\frac{\pi}{5}\right) - 1\right)}\ \sigma_0 \tag{3.4}$$

which is plotted in Fig. 3.3. Interestingly, this simple relation is in very good agreement with the radii of the equilibrium structures obtained from the simulations in Chapter 2.

In the limit of *big* shells the contribution of the 12 pentamers becomes negligible yielding

$$R_{th}^{sph}(T) = \sqrt{\frac{5\sqrt{3}}{\pi}T}\ \sigma_0 \approx 1.66\sqrt{T}\ \sigma_0 \tag{3.5}$$

which gives the expected scaling $R \sim T \sim \sqrt{N}$, and constitutes a much simpler relation. In Fig. 3.4 we plot the relative error of this approximation, where one can see that, except for very small T, the formula is almost exact. Indeed, for $T = 1$ the limit expression has an associated error around 10%, which decays very fast as $\sim 1/T$.

Polyhedral capsids

The same procedure can be applied to get the radius of a polyhedral capsid (see Fig. 3.1). The surface of the structure in terms of capsomers is $S_t = 10(T-1)A'_H + 12A'_P$, where A'_H and A'_P correspond to the effective area of the units for the polyhedral case. The total area of an icosahedron is $S_{ico} = \frac{40\sqrt{3}}{5+\sqrt{5}}R^2$,

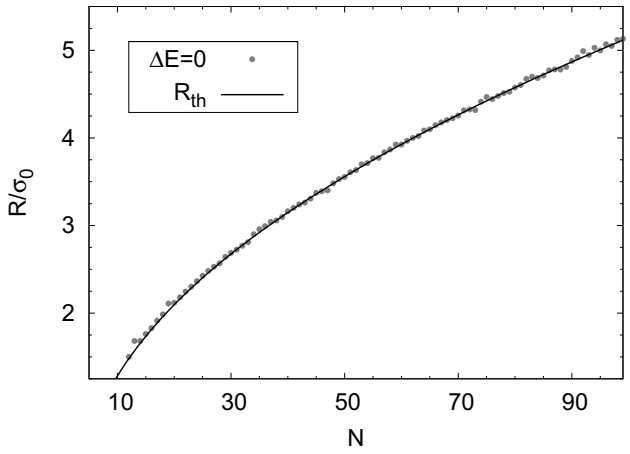


Figure 3.3: Theoretical and optimal radii for spherical capsids. The gray dots correspond to the radii obtained in the simulations of spherical capsids in the pentamer-hexamer scenario, i.e., $\Delta E = 0$ (see Chapter 2). The black curve plots the theoretical radii derived in the surface-covering model, Eq. (3.4). The agreement is excellent and confirms that the steric effects between coat units determine the size of the capsid (if the number of subunits is fixed).

where R is the radius of the polyhedron measured from the center to one vertex [56]. Equating both areas one obtains

$$R_{th}^{ico}(T) = \sqrt{\frac{5+\sqrt{5}}{40\sqrt{3}}\left[10(T-1)A_H + 12A_P\right]} \tag{3.6}$$

which is the general expression for polyhedral capsids.

Now, it is reasonable to assume an effective hexamer surface equivalent to the hexamer in spherical capsids[1],

$$A'_H = A_H. \tag{3.7}$$

However, for the pentamers in this polyhedral surface it is more convenient to assign them an effective area related specifically to the vertexes on the CK construction

$$A'_P = \frac{5}{6}A_H \tag{3.8}$$

[1]Indeed this choice is even more natural in the icosahedron, because the surface is recovered precisely by true hexagons, instead of "spherical" hexagons (see Fig. 3.1).

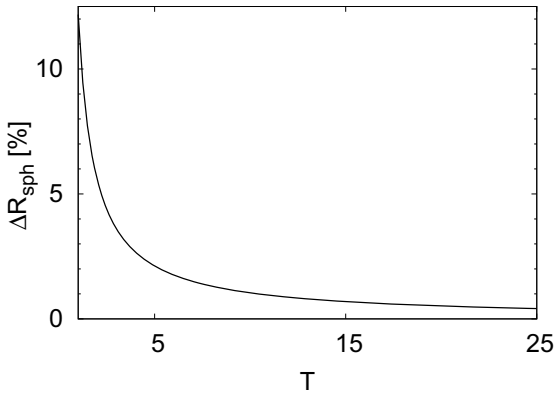


Figure 3.4: Relative error of the radius for spherical capsids between the exact expression, Eq. (3.4), and the approximation for big shells Eq. (3.5).

corresponding to the removal of one of the triangles of an hexamer (see Chapter 1). Note that this is precisely the area of a pentagonal pyramid at the vertex of the icosahedron (see Fig. 3.5). Then this leads to

$$R_{th}^{ico} = \sqrt{\frac{5+\sqrt{5}}{2}T}\,\sigma_0 \approx 1.90\sqrt{T}\,\sigma_0 \tag{3.9}$$

which is plotted in Fig. 3.5. In this case all capsids follow the strict scaling $R \sim T \sim \sqrt{N}$ for all N, because T is rigorously proportional to the surface of a triangular face of the polyhedral shell.

Let's define the ratio between the radii of the polyhedral and the spherical shapes as

$$\gamma_R = \frac{R_0^{ico}}{R_0^{sph}}. \tag{3.10}$$

For a fixed T, one gets that the radii of faceted shells are always slightly bigger than that of their spherical counterparts as shown in Fig. 3.5. In particular, in the limit of big shells, the radii of the polyhedral capsids tend to be almost 15% larger than that of the spherical shell, i.e.,

$$\gamma_R = \frac{R_{th}^{ico}}{R_{th}^{sph}} \approx 1.15 \tag{3.11}$$

as shown in Fig. 3.6.

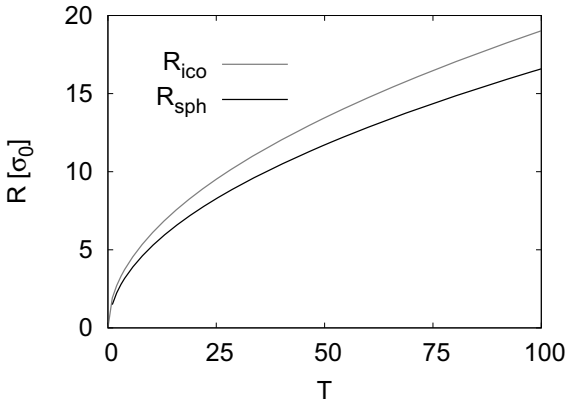


Figure 3.5: Theoretical radii for spherical, Eq. (3.4), and polyhedral shells, Eq. (3.9), as a function of the T-number.

It is interesting to compare the volumes of both shapes. In particular, the volume of the polyhedral shell is given by the classical formula of an icosahedron [56]

$$V_{th}^{ico} = \frac{80(3+\sqrt{5})}{3\left(\sqrt{10+2\sqrt{5}}\right)^3}(R_{th}^{ico})^3. \tag{3.12}$$

If we now compare it with the volume of a spherical capsid

$$V_{th}^{sph} = \frac{4\pi}{3}(R_{th}^{sph})^3 \tag{3.13}$$

the following ratio is obtained

$$\frac{V_{th}^{ico}}{V_{th}^{sph}} = \frac{20(3+\sqrt{5})}{\pi\left(\sqrt{10+2\sqrt{5}}\right)^3}\left(\frac{R_{th}^{ico}}{R_{th}^{sph}}\right)^3 \approx 0.61\left(\frac{R_{th}^{ico}}{R_{th}^{sph}}\right)^3 \tag{3.14}$$

This ratio is plotted in Fig. 3.7, where we observe that, although icosahedral radii are systematically bigger, $R_{th}^{ico} > R_{th}^{sph}$, the geometrical prefactor inverts the

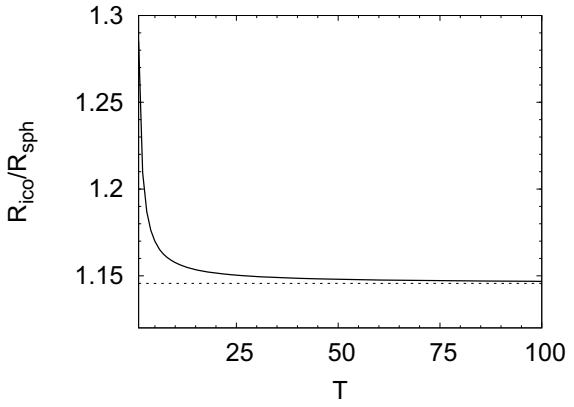


Figure 3.6: Ratio between the radii of the polyhedral, Eq. (3.9), and spherical capsids, Eq. (3.4). The dotted line is the limit for big shells, Eq. (3.11).

relation for the volumes so that $V_{th}^{ico} < V_{th}^{sph}$ (at least for $T \gtrsim 3$)[2]. In particular, in the limit of big shells, Eq. (3.11), one gets

$$\frac{V_{th}^{ico}}{V_{th}^{sph}} \approx 0.91 \tag{3.15}$$

So, except for very small shells, the volume of polyhedral capsids is almost 10% smaller compared to spherical shells of the same area. This might have important consequences in the mechanical properties of capsids, specially when they are pressurized due to the compaction of the viral genome, like for some dsDNA bacteriophages [8].

Note that in the comparison between spherical and icosahedral shells it was assumed that the structures had coat subunits with the same characteristic size σ_0. To compare structures with different subunit sizes, e.g., σ_0 and σ_0', it is necessary first to normalize their respective radii, i.e., R_0/σ_0 and R_0'/σ_0'. Then the results derived above will apply.

[2]Recall that the pentamers in the icosahedral shape have a larger effective area, i.e., $A_P' > A_P$ (see Fig. 3.1). For small capsids the total surface is dominated by the contribution of pentamers, and the area of the icosahedral shell becomes big enough to enclose a larger volume than the spherical counterpart. When the T-number increases the surface is dominated by the hexamers, which have the same effective area for both shapes, so the two capsids have essentially the same total effective area. At this point, since the sphere is the structure that maximizes the volume/surface ratio, it is natural to obtain $V_{sph} > V_{ico}$. Finally, notice that a $T = 3$ capsid is made by 12 pentamers and 20 hexamers, thus almost 2/3 of capsomers are hexamers, and it makes sense that the inversion of volumes takes place around $T \approx 3$.

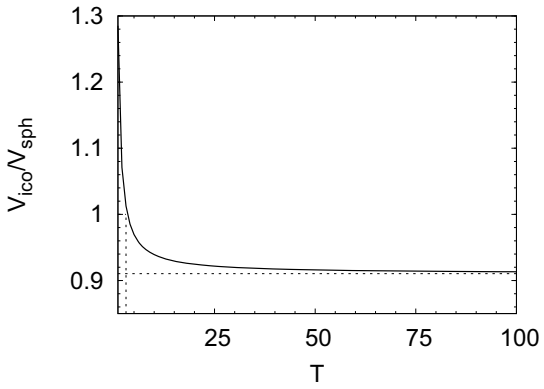


Figure 3.7: Theoretical prediction for the ratio of volumes of polyhedral and spherical capsids plotted following Eq. (3.14). The dotted line is the limiting value for big shells, Eq. (3.15). The intersection $V_{th}^{ico} = V_{th}^{th}$ takes place around $T \approx 3$, and it is indicated by a vertical dotted line.

3.2.2 Energy of capsids

Let's define the total energy at equilibrium of a capsid containing N capsomers as $E_e(N)$. This optimal energy is determined by the local interactions among coat proteins, and for each N corresponds to a minimum of the free energy versus radius (see Chapter 2). From a continuum elasticity perspective, this energy can be regarded also as the minimization of the elastic energy of the shell [22]. Interestingly, elastic constants of capsids can be computed in an indirect way by applying specific deformations to the capsids. Some of these deformations have also an intrinsic biological interest, because are involved in different steps of the virus cycle.

In particular, radial deformations are specially appealing because can be induced by several phenomena, like the internal pressure associated to the genetic material, variations of the pH, or osmotic shocks due to a sudden change of environmental conditions [4, 13, 57]. To understand the properties and to determine the potential advantages of the different viral shapes, we will connect the microscopic interactions with the macroscopic elastic description for this specific deformation.

The equilibrium energy $E_e(N)$ is a minimum, so for small radial deformations,

the total energy can be described by a parabolic expression[3] centered around $E_e(N)$, i.e.

$$E_N(\delta R) = E_e(N) + \frac{1}{2} a_E (\delta R)^2 \tag{3.17}$$

where $\delta R = R - R_0(N)$ is the deformation, $R_0(N)$ is the equilibrium radius, and a_E can be associated with the stiffness of the capsid. Our main goal here will be to connect this total energy with the microscopic interactions and deformations.

For small deformations, the local interaction between coat subunits can be described also as a quadratic minimum around the equilibrium position of the pair-interaction potential

$$u_{ij}(\delta r) = \varepsilon + \frac{1}{2} a_u (\delta r)^2 \tag{3.18}$$

where $\delta r = r - \sigma$ is the local deformation, σ is the equilibrium distance, ε is the binding energy at equilibrium, and a_u is the local stiffness of the bond. To connect the thermodynamic and the microscopic descriptions let's study a simple case. In particular, we assume a mean-field approach where all bonds interact on average with the same potential $u \approx u_{ij}$ associated to an effective interaction involving only the nearest neighbors with an average coordination number[4] z. Using this simplifying approximations, one can write

$$E_N(\delta R) = \frac{zN}{2} u_{ij}(\delta r) \tag{3.19}$$

Now we just need to relate the global and local deformations. For uniform deformations the relative global and local deformations are equal, yielding (see Appendix A)

$$\delta r = \frac{\sigma}{R_0}\, \delta R \tag{3.20}$$

This leads to a constitutive equation connecting the microscopic and macroscopic descriptions of capsids

$$E_N(\delta R) = \frac{zN}{2} (\varepsilon + \frac{1}{2} a_u \frac{\sigma^2}{R_0^2} (\delta R)^2) \tag{3.21}$$

[3] More precisely, the energy can be expressed as a Taylor expansion in terms of the deformation δR, i.e.

$$E(\delta R) = E_0 + \frac{\partial E}{\partial \delta R} \delta R + \frac{1}{2} \frac{\partial^2 E}{\partial \delta R^2} \delta R^2 + o(\delta R^3) \tag{3.16}$$

However, since the expansion is around a minimum, the first derivative is zero, and in a first approximation the deformation energy is a quadratic function.

[4] In a flat hexagonal sheet z would be six, but in icosahedral shells one has to take into account the influence of pentons, specially for small shells. For instance, a $T = 1$ has strictly a coordination $z_{(T=1)} = 5$.

where the minimum energy is given by

$$E_e(N) = \frac{zN}{2}\varepsilon \tag{3.22}$$

and the rigidity of the capsid is

$$a_E = \frac{zN}{2}\frac{\sigma^2}{R_0^2}\,a_u \tag{3.23}$$

Since $\sigma^2 N \sim R_0^2$, a_E is roughly a size-independent property of the capsid in this mean-field approximation (except for very small spherical structures).

Interestingly, it is now possible to show that the shape of capsids actually has an important influence in the response to radial deformations. More specifically, if one compares a spherical and a polyhedral capsid with the same number of capsomers N, and whose local interactions are similar, namely $a_u^{ico} \approx a_u^{sph}$, the ratio between their rigidities is just given by

$$\gamma_E = \frac{a_E^{ico}}{a_E^{sph}} = \left(\frac{R_0^{sph}}{R_0^{ico}}\right)^2 = \left(\frac{1}{\gamma_R}\right)^2 \tag{3.24}$$

Remarkably, since $R_0^{ico} > R_0^{sph}$, this ratio will be always smaller than one, $\gamma_E < 1$. This means that, subjected to similar radial deformations, the polyhedral capsid will accumulate less deformation energy. In particular, since for big shells $\gamma_R \approx 1.15$, a faceted shell will accumulate around a 30% less energy (although in terms of relative deformations $\delta R/R_0$, both shapes accumulate the same energy).

3.2.3 Pressure

The deformation energy can be related, for instance, to an effective pressure difference p between the exterior and the interior of the capsid. The classical thermodynamic definition of pressure is

$$p = -\frac{\partial E}{\partial V} \tag{3.25}$$

fixing the rest of thermodynamic magnitudes that describe the system, like the number of subunits N, the temperature T, or the solvent properties. For an empty capsid at equilibrium the pressure has to be zero, because the energy is a minimum

$$p_0 = -\left(\frac{\partial E}{\partial V}\right)_{R_0} = 0 \tag{3.26}$$

As mentioned before, the confinement of the genetic material and osmotic effects can induce non-zero pressure differences $p \neq 0$.

It is more convenient to express the pressure difference in terms of radial rather than volume changes. For both spherical and polyhedral shells the volumes are proportional to the cubic power of the radius, Eqs (3.13) and (3.12), which can be expressed in general as

$$V = a_V R^3 \tag{3.27}$$

where a_V is a geometrical constant. Now, it is easy to see that systems where $V \sim R^3$ will obey

$$\frac{\partial V}{\partial R} = 3\frac{V}{R} \tag{3.28}$$

so the pressure equation can be rewritten as

$$p = -\frac{R}{3V}\frac{\partial E}{\partial R} \tag{3.29}$$

In this way we obtain a direct formula for the pressure difference in terms of the radial deformations of the capsid.

Using the previous harmonic expressions connecting the macroscopic and microscopic deformations of the system, we can now reexpress the pressure in terms of the shape and the local interactions. In particular, the pressure associated to a deformation δR is

$$p = -\frac{R}{3V} a_E\, \delta R = -\frac{zNR}{6V}\frac{\sigma^2}{R_0^2}\, a_u\, \delta R \tag{3.30}$$

where in the second equality the explicit microscopic dependence has been used. It is useful to express this pressure in terms of the relative deformation $\delta R/R_0$ as

$$p = -a_p \frac{\delta R/R_0}{(1+\delta R/R_0)^2} \tag{3.31}$$

where

$$a_p = \frac{zN}{6V_0}\,\sigma^2 a_u \tag{3.32}$$

is a pressure constant that contains all the microscopic and geometrical details. Fig. 3.8 plots the reduced pressure p/a_p which has a universal dependence on the relative deformation that should hold for all quasi-spherical viruses, at least for *small* deformations where the quadratic approximation of the deformation energy is valid. Evidently, the actual pressure value for a particular virus is determined by its specific properties through a_p. This suggests that, by relating pressure

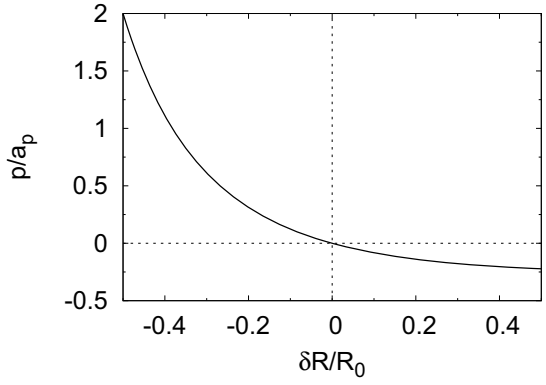


Figure 3.8: Pressure and radial deformation of a capsid. The reduced pressure p/a_p is plotted as a function of the relative capsid deformation $\delta R/R_0$, Eq. (3.31). Inwards pressure is positive, and outwards pressure negative.

differences with capsid deformations, one could obtain valuable information about the local interactions between coat subunits. It is worth to mention that there is a clear asymmetry depending on the direction of the deformation. Higher pressures are required to compress the shell than to expand it. Nevertheless, for very small deformations $\delta R/R_0 \ll 1$ the pressure reduces to $p \approx -a_p\, \delta R/R_0$, thus recovering a symmetric and linear behavior around equilibrium.

The maximum pressure that a capsid can resist is related to the maximum radial deformation that it can tolerate. In the linear regime, since $N \sim R_0^2$ and $V_0 \sim R_0^3$, this maximum pressure decays as[5]

$$p^\star \sim \frac{1}{R_0} \tag{3.33}$$

Therefore, for the same intermolecular interactions big capsids are in principle less efficient storing pressure.

Let's now compare the tolerance to pressure of polyhedral and spherical shells. As before, we consider capsids with the same number of capsomers N and the same local interactions. Then the ratio of pressures for the same relative defor-

[5]If the maximum pressure is set at the same relative deformation $\delta R^\star/R_0$ for different capsid sizes, the associated maximum pressures will also decay as $p^\star \sim 1/R_0$, even outside the linear regime.

mation is

$$\frac{p^{ico}}{p^{sph}} = \frac{a_p^{ico}}{a_p^{sph}} = \frac{V_0^{sph}}{V_0^{ico}} \tag{3.34}$$

Therefore, since $V_0^{sph} > V_0^{ico}$, except for very small shells, polyhedral capsids build up larger pressures when subjected to the same relative radial deformations. In particular, in the limit of big capsids, where $V_0^{ico}/V_0^{sph} \approx 0.91$, faceted capsids will accumulate around a 10% more pressure than the spherical counterparts. That might constitute an advantage, for instance, for viruses that rely on pressure driven translocations to initiate the infection.

3.2.4 Young modulus and effective spring constant

From classical elasticity theory, it is possible to relate also the internal pressure of a thin spherical shell with the radial deformation of the capsid [18]

$$p = -\frac{2Y}{1-\nu}\frac{h}{R_0}\frac{\delta R}{R_0} \tag{3.35}$$

where Y is the 3D Young modulus, ν is the Poisson ratio[6], and h is the effective thickness of the capsid. Comparing this expression with the linear approximation of Eq. (3.31), one gets

$$Y = \frac{z}{12}\frac{\sigma^2 N R_0}{V_0}(1-\nu)\,\frac{a_u}{h} \tag{3.36}$$

which provides a microscopic interpretation of the Young modulus. In solids the Young modulus is an intrinsic property of the material. However for thin shells we observe that, since $\sigma^2 N R_0 \sim V_0$, the effective Young modulus depends on the shell thickness h as $Y \sim 1/h$. Then to properly characterize viral capsids from an elasticity point of view it is more convenient to define a 2D Young modulus as follows [18]

$$Y^\star = Yh \tag{3.37}$$

For big capsids this elastic modulus will depend essentially on the microscopic interactions, although the changes in the effective coordination number and a_V for small shells may lead to some deviations, which have been also observed by nanoindentation simulations [37].

Interestingly, the Young modulus is also related to the elastic constant k measured in AFM experiments by [18]

$$k \approx \frac{Y\,h^2}{R_0} = Y^\star\frac{h}{R_0} \tag{3.38}$$

[6]The Poisson ratio is typically 1/3 for protein-like materials.

Using now Eq. (3.36) a microscopic interpretation for the elastic constant can be obtained

$$k \approx \frac{z}{12}\frac{N}{V_0}(1-\nu)\,\sigma^2 a_u\, h. \tag{3.39}$$

Reexpressing k in terms of the microscopic properties leads to a scaling

$$k \sim \frac{h}{R_0} a_u \tag{3.40}$$

Therefore, if one compares capsids of similar subunit interactions and shell thickness, the bigger the structure, the softer will be the elastic response. This scaling becomes of course invalid for large R_0, where k tends to a constant independent on the curvature [18, 37].

3.2.5 Bulk modulus

The bulk modulus is an elastic magnitude defined as the inverse of the compressibility of the capsid

$$K = -V\left(\frac{\partial p}{\partial V}\right) \tag{3.41}$$

which for quasi-spherical shells can be expressed in terms of the radius as

$$K = \frac{R^2}{9V}\left(\frac{\partial^2 E}{\partial R^2}\right) \tag{3.42}$$

Now, using Eq. (3.31), the bulk modulus can be rewritten in terms of the relative deformation as

$$K = \frac{1}{3} a_p \frac{1-\delta R/R_0}{(1+\delta R/R_0)^2}. \tag{3.43}$$

Fig. 3.9 plots the reduced bulk modulus K/a_p as a function of the relative radial deformation, which is a universal function, qualitatively similar to the pressure. Again the asymmetry in the elastic response when comparing compression and expansion originates by geometrical reasons, and does not require an intrinsic asymmetry in the local interactions.

At equilibrium, the value of the bulk modulus becomes

$$K_0 = \frac{1}{3} a_p \tag{3.44}$$

that similarly to the pressure decays as

$$K_0 \sim \frac{1}{R_0} \tag{3.45}$$

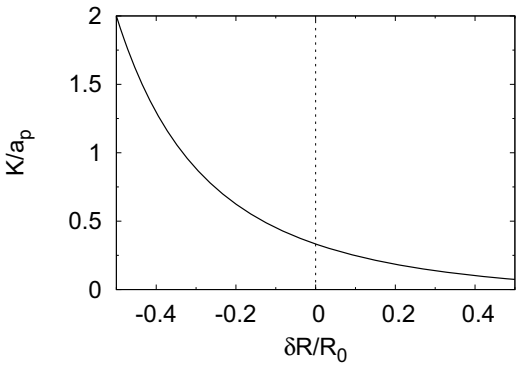


Figure 3.9: Bulk modulus. The reduced bulk modulus K/a_p is plotted as a function of the relative capsid deformation $\delta R/R_0$, Eq. (3.43).

since $a_p \sim 1/R_0$. Therefore, for identical interactions, big shells are expected to be more compressible.

Finally, when the bulk modulus of polyhedral and spherical shells are compared, the same relation as for the pressures is obtained

$$\frac{K^{ico}}{K^{sph}} = \frac{V_0^{sph}}{V_0^{ico}} \tag{3.46}$$

So, except for very small shells, polyhedral capsids are expected to be more incompressible than spherical ones.

3.2.6 Buckling transition

The fact that $R_0^{sph} < R_0^{ico}$ and the nearly parabolic dependence of the energies of capsids suggest that, upon expansion, there must be an intermediate radius R_c after which the energy of a compressed polyhedral shell becomes smaller than that of an expanded spherical shell (see Fig. 3.10). Therefore, the expansion of stable spherical shells should lead in general to a buckling transition where capsids will end up faceted and with polyhedral shape.

To estimate the transition radius, let's assume that the energy difference between both optimal energies is small compared to the energy barrier defined by the cut (see Fig. 3.10). Then the condition for the cut radius reads

$$\frac{1}{2}a_E^{sph}\,(R_c - R_0^{sph})^2 \approx \frac{1}{2}a_E^{ico}\,(R_c - R_0^{ico})^2 \tag{3.47}$$

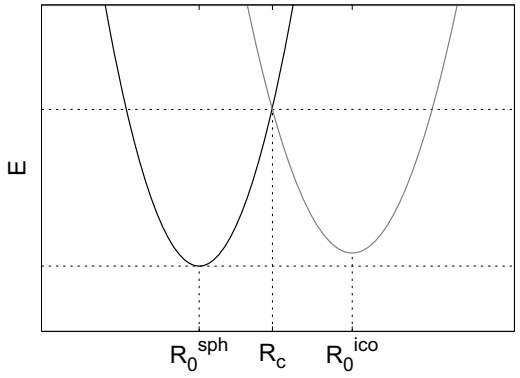


Figure 3.10: Schematic representation of the energies for a spherical and a polyhedral shell. The transition radius R_c is defined by the cut of both branches. The difference of optimal energies is assumed to be negligible, so the energy at the transition is dominated by deformation energies.

and from it one gets

$$R_c = \frac{1 + \gamma_R \sqrt{\gamma_E}}{1 + \sqrt{\gamma_E}} R_0^{sph} \tag{3.48}$$

In the case where the local interactions are similar, the ratio of both energy curvatures is only determined by the size of shells, Eq. (3.24), thus obtaining

$$R_c = \frac{2\gamma_R}{1 + \gamma_R} R_0^{sph} \approx 1.068\, R_0^{sph} \tag{3.49}$$

Therefore, if the local interactions are essentially the same in both capsids, the transition should take place around a 7% expansion in radius of the spherical shell.

3.3 Elasticity of a 2D lattice of Lennard-Jones particles

Let's particularize some of the results derived above for the specific case of a Lennard-Jones potential. The interest in this model comes from the fact that both the results of the previous chapter and the study of the mechanical properties of the next chapter are based on this type of interaction between capsomers.

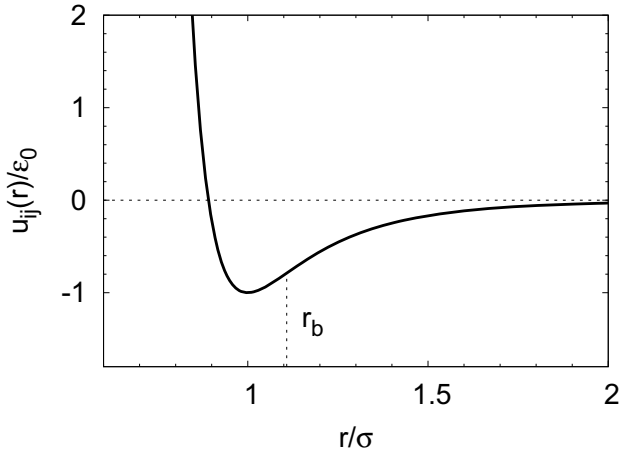


Figure 3.11: Lennard-Jones potential, Eq. (3.50). Intermolecular interaction u_{ij} versus the distance between particles, r. The units are normalized by the binding energy ε_0, and the equilibrium distance σ. The vertical line indicates the inflection point of the interaction r_b, Eq. (3.54), which is associated to the maximum bonding force.

For simplicity we are going to study some basic properties of a lattice of such particles, which will provide useful insights for the next chapter.

The Lennard-Jones potential used along the thesis is given by

$$u_{ij}(r) = \varepsilon_0 \left[\left(\frac{\sigma}{r}\right)^{12} - 2\left(\frac{\sigma}{r}\right)^{6} \right] \tag{3.50}$$

where σ is the equilibrium distance, ε_0 is the binding energy, and r is the distance between particles (see Fig. 3.11).

The main microscopic parameter in the mean-field theory introduced above corresponds to the interaction strength, a_u. This can be obtained for the Lennard-Jones by doing a Taylor expansion of Eq. (3.50) up two second order around the equilibrium position $r = \sigma$. So one gets

$$u_{ij}(r) = -\varepsilon_0 + \frac{1}{2}\,\frac{72\,\varepsilon_0}{\sigma^2}(\delta r)^2 + o(\delta r^3) \tag{3.51}$$

and the interaction strength is obtained by comparison with Eq. (3.18) as

$$a_u^{LJ} = 72\,\frac{\varepsilon_0}{\sigma^2} \tag{3.52}$$

With this value one can estimate the rest of intrinsic mechanical coefficients of a structure, i.e., the global energy strength a_E, the pressure scale a_p, and the 2D Young modulus $Y^\star$, respectively, Eqs (3.23), (3.32), and (3.37).

It is worth to compute the maximum force supported by a LJ interaction. The force between two particles is given by the derivative of the potential, $f_{ij} = -du_{ij}/dr$, and the maximum takes place when $df_{ij}/dr = -d^2u_{ij}/dr^2 = 0$, i.e., at the inflection point of the LJ potential (see Fig. 3.11). Thus, the value of the maximum force is

$$f_b = 12\frac{\varepsilon_0}{\sigma}\left[\left(\frac{7}{13}\right)^{13/6} - \left(\frac{7}{13}\right)^{7/6}\right] \approx -2.69\frac{\varepsilon_0}{\sigma} \tag{3.53}$$

that corresponds to an intermolecular distance

$$r_b = \left(\frac{13}{7}\right)^{1/6}\sigma \approx 1.11\,\sigma \tag{3.54}$$

Note that the cohesion force will drop for larger intermolecular separations, $r > r_b$. Thus if we apply an increasing tensile stress to a hexagonal network, when the local deformation reaches $r \approx r_b$, some local bonds will tend to disrupt and promote the formation of cracks in the solid lattice [58, 59].

In analogy for capsids, the maximum local deformation r_b can be associated to the bursting of the shell, and sets a maximum pressure resisted by a quasi-spherical structure, i.e., $p_b \approx p(r_b)$.

3.4 Conclusions

Based on simple theoretical insights, we have been able to connect the microscopic properties of capsids with their global (or macroscopic) response, showing also the relevance of the capsid shape in determining its mechanical properties.

In particular, we have computed the radii of spherical and icosahedral shapes, based on the assumption that the steric effects among coat proteins determine the packing of the shell. The results showed that polyhedral capsids tend to be $\sim 15\%$ larger in radius than their spherical counterparts. On the contrary, except for very small structures, the volume of a spherical capsid is almost $\sim 10\%$ bigger than that of the equivalent icosahedral shape.

A mean-field approach has allowed to characterize the total energy of capsids in terms of the intermolecular interactions among the coat subunits. In this way we have been able to compute the main mechanical properties of the

structures under radial deformations, like the internal pressure of the shell or its compressibility. In particular we have found a universal pressure behavior for quasi-spherical capsids as a function of the relative radial deformation. In all cases shells are harder to compress than to expand. Moreover, larger shells are generally less efficient in storing pressure than smaller ones, and, for the same microscopic conditions and relative deformations, icosahedral capsids are able to accumulate up to a $\sim 10\%$ more pressure than spherical shells, and are $\sim 10\%$ more incompressible than the spherical counterparts.

For spherical shells, it has also been possible to connect its elasticity with the underlying microscopic properties. In particular, a constitutive relation for the Young modulus was derived. This opens the door to connect the elastic constants typically measured in AFM experiments with the microscopic properties of the shell.

Finally, spherical capsids under expansion can show a buckling transition to polyhedral shaped shells, and we provided an estimation of the transition radius.

The next chapter will show, using the model introduced in Chapter 2, the validity of the predictions derived here. In addition, the possible biological implications of these results, in particular, with regard to viruses that suffer such buckling transitions during the maturation process, will be discussed.

Appendices

A Local and global deformations

If a spherical capsid is subjected to a uniform radial deformation, one can easily relate the intermolecular and global deformations (see Fig. A.1). The intermolecular distances are given by $\sigma = 2R_0 sin(\theta/2)$ and $r = 2Rsin(\theta/2)$, respectively, before and after the radial deformation. By dividing both expressions one gets

$$\frac{r}{\sigma} = \frac{R}{R_0} \tag{A-1}$$

or in terms of the deformations

$$\frac{\delta r}{\sigma} = \frac{\delta R}{R_0} \tag{A-2}$$

where $\delta r = r - \sigma$ and $\delta R = R - R_0$. Therefore, for a uniform radial deformation of a spherical capsid, the relative local and global deformations are equal.

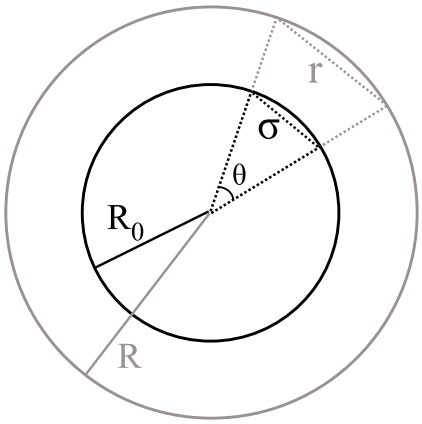


Figure A.1: Global and local deformations in a spherical capsid. The circumferences illustrates schematically a slice of a spherical capsid for two different radii. The initial state, characterized by a radius R_0 and a local distance between subunits σ, is shown in black. A uniform radial deformation of the initial sphere is represented in gray, where the radius is R and the new local distance is r. Since the deformation is uniform, in both cases the subtended angle between two interacting subunits is θ.

Notice that for a uniform expansion, the relation between the global and local deformations for a polyhedral capsid will be given also by Eq. (A-2).

References

[1] M. F. Moody, "Geometry of phage head construction," *J. Mol. Biol.*, vol. 293, pp. 401–33, 1999.

[2] G. Rice, K. Stedman, J. Snyder, B. Wiedenheft, D. Willits, S. Brumfield, T. McDermott, and M. J. Young, "Viruses from extreme thermal environments," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, pp. 13341–5, 2001.

[3] M. Prigent, M. Leroy, F. Confalonieri, M. Dutertre, and M. S. DuBow, "A diversity of bacteriophage forms and genomes can be isolated from the surface sands of the Sahara Desert," *Extremophiles*, vol. 9, pp. 289–296, 2005.

[4] A. Cordova, M. Deserno, W. M. Gelbart, and A. Ben-Shaul, "Osmotic shock and the strength of viral capsids," *Biophys. J.*, vol. 85, pp. 70–4, 2003.

[5] J. Kindt, S. Tzlil, A. Ben-Shaul, and W. M. Gelbart, "DNA packaging and ejection forces in bacteriophage," *Proc. Natl. Acad. Sci. USA.*, vol. 98, pp. 13671–13674, 2001.

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

[7] A. Evilevitch, L. Lavelle, C. M. Knobler, E. Raspaud, and W. M. Gelbart, "Osmotic pressure inhibition of DNA ejection from phage," *Proc. Natl. Acad. Sci. USA*, vol. 100, pp. 9292–9295, 2003.

[8] W. M. Gelbart and C. M. Knobler, "Pressurized viruses," *Science*, vol. 323, pp. 1682–1683, 2009.

[9] W. Jiang, Z. Li, Z. Zhang, M. L. Baker, P. E. Prevelige, and W. Chiu, "Coat protein fold and maturation transition of bacteriophage P22 seen at subnanometer resolutions," *Nat. Struct. Biol.*, vol. 10, pp. 131–5, 2003.

[10] K. N. Parent, R. Khayat, L. H. Tu, M. M. Suhanovsky, J. R. Cortines, C. M. Teschke, J. E. Johnson, and T. S. Baker, "P22 coat protein structures reveal a novel mechanism for capsid maturation: stability without auxiliary proteins or chemical crosslinks," *Structure*, vol. 18, pp. 390–401, 2010.

[11] A. Ionel, J. A. Velázquez-Muriel, D. Luque, A. Cuervo, J. R. Castón, J. M. Valpuesta, J. Martín-Benito, and J. L. Carrascosa, "Molecular rearrangements involved in the capsid shell maturation of bacteriophage T7," *J. Biol. Chem.*, vol. 286, pp. 234–42, 2011.

[12] X. Agirrezabala, J. Martín-Benito, J. R. Castón, R. Miranda, J. M. Valpuesta, and J. L. Carrascosa, "Maturation of phage T7 involves structural modification of both shell and inner core components," *EMBO J.*, vol. 24, pp. 3820–9, 2005.

[13] A. C. Steven, J. B. Heymann, N. Cheng, B. L. Trus, and J. F. Conway, "Virus maturation: dynamics and mechanism of a stabilizing structural transition that leads to infectivity," *Curr. Opin. Struc. Biol.*, vol. 15, pp. 227–36, 2005.

[14] I. Gertsman, L. Gan, M. Guttman, K. Lee, J. A. Speir, R. L. Duda, R. W. Hendrix, E. A. Komives, and J. E. Johnson, "An unexpected twist in viral capsid maturation," *Nature*, vol. 458, pp. 646–50, 2009.

[15] J. E. Johnson, "Virus particle maturation: insights into elegantly programmed nanomachines," *Curr. Opin. Struc. Biol.*, vol. 20, pp. 210–216, 2010.

[16] B. Bothner, D. Taylor, B. Jun, K. K. Lee, G. Siuzdak, C. P. Schlutz, and J. E. Johnson, "Maturation of a tetravirus capsid alters the dynamic properties and creates a metastable complex," *Virology*, vol. 334, pp. 17–27, 2005.

[17] J. Tang, K. K. Lee, B. Bothner, T. S. Baker, M. Yeager, and J. E. Johnson, "Dynamics and stability in maturation of a T=4 virus," *J. Mol. Biol.*, vol. 392, pp. 803–12, 2009.

[18] L. D. Landau and I. M. Lifshitz., *Theory of elasticity.* Pergamon, New York, 1975.

[19] F. Tama, , and C. L. Brooks, "The mechanism and pathway of pH induced swelling in cowpea chlorotic mottle virus," *J. Mol. Biol.*, vol. 318, pp. 733–747, 2002.

[20] F. Tama and C. L. Brooks, "Diversity and identity of mechanical properties of icosahedral viral capsids studied with elastic network normal mode analysis," pp. 299–314, 2005.

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

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

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

[24] T. Guérin and R. Bruinsma, "Theory of conformational transitions of viral shells," *Phys. Rev. E*, vol. 76, p. 061911, 2007.

[25] A. Y. Morozov and R. F. Bruinsma, "Assembly of viral capsids, buckling, and the Asaro-Grinfeld-Tiller instability," *Phys. Rev. E*, vol. 81, p. 041925, 2010.

[26] A. Šiber, "Buckling transition in icosahedral shells subjected to volume conservation constraint and pressure: Relations to virus maturation," *Phys. Rev. E*, vol. 73, p. 061915, 2006.

[27] A. Šiber and R. Podgornik, "Stability of elastic icosadeltahedral shells under uniform external pressure: Application to viruses under osmotic pressure," *Phys. Rev. E*, vol. 79, p. 011919, 2009.

[28] C. Q. Ru, "Buckling of empty spherical viruses under external pressure," *J. Appl. Phys.*, vol. 105, p. 124701, 2009.

[29] M. Widom, J. Lidmar, and D. Nelson, "Soft modes near the buckling transition of icosahedral shells," *Phys. Rev. E*, vol. 76, p. 031911, 2007.

[30] Z. Yang, I. Bahar, and M. Widom, "Vibrational dynamics of icosahedrally symmetric biomolecular assemblies compared with predictions based on continuum elasticity," *Biophys. J.*, vol. 96, pp. 4438–48, 2009.

[31] B. Stephanidis, S. Adichtchev, P. Gouet, a. McPherson, and a. Mermet, "Elastic properties of viruses," *Biophys. J.*, vol. 93, pp. 1354–9, 2007.

[32] M. Zink and H. Grubmuller, "Mechanical properties of the icosahedral shell of southern bean mosaic virus: A molecular dynamics study," *Biophys. J.*, vol. 96, pp. 1350–1363, 2009.

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

[34] I. L. Ivanovska, P. J. de Pablo, B. Ibarra, G. Sgalari, F. C. MacKintosh, J. L. Carrascosa, C. F. Schmidt, and G. J. L. Wuite, "Bacteriophage capsids: tough nanoshells with complex elastic properties," *Proc. Natl. Acad. Sci. USA.*, vol. 101, pp. 7600–7605, 2004.

[35] C. Carrasco, A. Carreira, I. A. T. Schaap, P. A. Serena, J. G.-H. M. G., Mateu, and P. J. de Pablo, "DNA-mediated anisotropic mechanical reinforcement of a virus," *Proc. Natl. Acad. Sci. USA.*, vol. 103, pp. 13706–11, 2006.

[36] C. Carrasco, M. Castellanos, P. J. de Pablo, and M. G. Mateu, "Manipulation of the mechanical properties of a virus by protein engineering," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, pp. 4150–5, 2008.

[37] G. A. Vliegenthart and G. Gompper, "Mechanical deformation of spherical viruses with icosahedral symmetry," *Biophys. J.*, vol. 91, pp. 834–41, 2006.

[38] W. Klug, R. Bruinsma, J.-P. Michel, C. Knobler, I. Ivanovska, C. Schmidt, and G. Wuite, "Failure of viral shells," *Phys. Rev. Lett.*, vol. 97, p. 228101, 2006.

[39] M. Gibbons and W. Klug, "Nonlinear finite-element analysis of nanoindentation of viral capsids," *Phys. Rev. E*, vol. 75, p. 031901, 2007.

[40] M. M. Gibbons and W. S. Klug, "Mechanical modeling of viral capsids," *J. Mater. Sci.*, vol. 42, pp. 8995–9004, 2007.

[41] M. M. Gibbons and W. S. Klug, "Influence of nonuniform geometry on nanoindentation of viral capsids," *Biophys. J.*, vol. 95, pp. 3640–9, 2008.

[42] M. Buenemann and P. Lenz, "Mechanical limits of viral capsids," *Proc. Natl. Acad. Sci. USA.*, vol. 104, pp. 9925–9930, 2007.

[43] M. Buenemann and P. Lenz, "Elastic properties and mechanical stability of chiral and filled viral capsids," *Phys. Rev. E*, vol. 78, p. 051924, 2008.

[44] J. P. Michel, I. L. Ivanovska, M. M. Gibbons, W. S. Klug, C. M. Knobler, G. J. L. Wuite, and C. F. Schmidt, "Nanoindentation studies of full and empty viral capsids and the effects of capsid protein mutations on elasticity and strength," *Proc. Natl. Acad. Sci. USA.*, vol. 103, pp. 6184–6189, 2006.

[45] C. Uetrecht, C. Versluis, N. R. Watts, W. H. Roos, G. J. L. Wuite, P. T. Wingfield, A. C. Steven, and A. J. R. Heck, "High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, pp. 9216–20, 2008.

[46] W. H. Roos, K. Radtke, E. Kniesmeijer, H. Geertsema, B. Sodeik, and G. J. L. Wuite, "Scaffold expulsion and genome packaging trigger stabilization of herpes simplex virus capsids," *Proc. Natl. Acad. Sci. USA.*, vol. 106, pp. 9673–9678, 2009.

[47] W. H. Roos, M. M. Gibbons, A. Arkhipov, C. Uetrecht, N. R. Watts, P. T. Wingfield, A. C. Steven, A. J. R. Heck, K. Schulten, W. S. Klug, and G. J. L. Wuite, "Squeezing protein shells: how continuum elastic models, molecular dynamics simulations, and experiments coalesce at the nanoscale," *Biophys. J.*, vol. 99, pp. 1175–1181, 2010.

[48] M. Baclayon, G. J. L. Wuite, and W. H. Roos, "Imaging and manipulation of single viruses by atomic force microscopy," *Soft Matter*, vol. 6, p. 5273, 2010.

[49] W. H. Roos, R. Bruinsma, and G. J. L. Wuite, "Physical virology," *Nature Phys.*, vol. 6, pp. 733–743, 2010.

[50] A. Ahadi, J. Colomo, and A. Evilevitch, "Three-dimensional simulation of nanoindentation response of viral capsids. Shape and size effects," *J. Phys. Chem.. B*, vol. 113, pp. 3370–8, 2009.

[51] Y. Zhao, Z. Ge, and J. Fang, "Elastic modulus of viral nanotubes," *Phys. Rev. E*, vol. 78, p. 031914, 2008.

[52] N. Kol, M. Gladnikoff, D. Barlam, R. Z. Shneck, A. Rein, and I. Rousso, "Mechanical properties of murine leukemia virus particles: effect of maturation," *Biophys. J.*, vol. 91, pp. 767–774, 2006.

[53] R. Zandi and D. Reguera, "Mechanical properties of viral capsids," *Phys. Rev. E*, vol. 72, p. 021917, 2005.

[54] A. Arkhipov, W. H. Roos, G. J. L. Wuite, and K. Schulten, "Elucidating the mechanism behind irreversible deformation of viral capsids," *Biophys. J.*, vol. 97, pp. 2061–9, 2009.

[55] V. Krishna, G. S. Ayton, and G. A. Voth, "Role of protein interactions in defining HIV-1 viral capsid shape and stability: a coarse-grained analysis," *Biophys. J.*, vol. 98, pp. 18–26, 2010.

[56] H. S. M. Coxeter, *Introduction to geometry.* Wiley, 2nd ed., 1989.

[57] J. M. Fox, G. W. 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*, vol. 244, pp. 212–218, 1998.

[58] R. L. B. Selinger, Z.-G. Wang, and W. M. Gelbart, "Effect of temperature and small-scale defects on the strength of solids," *J. Chem. Phys.*, vol. 95, p. 9128, 1991.

[59] R. L. B. Selinger, R. M. Lynden-Bell, and W. M. Gelbart, "Stress-induced failure and melting of ideal solids," *J. Chem. Phys.*, vol. 98, p. 9808, 1993.

Chapter 4

Buckling and Maturation of Spherical Viruses

4.1 Introduction

As mentioned in the previous chapter, after the self-assembly of the capsid and before becoming infective, many viruses suffer a maturation process that properly tunes the properties of the viral shell and which often involves a buckling transition [1–10]. In this transition, the initial spherical capsid with icosahedral symmetry undergoes a transformation into a polyhedral shell with flat triangular faces and the shape of an icosahedron (see Fig. 4.1).

The occurrence of the icosahedral faceting during virus maturation can be associated to different biological pathways. For instance, bacteriophage T7 assembles a procapsid helped by scaffolding proteins, and then suffers an expansion and a buckling transition during the packaging of the viral dsDNA [4,5]. Instead, bacteriophage HK97 suffers a cleavage in the coat protein of the prohead and the buckling transition can be triggered *in vitro* by a change in the pH [6–8]. Once triggered, the buckling transition seems to be related to an energy cascade of steps that lowers the energy of the capsid [7, 8]. The spontaneous nature of the transition combined with the fact that it can be activated by different signals suggests that common features regarding the structure or interactions of the coat proteins could play an important role determining the viability of buckling, and the potential mechanical advantages of the final capsid.

The occurrence of this buckling transition has been addressed in physical terms by using continuum elasticity theory [11–21]. In this context a viral capsid is considered a thin elastic shell with two competing energies: the bending

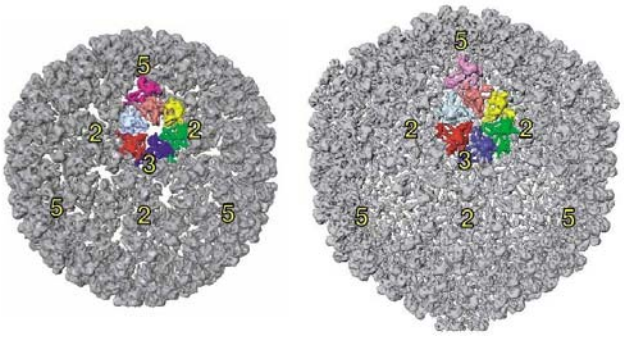


Figure 4.1: Buckling during virus maturation. The left panel corresponds to the P22 bacteriophage *spherical* procapsid. The right panel is the mature capsid, which has expanded and shows a polyhedral shape. The reconstruction have been extracted from Ref. [2].

energy, which takes into account the deviations from the spontaneous curvature imposed by the proteins, and the stretching energy, which involves the in-plane deformations of the network of proteins. The ratio of these contributions defines the Föppl-von Kárman number, $\gamma = Y^{\star}R^2/\kappa$, a dimensionless parameter that determines the shape of the virus. Here $Y^{\star}$ is the 2D Young modulus, κ is the bending rigidity, and R is the radius of the spherical shell. For small values of γ the spherical shape is energetically favored. But when γ exceeds a certain threshold, a buckling transition takes place and the polyhedral shape, with flat faces, becomes more stable because it reduces the stretching energy. Since $\gamma \sim R^2$, big icosahedral capsids are expected to be polyhedral rather than spherical, which is in agreement with experimental observations [22]. Nevertheless, the discrete nature and arrangement of capsomers in the viral shell might have a significant influence in the buckling phenomenon, specially for small viruses, which cannot be easily described in the framework of continuum elasticity theory.

In this chapter we will compare the energetics and mechanics of spherical and icosahedral viral capsids for different triangulation numbers T (see below), and determine how sensitive is the buckling transition to the discrete nature and precise arrangement of protein subunits in the shell. We will also analyze the tendency to buckle in terms of the distribution of lateral stresses, and discuss the potential mechanical and biological advantages of polyhedral versus spherical capsids.

4.2 Spherical and polyhedral shells

The CK construction described in Chapter 1 determines all possible icosahedral capsids that can be built. In order to understand the benefits of the buckling phenomenon, we will compare the properties of the spherical and polyhedral capsid shapes for the same T-number (see Table 4.1).

To build spherical shells we use the exact coordinates of the icosahedral spherical codes from Ref. [23], which in general correspond to local energy minima of the capsomer-capsomer interaction model studied in Chapter 2, and Ref [24]. Using these fixed angular coordinates, we have explored the energies and mechanical properties of spherical capsids under radial deformations.

To form the polyhedral shells, an icosahedron is reconstructed using the basic triangles defined by the CK construction (see Appendix A). Starting from the h and k values of a given T-number, we build the associated equilateral triangle and keep track of the coordinates of all capsomers, i.e., nodes of the hexagonal lattice, in the triangular face. By applying the proper rotations and translations, we build an icosahedron out of 20 of these triangles, whose centers are placed on the vertexes of its dual dodecahedron. This procedure generates the positions of all capsomers on an icosahedron of edge $l = \sqrt{T}$ and radius $R = \frac{1}{4}\sqrt{5+\sqrt{5}}\,l$, defined as the distance from the center to any vertex [25]. Finally, the optimal polyhedral shell is obtained by varying the radius R to minimize the total energy, as shown in Section 4.4. Note that we keep the same icosahedral shape for all radii, but in each case the distances between capsomers are uniformly rescaled.

4.3 Mechanical characterization

The energy of the resulting capsids can be simply evaluated as the sum over all pairs of capsomer-capsomer interactions, using as intermolecular potential the

	T							
$P \setminus f$	1	2	3	4	5	6	7	8
1	1	4	9	16	25	36	49	64
3	3	12	27	48	75			
7	7	28	63	112				
13	13	52	117					

Table 4.1: T-numbers for the first four P classes (rows) classified by the multiplicity f (columns), $T = Pf^2$ (see Eq. (1.2) in Chapter 1).

Lennard-Jones-like model described in Chapter 2. For each capsid shape we evaluate the energy for different values of the radius as

$$E(R) = \sum_{i=1}^{N} \sum_{j<i} V_{ij}(R) \tag{4.1}$$

where the distance between capsomers (and so the interaction) depends on the radius of the shell.

In addition to their energy, we evaluate their mechanical properties through the stress tensor [26]. For discrete systems, the *global* stress is commonly measured as the average of the virial stress tensor, $\sigma_{\alpha\beta}$, and can be decomposed in the kinetic and force contributions of each morphological unit i

$$\Omega^d \sigma^d_{\alpha\beta} = \sum_i \Omega^d_i (\sigma^d_{\alpha\beta})_i = \sum_i \left[\sum_j \frac{1}{2} m_i v_i^\alpha v_j^\beta - \frac{1}{2} \sum_{i \neq j} \frac{dV(r_{ij})}{dr_{ij}} \frac{r_{ij}^\alpha r_{ij}^\beta}{r_{ij}} \right] \tag{4.2}$$

Here d is the dimensionality, usually 2D or 3D, and Ω^d is the *d-volume*, i.e., the surface in 2D or the volume of the shell in 3D. In the kinetic term, m_i and v_i^α are the mass and the α component of the velocity of capsomer i, respectively. In the force term, r_{ij} is the module of the vector $\vec{r}_j - \vec{r}_i$ that joins particles i and j, and r_{ij}^α is the α component of this vector. Finally, Ω_i^d corresponds to the d-volume per capsomer, which can be defined in different ways, all leading to similar qualitative results [26]. Here we use the simplest choice where each capsomer has the same d-volume $\Omega_i^d = \Omega^d/N$. Using these definitions, the local, $(\sigma^d_{\alpha\beta})_i$, and global, $\sigma^d_{\alpha\beta}$, mechanical stresses can be computed.

Giving the high relative strength of the interactions compared to the thermal energy, $\varepsilon_0 = 15\, k_B T$, the kinetic contribution of the stress tensor in Eq. (4.2) will be neglected. In a first approximation the interactions of the capsid determine the main properties of the system, whereas thermal fluctuations would be relevant only in more accurate studies, e.g., the role of soft modes in the activation of buckling [27]. For spherical shells the stress tensor, Eq. (4.2), will be expressed in spherical coordinates, and due to the symmetry of viral capsids it is worth to recombine its components in three terms [26]: the lateral stress

$$(\sigma_T^{2D})_i = -\frac{N}{2A} \sum_j \frac{dV(r_{ij})/dr_{ij}}{r_{ij}} \frac{(\vec{r}_{ij} \cdot \widehat{e_\theta})^2 + (\vec{r}_{ij} \cdot \widehat{e_\phi})^2}{2} \tag{4.3}$$

the (45^o) shear stress

$$(\sigma_{\theta\phi}^{2D})_i = -\frac{N}{2A} \sum_j \frac{dV(r_{ij})/dr_{ij}}{r_{ij}} (\vec{r}_{ij} \cdot \widehat{e_\theta})(\vec{r}_{ij} \cdot \widehat{e_\phi}) \tag{4.4}$$

and the radial stress

$$(\sigma_R^{3D})_i = -\frac{N}{2V}\sum_j \frac{dV(r_{ij})/dr_{ij}}{r_{ij}}(\vec{r}_{ij}\cdot\widehat{e_r})^2 \tag{4.5}$$

where A and V are the surface and volume of the shell. For icosahedral shells the formulas are still valid but it is better to use a different set of basis vectors. In particular, for each face we replace $(\widehat{e_\theta}, \widehat{e_\phi})$ by two perpendicular vectors that are tangent to the surface, and $\widehat{e_r}$ by a vector perpendicular to the face at all points. For both capsid shapes, Eq. (4.3), (4.4), and (4.5) characterize the local stresses, and using them into Eq. (4.2) similar expressions for the global mechanics of capsids are obtained.

Finally, we will analyze the pressure p and the bulk modulus K, which measures the incompressibility of the capsid. It is useful to rewrite both in terms of energy variations as (see Chapter 3)

$$p = -\frac{\partial E}{\partial V} = -\frac{R}{3V}\frac{\partial E}{\partial R} \tag{4.6}$$

and

$$K = -V\left(\frac{\partial p}{\partial V}\right) = \frac{R^2}{9V}\left(\frac{\partial^2 E}{\partial R^2}\right) \tag{4.7}$$

where we have transformed volume derivatives into radial derivatives, using an expression valid for any system where $V \sim R^3$. We remark that for all optimal structures the pressure is zero [26], and their compressibility will be labeled as K_0.

4.4 Energy of spherical vs polyhedral capsids

The first step was to find the optimal radii, R_0^{sph} and R_0^{ico}, for the spherical and polyhedral capsids with T-numbers listed in Table 4.1. To that end, starting from the exact angular coordinates of the spherical and icosahedral shells described in Section 4.2, we subjected the capsids to a radial deformation and calculated the associated energy, as shown in Fig. 4.2 for the $T = 13$ case. For each shape and T-number we obtained a nearly parabolic behavior around the energy minima. For larger expansions, the compression branch is steeper than the expansion branch, due to the details of the Lennard-Jones interaction.

Fig. 4.3 plots the optimal radii of spherical and polyhedral shells as a function of the T-number. In both cases the radii increase as $R_{opt} \sim \sqrt{T}$, since the T-number is proportional to the shell surface, which grows as $\sim R^2$. Comparing the

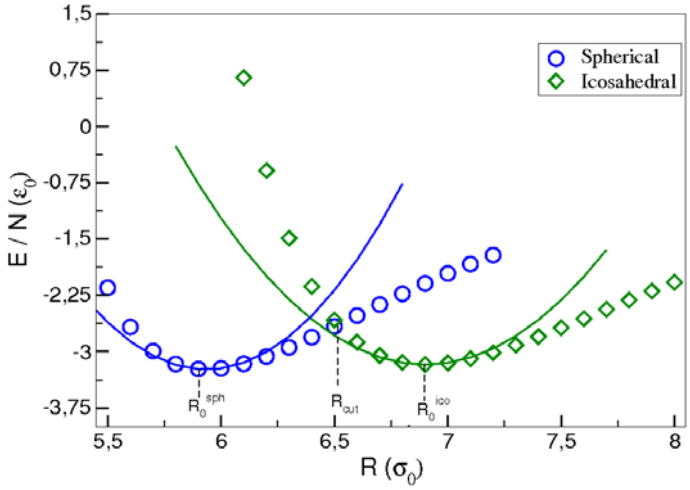


Figure 4.2: Energy per capsomer for the spherical and polyhedral $T = 13$ shells under radial deformation. The optimal radii of the sphere, R_0^{sph}, the icosahedron, R_0^{ico}, and the cut radius where both energies intersect, R_c, are also indicated. The lines show a parabolic fit $E = E_0 + \frac{1}{2}a_E(R - R_0)^2$ around the respective minima (see text).

optimal radii, we systematically observe that the polyhedral shells are larger than the spherical capsids typically by $\sim 15\%$. This is precisely the result predicted by the model introduced in Chapter 3. Indeed, the surface covering model is in excellent agreement with the equilibrium results, which is a direct consequence of the prevalence of steric effects between capsomers. Therefore, the results obtained for the optimal radii are essentially independent of the details of the interaction model.

Fig. 4.4 plots the energy per capsomer of the optimal spherical and polyhedral structures as a function of the T-number. For $T \leq 4$, the energy is dominated by the 12 pentamers rather than by the arrangement of the few existing hexamers, leading to a behavior that deviates from that of the remaining shells of a given class. Instead, for $T > 4$, an important dependence on the icosahedral class P can be observed. In particular, T-structures with $P = 1$ (see Table 4.1) are local maxima in the energy landscape of spherical shells, but local minima in the polyhedral one. This suggests that $P = 1$ capsids are not particularly satisfied in the spherical shape and will have a strong tendency to become faceted. On the other hand, $P = 3$ capsids show the opposite behavior of $P = 1$: they are local minima in the spherical landscape and maxima in the faceted one. The origin of these

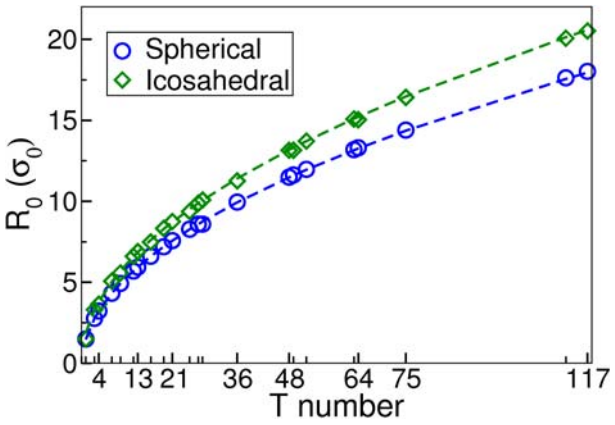
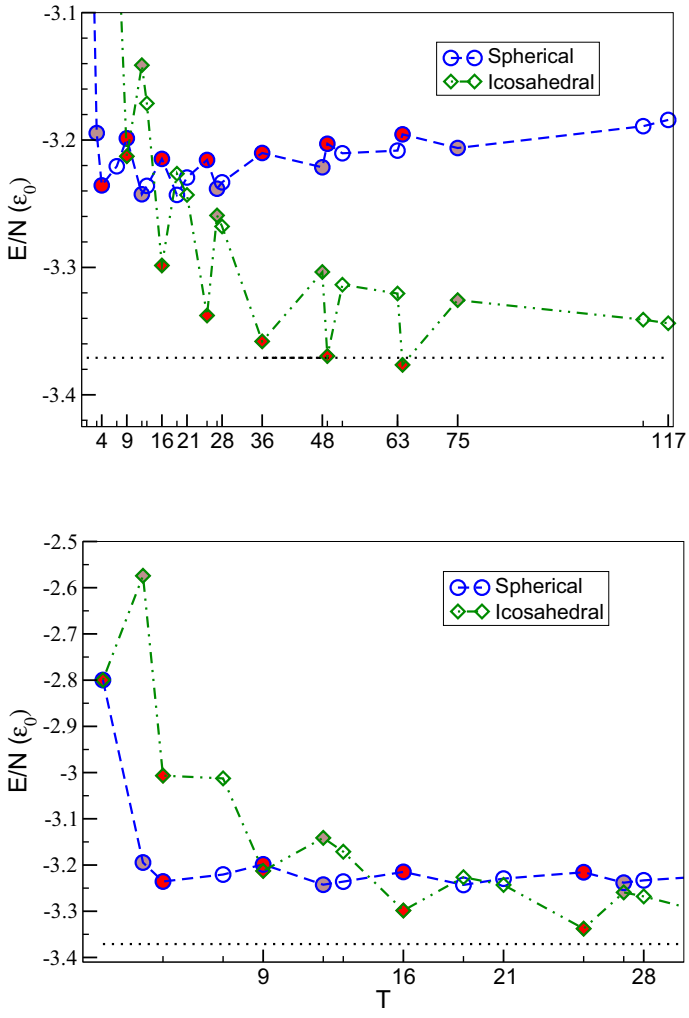


Figure 4.3: Spherical and polyhedral optimal radii as a function of the T-number. The lines are the theoretical predictions of the surface-covering model (see Chapter 3), and the points are obtained by minimizing the energy of the different structures (see text).

energy differences comes from the specific distribution of capsomers for each class P, as will become evident in the analysis of the lateral stress (see Section 4.7). For big capsids, the energy dependence on P attenuates, and polyhedral shells tend to the energy of a flat hexagonal lattice of hexamers. Furthermore, for big capsids ($T > 28$ in our case) spherical shells have systematically a higher energy than icosahedral ones, in agreement with the predictions of continuum elasticity theory [11]. The energetic frustration of the spherical shape comes from the pentameric disclinations that introduce a stretching energy penalty proportional to R^2, which yields a near constant jump in the energy per capsomer with respect to the flat hexagonal lattice. On the other hand, for small structures ($T \leq 28$) one has a complex scenario with a strong competition between spherical and polyhedral shells. This structural dependence of the optimal energies is difficult to justify in the framework of continuum elasticity, since its origin lies on the discrete and specific distribution of capsomers for the different P-shells rather than on generic elastic properties.

4.5 Buckling of spherical capsids upon expansion

Interestingly, the energy analysis for the different shapes under radial deformation already reveals the tendency of spherical capsids to become icosahedral upon



(b)

Figure 4.4: Energy per capsomer of quasi-spherical capsids. (a) Optimal energy per capsomer for the spherical (blue circles) and icosahedral (green diamonds) T shells. The T numbers corresponding to the $P = 1$ and $P = 3$ classes are highlighted in red and brown, respectively. (b) Same data as in (a), showing the T-numbers missing in (a), and facilitating the appreciation of the competition between spherical and icosahedral shapes for $T < 28$.

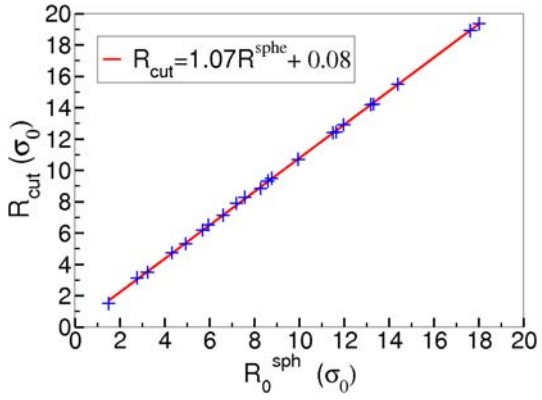


Figure 4.5: Cut radius R_c as a function of R_0^{sph}. The line corresponds to a linear fit.

expansion, as predicted using simple arguments in the previous chapter. This is illustrated in Fig. 4.2 for a $T = 13$ capsid. Upon expansion, the energy of the spherical shell increases and eventually crosses the compression branch of the polyhedral structure. This crossing defines a transition radius R_c beyond which the energy of the expanded spherical shell is larger than that of the compressed faceted capsid, thus signaling the energetic onset of buckling for this structure, E_c. Plotting the radii R_c as a function of the optimal radii R_0^{sph} for different T-numbers (see Fig. 4.5) a linear dependence is obtained, which suggests that spherical capsids will undergo a buckling transition when they are subjected to a radial expansion of $\sim 7\%$. This is precisely the result predicted in the theoretical analysis of Chapter 3[1]. Notice that, in the physical model, the deformation energy plays a major role to determine the transition point, and the energy difference between optimal structures introduces only a minor correction, which is precisely the scenario discussed in the theoretical analysis. However, if we look carefully, the cut radius of the physical model is slightly bigger, due to the small deviations with respect to the parabolic behavior of the energy around the transition point. In any case, our observations are consistent with the capsid expansions observed during the maturation process of many viruses, which lead to the faceting of an initially spherical procapsid [6] (see Fig. 4.1).

[1]It is worth to notice that Mannige and Brooks suggested in Ref. [28], based on geometrical considerations, that only capsids with $T \geq 7$ could undergo a buckling transition with the help of auxiliary proteins. Instead, our results indicate that under radial expansion any spherical T-shell could in principle undergo such transition.

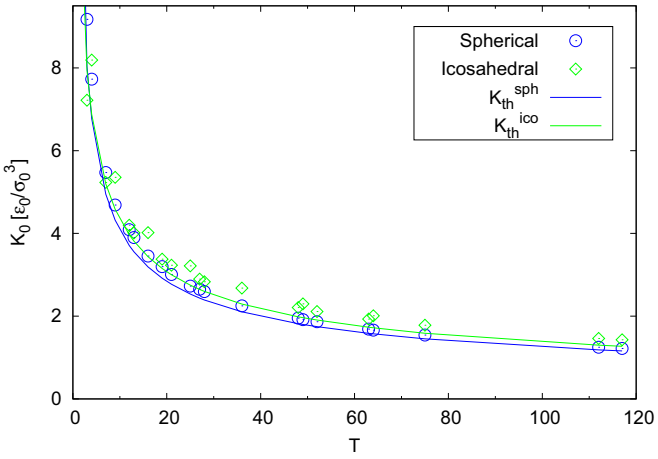


Figure 4.6: Theoretical and numerical results for the bulk modulus at the optimal radii. The blue circles and the green diamonds are the results obtained in this chapter for the spherical and icosahedral T-shells, respectively. The two curves are the theoretical predictions derived in Chapter 3 for a Lennard-Jones like interaction at first neighbors. The $T = 1$ capsid is not shown.

4.6 Bulk modulus

From the exploration of the energy under radial deformation, the bulk modulus of the optimal structures, K_0, can also be extracted. As shown in Fig. 4.2, the energy is parabolic around the energy minimum, and can be expressed as $E = E_0 + \frac{1}{2}a_E(R - R_0)^2$, Eq. (3.17). Then, by substituting $a_E = \left(\partial^2 E/\partial R^2\right)_{R_0}$ into Eq. (4.7), we determine K_0 for the different shapes and T-numbers (see Fig. 4.6).

The theoretical prediction of the bulk modulus derived in Chapter 3, and based on a mean-field approximation, is also plotted in Fig. 4.6. Despite the simplicity of the arguments introduced in the mechanical theory, the predictions are in a very good agreement with the numerical results. For both capsids shapes the bulk modulus decays inversely proportional to their corresponding optimal radii, i.e., $K_0 \sim 1/R_0$, as predicted, so in general larger capsids will be more compressible. This is a a direct consequence of the geometrical prefactor $R^2/V \sim 1/R$ in Eq. (4.7) and the fact that the energy is proportional to R^2 under radial deformation. Except for $T = 3$ and $T = 7$, the bulk modulus of faceted shells, K_0^{ico}, is systematically larger than that of the spherical ones, K_0^{sph}, by a factor

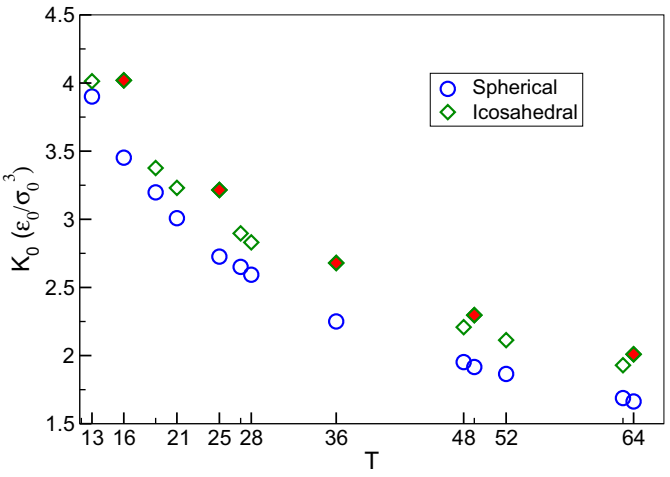


Figure 4.7: Bulk modulus for the spherical and polyhedral shells at the optimal radii. A zoom of Fig. 4.6 for the region $13 < T < 64$ is shown to highlight the fact that polyhedral $P = 1$ shells (colored in red) are comparatively stiffer.

ranging from $2-10\%$ for small capsids to almost 20% for bigger ones, which is also in agreement with our predictions (note that the theory also predicts an inversion of behavior at $T \sim 4$). The values of the bulk modulus obtained in the numerical exploration are slightly higher than the prediction. This is probably due to the fact that, in the theoretical model, the LJ interaction was only considered up to the nearest neighbors and the deformation energies of the ridges of the icosahedron where not taken into account. In any case, it is surprising that such a simple theory (without any fitting parameter), could describe the main mechanical properties of quasi-spherical shells. Finally, there seems to be no special structural dependence in the bulk modulus of spherical shells. Contrarily, polyhedral $P = 1$ capsids are comparatively stiffer than those of any other class (see Fig. 4.7). This dependence on the particular P cannot be captured by the theory.

4.7 Local distribution of stress

To elucidate the relevance of the structure in the mechanical response of viral capsids, we computed the local distribution of stresses at the level of capsomers for the different T-shells. Our analysis is based on the previous work done by Zandi and Reguera in Ref. [26], where the mechanical properties of spherical capsids were studied in detail using a similar model. We will focus on the distribution of local lateral stresses, using the convention that a positive or negative stress means that a capsomer is being compressed or stretched, respectively.

The distribution of stresses for the different T-shells can be naturally grouped in classes that show a similar pattern. To illustrate how the precise arrangement of capsomers influences the tendency of a capsid to remain spherical or to become faceted, we have analyzed series of T-numbers corresponding to the first four P-classes (see Table 4.1).

The simplest case is $P = 1$, where each pair of neighboring pentamers are connected by a straight line of hexamers (see Fig. 4.8). This structural feature clearly dictates the local stress pattern. In the spherical shells, positive stress concentrates on the lines connecting pentamers, which delimits triangular areas with stretched hexamers. The absolute values of those stresses get larger as the the T-number of the shell increases. Compared with the rest of spherical P-shells (see below), the $P = 1$ class shows the highest local stretching and compression. This stress frustration, associated to the geometrical configuration of capsomers in $P = 1$ capsids, is the underlying reason why they are local maxima in the optimal energy landscape of Fig. 4.4. Moreover, the stress distribution clearly highlights the energetic advantage of adopting a polyhedral shape, since the hexamer stretching will be relieved if the triangular regions between pentamers are flattened. In fact, the distribution of stress in the polyhedral $P = 1$ capsids shows that the local lateral stresses at the triangular faces are close to zero and get smaller as the size of the shell increases. Furthermore, compared to the other classes, polyhedral $P = 1$ capsids have the lowest stress frustration, which justifies why they are local minima in the optimal energy landscape (see Fig. 4.4).

The other achiral class, $P = 3$, is characterized by zig-zag lines of hexamers connecting pentamers, instead of straight lines (see Fig. 4.9), which leads to an inversion of the stress behavior. In spherical $P = 3$ shells, compressions concentrates on the pentamers, whereas the hexamers surrounding them are slightly stretched, leading to a dodecahedral pattern instead of the icosahedral one observed in $P = 1$ (see Fig. 4.8). In this case, nothing indicates any particular relief of the stress on the hexamers upon adopting an icosahedral shape. On the contrary, the stress pattern suggests that the shell will flatten along the lines

connecting second neighbor pentamers, adopting a dodecahedral shape rather than the icosahedral one observed in $P = 1$. It is worth to mention that such a dodecahedral faceting seems to be present in the reconstruction images of *Chilo iridescent virus* (CIV), a large $T = 147$ ($P = 3, f = 7$) virus [29]. In the spherical $P = 3$ shell, due to the symmetric capsomer arrangement, the local stress is smoothly shared among all capsomers and its values are low enough to produce local minima in the optimal energy landscape of Fig. 4.4. Contrarily, in the polyhedral shape, the $P = 3$ class has a higher stress frustration compared to the faceted $P = 1$, showing a less uniform pattern, where stress accumulates at the zig-zag lines along the edges. In fact, the icosahedron $P = 3$ construction is the most stressed of all polyhedral classes leading to local maxima in the optimal energy per capsomer (see Fig. 4.4).

The remaining P classes are chiral, i.e., for each shell we have two specular structures with different handedness. The arrangement of hexamers along the line connecting neighboring pentamers is now skewed, and the resulting stress patterns show neither a clear advantage or disadvantage on adopting the shape of an icosahedron (see Figs. 4.10 and 4.11). Accordingly, the values of optimal energies and stresses are intermediate between those of $P = 1$ and $P = 3$. Therefore, classes with $h_0 \sim k_0$ and $h_0 \gg k_0$ will behave similarly to $P = 3$ and $P = 1$, respectively.

Finally, a common feature for all spherical classes is that positive stress accumulates on the pentameric disclinations as the shell gets bigger, in agreement with the predictions of continuum elasticity theory [11]. The accumulation of stress on the pentamers of big shells will eventually lead to a buckling transition towards a faceted shape, and provides a mechanical justification of why for all larger capsids the polyhedral shape seems to be energetically favored.

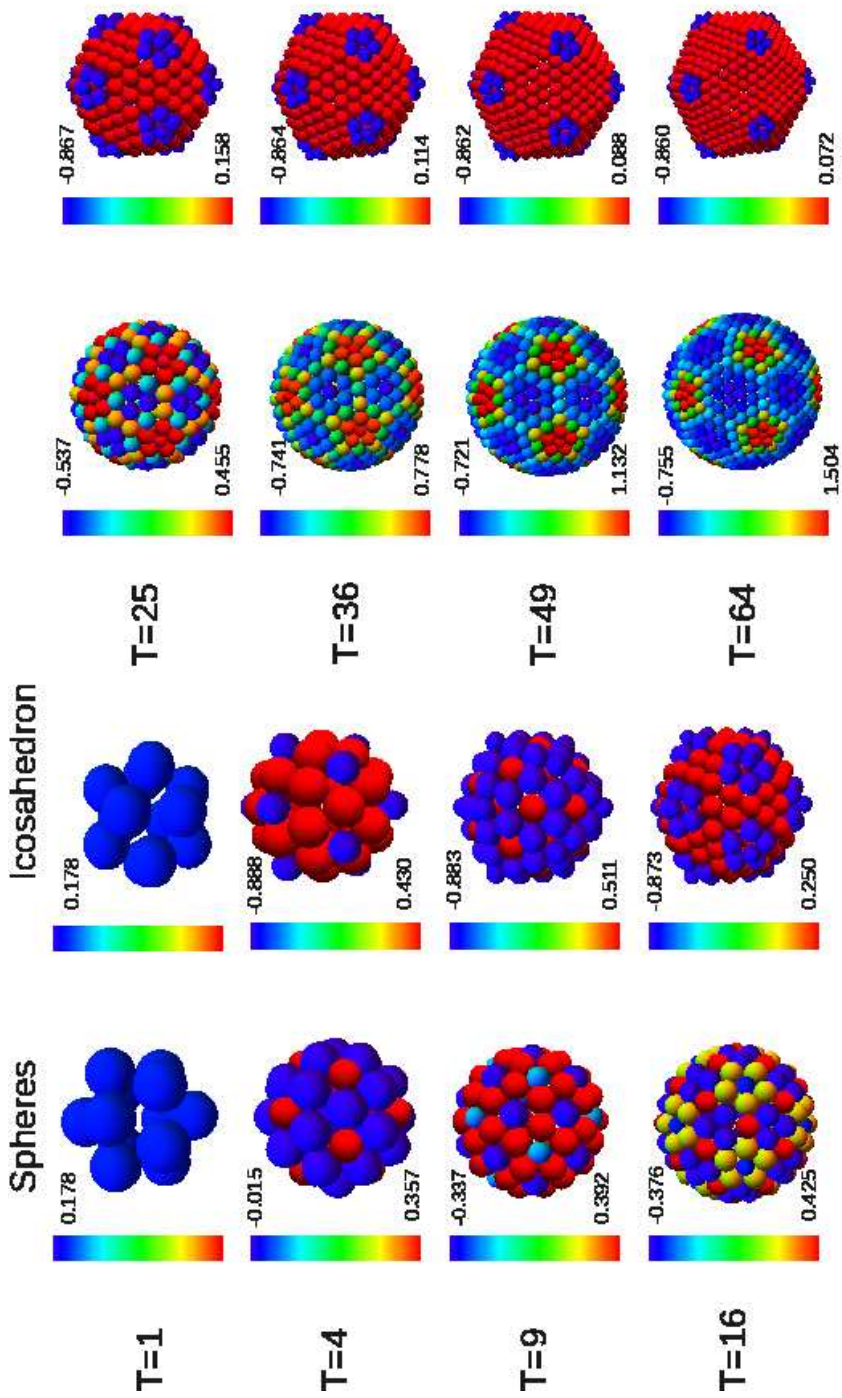


Figure 4.8: Lateral stress distribution for $P = 1$ spherical and polyhedral shells. The color bar indicates the values of the local stresses, and it is rescaled for each structure using the most compressed (positive, red) and most stretched (negative, blue) capsomers.

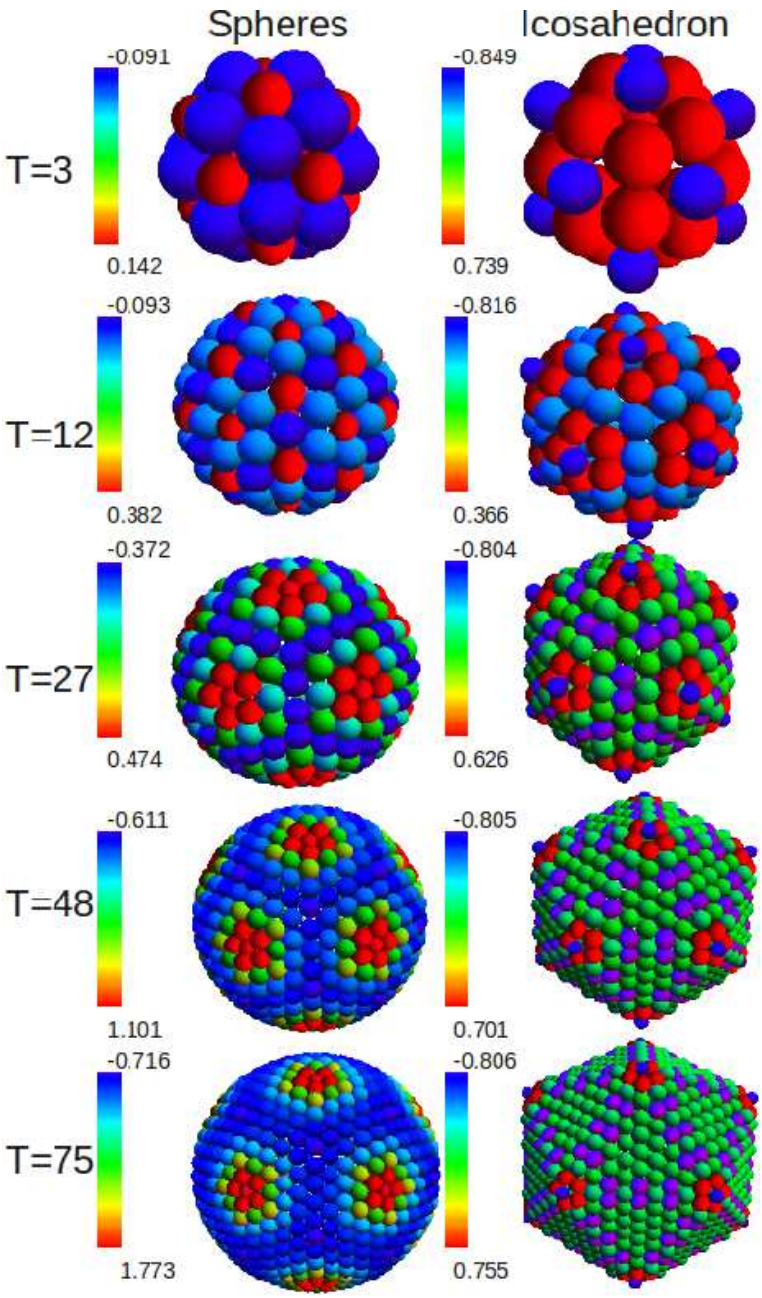


Figure 4.9: Lateral stress distribution for $P = 3$ spherical and polyhedral shells. The color code is described in Fig. 4.8.

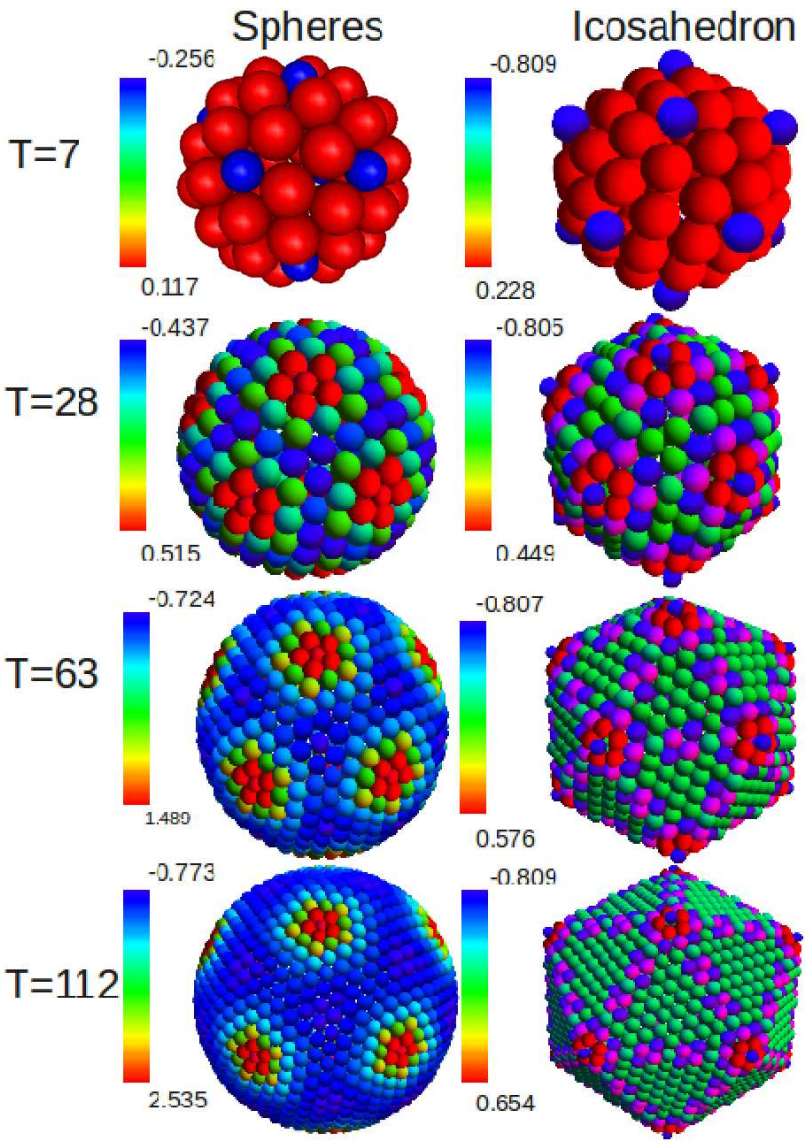


Figure 4.10: Lateral stress distribution for $P = 7$ spherical and polyhedral shells. The color code is described in Fig. 4.8.

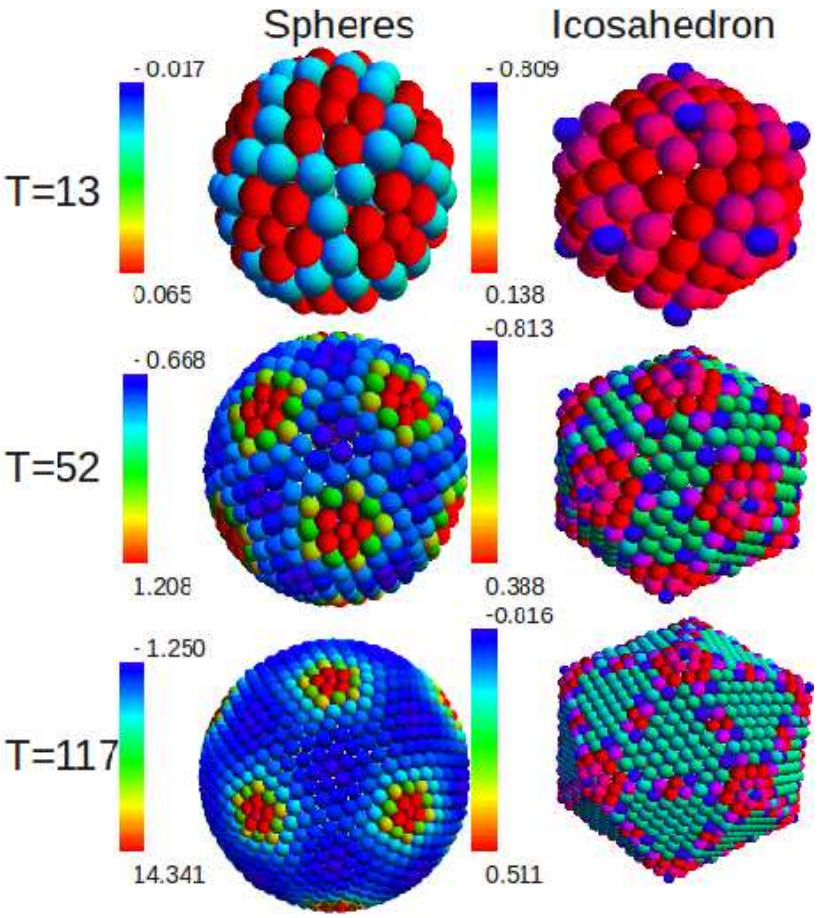


Figure 4.11: Lateral stress distribution for $P = 13$ spherical and polyhedral shells. The color code is described in Fig. 4.8.

4.8 Pressure and bursting

An important property of capsids is their tolerance to internal pressure before bursting. To compare the resistance of both capsid shapes, we calculated the maximum pressure that will lead to lateral failure of the shells.

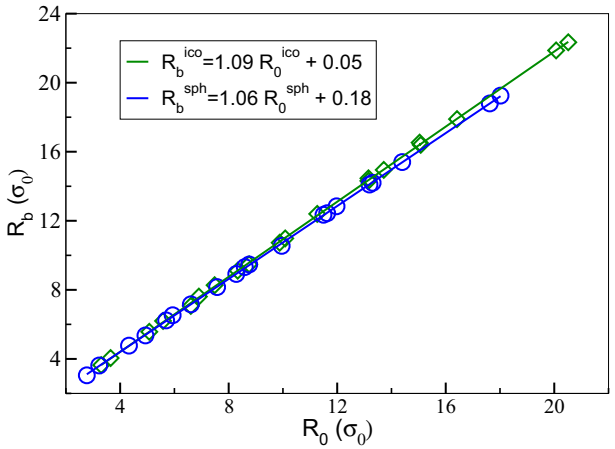
In the previous chapter it was shown that the maximum tensile force that two LJ particles can tolerate is given by $f_b \approx -2.69\varepsilon_0/\sigma_0$, Eq. (3.53), that takes place at an intermolecular distance $r_b \approx 1.11\sigma_0$, Eq. (3.54). For a 2D LJ lattice subjected to a uniform deformation given by r_b, one can compute now the local 2D lateral stress for a capsomer in this situation using Eq. (4.3), which gives $\sigma_b^{2D} \approx -1.1\varepsilon_0/\sigma_0^2$. Remarkably, this rough limit of resistance of a flat hexagonal lattice of LJ particles [30], was found to be a good approximation for the onset of bursting in the simulations of spherical capsids under expansion in Ref. [26]. Thus to determine the bursting pressure of a capsid, radial expansions on each quasi-spherical shell were performed until a single capsomer reaches the lateral stress limit σ_b^{2D}.

The associated bursting radii, R_b, are plotted in Fig. 4.12, and show a linear dependence on the optimal radii, which correspond to expansions of roughly $\sim 6\%$ and $\sim 9\%$ for spherical and polyhedral shells, respectively. Hence, polyhedral capsids resist larger relative expansions than their spherical counterparts. The associated internal pressures at the onset of bursting are plotted in Fig. 4.13. We observe that polyhedral capsids systematically tolerate higher values of the bursting pressure, p_b, than their spherical counterparts (except for $T = 3$ and $T = 7$), ranging from $\sim 10 - 20\%$ more for the smallest T-numbers up to a $\sim 35\%$ for the largest viruses. This advantage is specially evident for the $P = 1$ capsids. However, in general the bursting pressure decays with shell's size as $p_b \sim 1/R_b$, because $V \sim R^3$ and under deformation $E \sim R^2$ (see Eq. 4.6 and Chapter 3).

4.9 Potential biological implications

The aim of this section is to discuss the biological implications of the results previously derived, and their relation with virus maturation.

The maturation process of a virus can be very complex, involving capsid expansion, cleavage of proteins, and even covalent reinforcement of the shell [2,4,6, 8]. Therefore, to isolate the fundamental mechanisms responsible for the buckling phenomenon, we focus first on the simple case of bacteriophage P22, a $T = 7$ virus with a maturation pathway that does not require auxiliary proteins, cleavage of



(a)

Figure 4.12: Bursting radius as a function of the optimal radii for the spherical and polyhedral shells. The lines represent linear fits to the data.

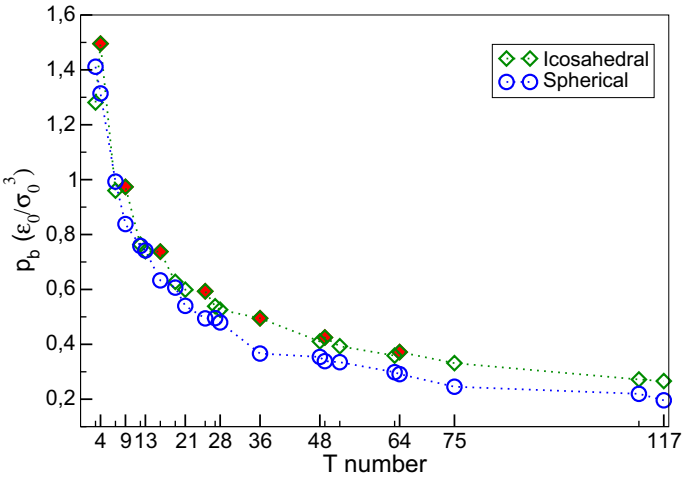


Figure 4.13: Bursting pressure as a function of the T-number. Icosahedral $P = 1$ shells are highlighted in red.

the coat protein, nor cross-linking between subunits [3]. Comparing the spherical procapsid and the larger mature icosahedral shell of P22 [2] (see Fig. 4.1), one obtains a ratio 1.16 ± 0.04 between their radii, which is in excellent agreement with the expected ratio 1.17 obtained in the simulations for $T = 7$ shells. We stress that our simple surface-covering model accurately reproduces the radii obtained in the simulations (see Fig. 4.3 and in Chapter 3), meaning that the result should be robust to different types of capsomer-capsomer interaction. However, if during virus maturation the interactions or the effective size of capsomers are strongly altered by conformational changes or chemical processes, important deviations could be expected. For instance, *Nudaurelia capensis* ω virus (NωV) is a $T = 4$ virus that assembles in spherical shape that after a reduction of the pH and a cleavage in the coat protein leads to an irreversible final faceted capsid 15% smaller in radius than the initial procapsid, i.e., the opposite behavior of the one expected in our study [9, 10]. The realistic scenario for viruses that trigger the maturation and the buckling transition by important changes in the viral coat proteins, and accordingly in their effective interactions, goes beyond the scope of the present analysis.

As mentioned in a previous section, buckling of T-shells produces polyhedral capsids with radii $\sim 15\%$ bigger than the spherical ones. However, the volume of the resulting icosahedron V_0^{ico} is $\sim 10\%$ smaller than the spherical one V_0^{sph}. This counterintuitive result is coherent with the fact that both capsids have the same number of capsomers, i.e., the same surface, and the sphere is the shape that maximizes the volume/surface ratio. In addition, this result brings up an interesting question. Many viruses that undergo a maturation process with buckling end up storing dsDNA at high density [31], so the capsid must sustain an important internal pressure due to the confinement of this semi-flexible and electrostatically charged polymer. The dsDNA density is correlated with the internal pressure [32], which means that the same amount of genetic material will generate a higher pressure in the polyhedral shell than in the spherical capsid. This might be a desirable effect in viruses that take advantage of this internal pressure to initiate the infection [33].

Another interesting consequence of our results is that for $T \leq 28$, the class P clearly dictates the energetic preference in adopting either a spherical or a polyhedral capsid. For instance, $T = 13$ shells tend to be more stable in the spherical shape, and in fact some $T = 13$ dsRNA viruses like bluetongue virus (BTV) or rotavirus remain spherical even after maturation [34]. On the other hand, in viruses where the icosahedral shell is the most stable, like for $T = 25$, the capsid assembly might produce a polyhedral shell already as a procapsid, as

has been observed in the $T = 25$ bacteriophage PRD1 [35].

4.10 Conclusions

In this chapter we have analyzed the relevance of the T-number of viruses in the buckling phenomenon, using a simple physical model that captures the essential ingredients of capsomer-capsomer interactions and that successfully reproduces the equilibrium structures of viral capsids, as we showed in Chapter 2. Despite the simplicity of the model, it is important to point out that the qualitative results obtained do not depend on the particular choice of the details of the intercapsomer potential, but are rather dictated purely by the geometrical arrangement of the capsomer in the T-shells. Obviously, the actual numerical values of properties like the local stresses, energies, pressures or bursting radii do depend on the details of the potential; but the trends and general conclusions that we summarize below seem to be insensitive to them and they are just determined by general physical and geometrical considerations.

We have found that, for small T-numbers, the tendency to buckle strongly depends on the class P. In particular, structures from the class $P = 1$, e.g., $T = 9\,, 16$ or 25 ($T = 4$ is the only exception), are the most favorable to produce polyhedral shells, whereas capsids from the class $P = 3$ e.g., $T = 3$, 12 or 27, are more stable as spheres. For the chiral classes $P > 3$ an intermediate behavior is obtained. Nevertheless, independently on P, for big capsids ($T > 28$) the polyhedral shell is always more stable than the spherical shell, in agreement with continuum elastic theory [11].

Furthermore, the analysis of the local lateral stress distribution unveils the microscopic explanation of the different tendency to buckle observed in different classes P. For spherical shells one generally observes highly squeezed pentameric zones and stretched hexamers, which is in agreement with the theory of disclinations in hexagonal lattices [36]. However, the class P dictates the capsomer arrangement and the resulting pattern of stress, which for spherical shells indicate the tendency of a structure to produce a polyhedral shell.

Interestingly, even in the cases where the spherical shape is more stable, viruses tend to undergo a buckling transition and to become polyhedral upon expansion, as it is often the case during virus maturation. In fact, several $T = 7$ viruses adopt spherical procapsids but end up becoming faceted after suffering a buckling transition triggered by an expansion during virus maturation [2, 4, 6, 8].

More importantly, the choice of a polyhedral instead of spherical shape seems to have mechanical advantages that might play an important biological role.

Faceted icosahedral shells have higher bulk moduli and tolerate larger expansions before bursting. In addition, they are able to withstand internal pressures $\sim$ 20% higher than spherical capsids. They also have smaller volumes than their spherical counterparts, which could increase the pressurization of the confined genetic material. All these enhanced properties could be advantageous for viruses that rely on a pressurized capsid to initiate the genetic material ejection [33]. However, since the maximum tolerable pressure is inversely proportional to the radius of the shell [26], this suggests that viruses could only take advantage of an internal pressurization mechanism for a specific range of low T's.

Finally, even though many viruses show a buckling transition during maturation, the pathways and processes involved can be very complex, including auxiliary proteins, cleavage or cross-linking. Obviously, our simple model cannot describe these pathways nor capture those complications. But our results could be helpful in understanding the biophysical advantages of undergoing a maturation and buckling process and adopting a faceted shape.

Appendices

A Polyhedral shell construction

The position of all capsomers in a polyhedral shell is obtained by reconstructing an icosahedron using as faces the basic triangles of the Caspar and Klug model. Starting from the h and k values of a given T-number, we build the associated equilateral triangle on a hexagonal lattice and keep track of the x, y coordinates of all capsomers, i.e. nodes of the hexagonal lattice, that lie inside this triangle of edge $l = \sqrt{T}$. By applying a rotation of $\arctan(\frac{1}{\sqrt{3}}\frac{k-h}{h+k})$ around the z-axis, and a translation of $(-l/2, -l/(2\sqrt{3}), R)$ we place the center of the equilateral triangle at $(0, 0, R)$, where $R = \frac{\sqrt{3}}{12}(3 + \sqrt{5})l$ is the distance from the center to each face of an icosahedron of edge length l. We use then 20 copies of this triangle to reconstruct the icosahedron.

The angular coordinates of the center of those triangles in the final icosahedron are given by the position of the vertexes of the dodecahedron of unit edge, which is the dual structure of the icosahedron: $(\pm 1, \pm 1, \pm 1)$, $(0, \pm 1/\phi, \pm\phi)$, $(\pm 1/\phi, \pm\phi, 0)$, and $(\pm\phi, 0, \pm 1/\phi)$, where $\phi = (1 + \sqrt{5})/2$ is the "golden ratio". After transforming those points into spherical coordinates (r, θ, φ), the icosahedron is reconstructed from the basic CK triangle located at $(0, 0, R)$ by applying a custom rotation around the z-axis, followed by one of an angle θ around the y-axis and one of φ around the z-axis.

This procedure was implemented using Mathematica© [37]. In this way we generated the (x,y,z) coordinates of all capsomers on an icosahedron of edge $l = \sqrt{T}$ and radius $R = \frac{1}{4}\sqrt{10 + 2\sqrt{5}}\, l$, defined as the distance from the center to any vertex.

References

[1] M. F. Moody, "Geometry of phage head construction," *J. Mol. Biol.*, vol. 293, pp. 401–33, 1999.

[2] W. Jiang, Z. Li, Z. Zhang, M. L. Baker, P. E. Prevelige, and W. Chiu, "Coat protein fold and maturation transition of bacteriophage P22 seen at subnanometer resolutions," *Nat. Struct. Biol.*, vol. 10, pp. 131–5, 2003.

[3] K. N. Parent, R. Khayat, L. H. Tu, M. M. Suhanovsky, J. R. Cortines, C. M. Teschke, J. E. Johnson, and T. S. Baker, "P22 coat protein structures reveal a novel mechanism for capsid maturation: stability without auxiliary proteins or chemical crosslinks," *Structure*, vol. 18, pp. 390–401, 2010.

[4] A. Ionel, J. A. Velázquez-Muriel, D. Luque, A. Cuervo, J. R. Castón, J. M. Valpuesta, J. Martín-Benito, and J. L. Carrascosa, "Molecular rearrangements involved in the capsid shell maturation of bacteriophage T7," *J. Biol. Chem.*, vol. 286, pp. 234–42, 2011.

[5] X. Agirrezabala, J. Martín-Benito, J. R. Castón, R. Miranda, J. M. Valpuesta, and J. L. Carrascosa, "Maturation of phage T7 involves structural modification of both shell and inner core components," *EMBO J.*, vol. 24, pp. 3820–9, 2005.

[6] A. C. Steven, J. B. Heymann, N. Cheng, B. L. Trus, and J. F. Conway, "Virus maturation: dynamics and mechanism of a stabilizing structural transition that leads to infectivity," *Curr. Opin. Struc. Biol.*, vol. 15, pp. 227–36, 2005.

[7] I. Gertsman, L. Gan, M. Guttman, K. Lee, J. A. Speir, R. L. Duda, R. W. Hendrix, E. A. Komives, and J. E. Johnson, "An unexpected twist in viral capsid maturation," *Nature*, vol. 458, pp. 646–50, 2009.

[8] J. E. Johnson, "Virus particle maturation: insights into elegantly programmed nanomachines," *Curr. Opin. Struc. Biol.*, vol. 20, pp. 210–216, 2010.

[9] B. Bothner, D. Taylor, B. Jun, K. K. Lee, G. Siuzdak, C. P. Schlutz, and J. E. Johnson, "Maturation of a tetravirus capsid alters the dynamic properties and creates a metastable complex," *Virology*, vol. 334, pp. 17–27, 2005.

[10] J. Tang, K. K. Lee, B. Bothner, T. S. Baker, M. Yeager, and J. E. Johnson, "Dynamics and stability in maturation of a T=4 virus," *J. Mol. Biol.*, vol. 392, pp. 803–12, 2009.

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

[12] T. T. Nguyen, R. F. Bruinsman, and W. M. Gelbart, "Elasticity theory and shape transitions of viral shells," *Phys. Rev. E*, vol. 72, p. 051923, 2005.

[13] A. Šiber, "Buckling transition in icosahedral shells subjected to volume conservation constraint and pressure: Relations to virus maturation," *Phys. Rev. E*, vol. 73, p. 061915, 2006.

[14] M. Buenemann and P. Lenz, "Mechanical limits of viral capsids," *Proc. Natl. Acad. Sci. USA*, vol. 104, pp. 9925–30, 2007.

[15] M. Buenemann and P. Lenz, "Elastic properties and mechanical stability of chiral and filled viral capsids," *Phys. Rev. E*, vol. 78, p. 051924, 2008.

[16] G. A. Vliegenthart and G. Gompper, "Mechanical deformation of spherical viruses with icosahedral symmetry," *Biophys. J.*, vol. 91, pp. 834–41, 2006.

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

[18] T. Guerin and R. Bruinsma, "Theory of conformational transitions of viral shells," *Phys. Rev. E*, vol. 76, p. 061911, 2007.

[19] M. Gibbons and W. Klug, "Nonlinear finite-element analysis of nanoindentation of viral capsids," *Phys. Rev. E*, vol. 75, p. 031901, 2007.

[20] A. Ahadi, J. Colomo, and A. Evilevitch, "Three-dimensional simulation of nanoindentation response of viral capsids:shape and size effects," *J. Phys. Chem. B*, vol. 113, pp. 3370–8, 2009.

[21] C. Q. Ru, "Buckling of empty spherical viruses under external pressure," *J. Appl. Phys.*, vol. 105, pp. 3370–8, 2009.

[22] T. S. Baker, N. H. Olson, and S. D. Fuller, "Adding the third dimension to virus life cycles: Three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs," *Microbiol. Mol. Biol. Rev.*, vol. 63, pp. 862–922, 1999.

[23] R. H. Hardin, N. J. A. Sloane, and W. D. Smith, "Tables of spherical codes with icosahedral symmetry," *published electronically at http://www.research.att.com/ njas/icosahedral.codes/*

[24] R. Zandi, D. Reguera, R. F. Bruinsma, W. M. Gelbart, and J. Rudnick, "Origin of icosahedral symmetry in viruses," *Proc. Natl. Acad. Sci. USA.*, vol. 101, pp. 15556–15560, 2004.

[25] H. S. M. Coxeter, *Introduction to geometry.* Wiley, 2nd ed., 1989.

[26] R. Zandi and D. Reguera, "Mechanical properties of viral capsids," *Phys. Rev. E*, vol. 72, p. 021917, 2005.

[27] M. Widom, J. Lidmar, and D. R. Nelson, "Soft modes near the buckling transition of icosahedral shells," *Phys. Rev. E*, vol. 76, p. 031911, 2007.

[28] R. V. Mannige and C. L. Brooks, "Geometric considerations in virus capsid size specificity, auxiliary requirements, and buckling," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, pp. 8531–6, 2009.

[29] X. Yan, N. H. Olson, J. L. V. Etten, M. Bergoin, M. G. Rossmann, and T. S. Baker, "Structure and assembly of large lipid-containing dsdna viruses," *Nature structural biology*, vol. 7, pp. 101–3, 2000.

[30] R. L. B. Selinger, Z.-G. Wang, and W. M. Gelbart, "Effect of temperature and small-scale defects on the strength of solids," *J. Chem. Phys*, vol. 95, p. 9128, 1991.

[31] A. Evilevitch, L. Lavelle, C. M. Knobler, E. Raspaud, and W. M. Gelbart, "Osmotic pressure inhibition of DNA ejection from phage," *Proc. Natl. Acad. Sci. USA*, vol. 100, pp. 9292–9295, 2003.

[32] P. K. Purohit, J. Kondev, and R. Phillips, "Mechanics of dna packaging in viruses," *Proc. Natl. Acad. Sci. USA*, vol. 100, pp. 3173–3178, 2003.

[33] W. M. Gelbart and C. M. Knobler, "Pressurized viruses," *Science*, vol. 323, pp. 1682–1683, 2009.

[34] M. Carrillo-Tripp, C. Shepherd, I. A. Borelli, S. Venkataraman, G. Lander, P. Natarajan, J. E. Johnson, C. L. B. III, and V. S. Reddy, "Viperdb2: an enhanced and web api enabled relational database for structural virology," *Nucleic Acid Research*, vol. 37, pp. D436–D442, 2009.

[35] S. J. Butcher, D. H. Bamford, and S. D. Fuller, "DNA packaging orders the membrane of bacteriophage PRD1," *EMBO J.*, vol. 14, pp. 6078–6086, 1995.

[36] L. D. Landau and I. M. Lifshitz., *Theory of elasticity.* Pergamon, New York, 1975.

[37] I. Wolfram Research, *Mathematica edition: version 7.0.* Wolfram Research, Inc., 2008.

Chapter 5

Built-in Stress in Bacteriophage $\phi 29$

5.1 Introduction

Bacillus subtilis $\phi 29$ is a dsDNA bacteriophage, which, due to its relative simplicity, has been placed in the spotlight of research, providing deep insights into the strategies used by tailed viruses [1–3] (see Fig. 5.1). In general dsDNA phages first assemble in an empty prohead that later is filled with the genetic material. This process requires the participation of a packing motor and the consumption of chemical energy in the form of ATP. Interestingly, the energy stored during the packaging of dsDNA usually gives rise to an internal pressure, which can be used to initiate the translocation of the genetic material into the new host [4]. However, the principles and biological relevance of this mechanism remain controversial [5].

Bacteriophage $\phi 29$ is a prolate virus, whose capsid has dimensions $54\,nm \times 42\,nm$ and adopts an architecture $T_{end} = 3$ and $Q_{5F} = 5$ centered on a 5-fold axis, as have been discussed in Chapters 1 and 2 (see Fig. 5.1a). The shell of the prohead is constructed from 235 *gp*8 protein subunits and a dodecameric connector of protein *gp*10, which is placed on one end of the 5-fold axis, and plays the structural role of a penton [8]. Therefore, the bacilliform shell is build by 12 pentons (11 *gp*8-pentamers and the connector) plus 20 *gp*8-hexamers forming the icosahedral caps, and 10 *gp*8-hexamers forming the zigzag equatorial zone of the body [1, 9] (see Fig. 5.1b). The correct assembly of the prohead requires the interaction of the connector protein (*gp*10), the scaffolding protein (*gp*7) and the major head protein (*gp*8). The absence of the connector, the scaffolding proteins,

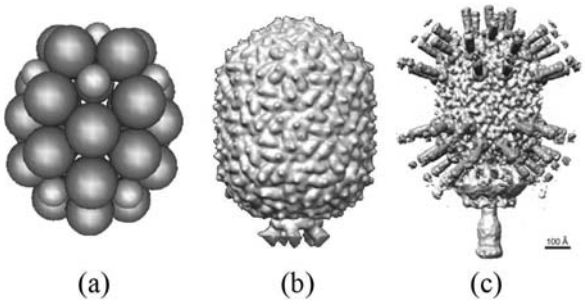


Figure 5.1: Bacteriophage $\phi 29$. (a) The $T_{end} = 3$ (5-fold) $Q_{5F} = 5$ bacilliform architecture derived from the physical model (see Chapter 2). (b) The prohead of the $\phi 29$ phage reconstructed by cryo-em [6]. (c) Mature virion structure derived from cryo-em [7].

or the presence of mutants of these proteins yield aberrant structures made of *gp*8 (open rounded shells, icosahedral capsids, tubular assemblies) [10]. This indicates that the *information* of the major head protein is not sufficient to define the shape and size of the virus head, but rather it is the interaction of these components what directs the precise curvature and extension of the contacts to generate the shell architecture [11, 12].

To produce the mature virion, the DNA is packaged into the prohead, which also ensures the complete release of the scaffolding protein [13]. Interestingly, contrarily to many other phages [14, 15], $\phi 29$ shows no expansion or structural differences in the capsid when the prohead and the filled mature particles are compared. Once the DNA has been completely stored, the connector complex interacts with other tail components (*gp*11, *gp*12 and *gp*9) to secure the DNA inside the head shell. In addition, head fibers made of protein *gp*8.5 are incorporated on the pentamers of the shell to enhance the attachment to the cell wall, even though they are dispensable for virus infectivity [7] (see Fig. 5.1c). Thus it is likely that, beyond the absence of head fibers and the tail, there are not major mechanically relevant differences between the prohead and the empty mature head particles [16]. Remarkably, the structure of the *gp*8 shell seems the same in both the prohead and the final mature virion.

To further infect a new host, a $\phi 29$ virion must attach to a bacillus subtilis cell and perform a two step injection mechanism. Initially, up to a 65% of the genome is released into the host driven probably by the high pressure inside the bacteriophage capsid. Subsequently, there is a push-pull mechanism were at least

one viral protein that has been expressed after the initial translocation step is involved [2].

The ubiquity of the pressure mechanism as a promoter of the initial injection of the viral genome for dsDNA bacteriophages, remains controversial [5, 17]. However, the conditions at which dsDNA is packed within the capsid generates an internal pressure that, independently on the ultimate role of this stored energy, must be withstand by the capsid [18]. Thus these viral particles must be designed in a way to be able to sustain this outward force that can be up to several tens of atmospheres.

The ultimately purpose of this chapter is to shade some light on the reinforcement mechanism of ϕ29. To that end, we have investigated the mechanical properties of empty ϕ29 proheads combining theory, simulations, and experiments. We will show that empty prolated ϕ29 bacteriophage proheads exhibit an intriguing anisotropic stiffness which behaves counter-intuitively different from standard continuum elasticity predictions. By using Atomic Force Microscopy (AFM), we find that the ϕ29 shells are about twice stiffer along the short than along the long axis. This result can be attributed to the existence of a residual stress due to the discrete nature of the capsid, a hypothesis that we confirm by coarse-grained simulations. This built-in stress of the virus prohead could be a strategy to provide extra mechanical strength to withstand the DNA compaction during and after packing, or a variety of extracellular conditions, such as osmotic shocks or dehydratation.

The chapter is organized as follows: in Section 5.2 we introduce the AFM experiments that measure the mechanical strength of ϕ29 proheads. Section 5.3 shows that basic continuum elasticity theory is not able to explain the experimental results. Simulations of the physical model introduced in Chapter 2 unveil, in Section 5.4, the mechanism of reinforcement associated to the existence of a residual stress. In Section 5.5 new AFM experiments are performed using a cross-linker that homogenizes the capsid shell leading to a continuum like mechanical response, which supports the residual stress mechanism. Finally, in Section 5.6 we summarize our findings and remark the implications on the bacteriophage ϕ29 infection strategy.

5.2 AFM experiments

In order to measure the mechanical properties of ϕ29, nanoindentation experiments on single procapids were performed by the group of P.J. de Pablo at the Universidad Autonoma de Madrid.

Atomic force microscopy (AFM) is a high-resolution technique that allows to probe single capsids. The basic set up of an AFM experiment is illustrated in Fig. 5.2a. In essence the microscope consists of a small tip ($15\,nm$ radius in our case) that is ultimately responsible for *touching* (probing) the sample. This tip is attached to a cantilever that can exert a force on the viral capsids. The position of the tip is controlled using a piezolectric device that can move the sample in the three spacial directions by an applied electric voltage. When a capsid is probed by the AFM tip, the cantilever suffers a deflection that is measured by a laser beam and translated into force units (once the system has been properly calibrated). The AFM experiments are in essence divided in two parts. First the tip scans the sample, and records a precise topographic image of a capsid (see Fig. 5.2b); subsequently, the shell can be nanoindented at a specific locations to measure its mechanical response.

More specifically, the first step of the experiment is the adsorption of the proheads on a properly modified glass surface (see Fig. 5.3b). Then using the AFM is possible to scan the surface and obtain several structural and mechanical properties of the proheads. In particular, from the topographical profiles, the distribution of heights of the adsorbed proheads is obtained, which shows two clear peaks at 42 ± 2 and $55 \pm 2nm$ (inset Fig. 5.3b). They can be ascribed to laid down on the side (29 proheads) and upright particles (28 proheads), respectively [19]. In general, to obtain topographical features of the proteinaceous structure of each isolated prohead with AFM in buffer conditions is a daunting task which strongly depends on the tip conditions, the stability of the anchorage of the particle to the substrate and the adsorption geometry. Interestingly, while none of the laid down particles showed further details beyond some triangular facets (see Fig. 5.3c), 6 out of the 28 upright proheads presented definite recognizable details corresponding to the 5-fold axial symmetry of the capped end of proheads adsorbed on the connector (see Fig. 5.3e). Likewise, we found just two upright proheads showing the connector facing up. Fig. 5.3h presents an example of a prohead absorbed through the capped end, showing a small hole surrounded by a ring (see Fig. 5.3g) that can be associated to the connector (see Fig. 5.3h). The other unidentified 20 upright proheads do not present enough clues to recognize their adsorption geometry. Therefore they could be either resting on the capped end or on the connector. For the sake of clarity, the AFM data was compared with the $\phi 29$ prohead cryo-EM volume [6] (Fig. 5.3a). Thus, the inherent geometrical dilation effect between tip and sample was considered by processing the EM data, conveniently oriented, with a dilation algorithm using a 15 nm radius tip [20]. The resulting dilated cryo-EM data (see Figures 5.3d, 5.3f, and 5.3h) present

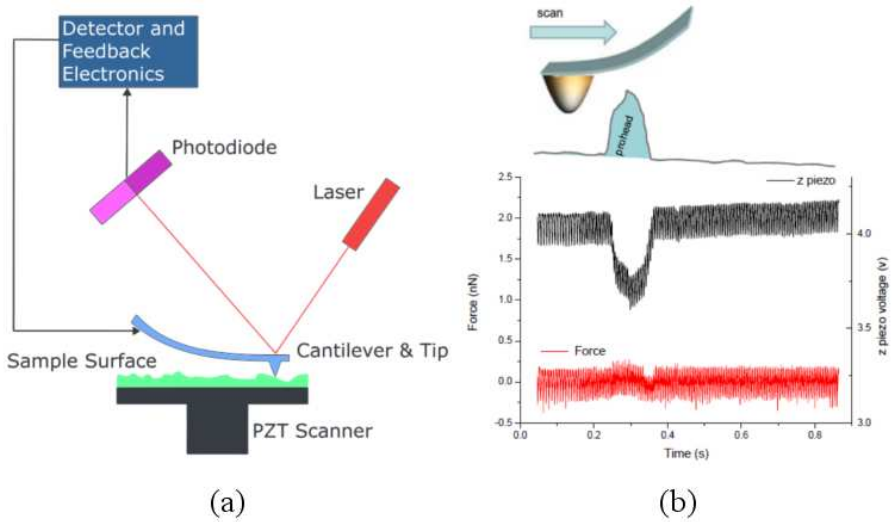


Figure 5.2: Scheme of the AFM experiment. (a) Block diagram of the atomic force microscope. The sample is mounted on a piezoelectric tube that can move the sample in the z direction using an electric voltage. The deflection of the cantilever is measured by a laser-photodiode system properly calibrated to measure the force applied by the tip. (b) The AFM can scan the surface to find the bacteriophage proheads. A jumping mode protocol is used to scan the surface keeping the force constant on average. The variations in the voltage of the piezo tracks the height of the sample. Once the profile of prohead has been identified, the AFM can nanoindent a specific point of the capsid.

topographical features which agree with the corresponding AFM images of a prohead laid on the side (see Fig. 5.3c), the 5-fold symmetry of the capped end of a particle adsorbed on the connector (see Fig. 5.3e) and the connector facing up of a prohead adsorbed on the capped end (see Fig. 5.3g).

As commented on Appendix A, to analyze the AFM experiments, we consider only proheads showing a stable mechanical behavior. Typical indentation curves performed on glass (dotted line), on laid down (gray) and on upright (dark) $\phi 29$ proheads are shown in Fig. 5.4a. In order to obtain the prohead stiffness (spring constant) k along the perpendicular direction to the substrate the recorded nanoindentation curves were fitted linearly [19]. The spring constants are sorted out depending on the selected prohead adsorption geometry [21] (upright or laid down). In the histogram of Fig. 5.4b we classified the slopes of 56 indentations carried out on 5 upright (black) and 6 laid down (gray) proheads.

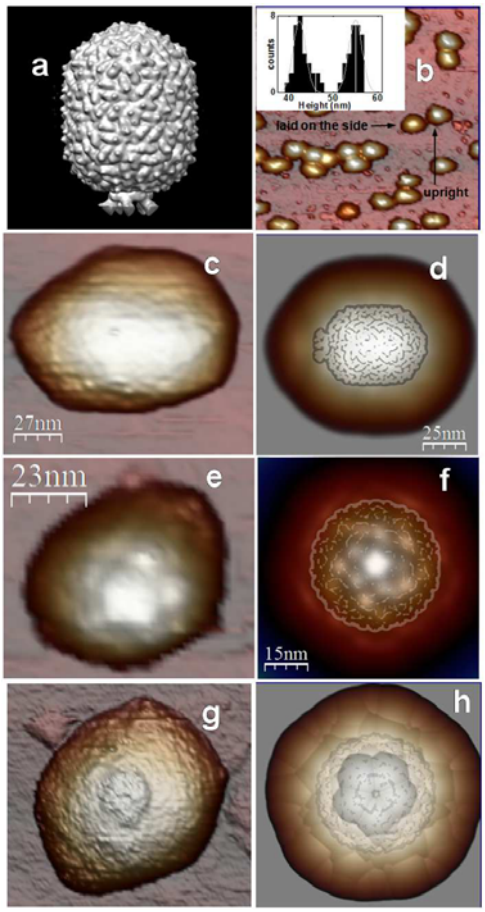


Figure 5.3: ϕ29 prohead orientation geometry. Panel a shows the EM reconstruction model EMD-1117 [6]. Panel b shows an AFM image of a typical population of ϕ29 proheads adsorbed on the silanized glass. The inset shows a histogram of heights with two peaks (see text). Panel c shows a typical AFM image of a laid down on the side prohead. Panels e and g present typical AFM images of upright proheads adsorbed by the connector and by the capped end, respectively. Panels d, f and h are the dilation images of the cryo-EM data of panel a conveniently oriented to be laid down on the side, adsorbed through the connector and through the capped end, respectively. The dilated structures compare fairly well with the corresponding AFM data. EM data are lightly superimposed on panels d, f and h.

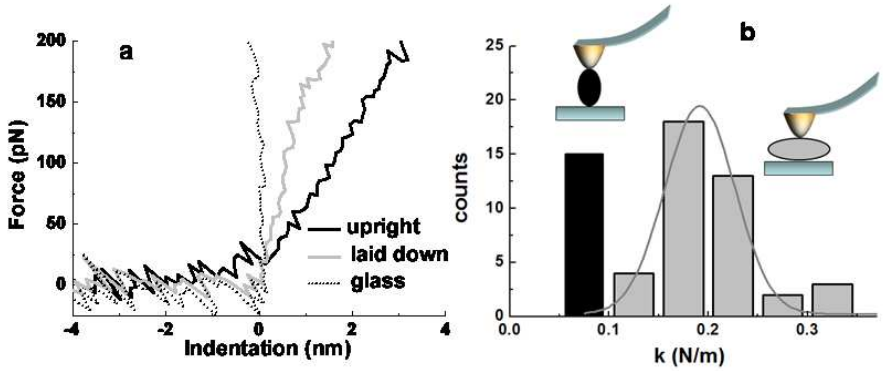


Figure 5.4: Nanoindentation experiments. Panel a shows the typical forward indentation curves on different locations: glass (dotted), laid-down shell (gray) and upright shell (black). Panel b presents the classification of the indentation curve slopes (spring constants) on upright (black) and laid-down (gray) proheads.

In the upright proheads there are included 2 adsorbed through the connector, 2 through the capped end and 1 unidentified. Gaussian fitting of the data results in spring constants of $0.075 \pm 0.020\, N/m$ and $0.192 \pm 0.034\, N/m$ for upright and laid down proheads, respectively. The spring constant of the laid down proheads corresponds to the lowest one of the two values reported in [19] within the experimental error.

Therefore, laid down proheads are $\sim 2.6 \pm 0.8$ times stiffer than upright ones. Interestingly, the spring constant of the upright proheads do not depend on the particle geometry adsorption, indicating that the position of the connector, either on top or below the particle, is not relevant for the spring constant within the experimental error (see discussion in Appendix B).

5.3 Elastic properties of empty prolate capsids

Let's analyze first the expected properties of a prolate shell based on a continuum elasticity theory description. The capsid of $\phi 29$ can be roughly considered as a shell made by a cylindrical body closed by two spherical caps. By analyzing the elastic response of the cylindrical and spherical part of the virus as independent entities one would expect that the laid-down virus would be softer than the upright one, since it is easier to deform a cylindrical than a spherical shell of the same radius [22]. The reason is that a cylindrical shell, unlike a spherical one,

can be bent without much stretching.

In a spherical shell of thickness h and radius R subjected to a concentrated force f, the stretching energy scales as $E_{str} \sim Y\, h\, (\xi/R)^2 d^2$, whereas the bending energy is $E_{bend} \sim Y\, h^3 (\xi^2/d^2)$, where ξ is the deformation, Y the 3D Young Modulus, and d the length scale over which the deformation extends. Minimization of the total energy leads to $d \sim \sqrt{hR}$, which implies that the deformation is proportional to the force with an effective spring constant [22]

$$k_{eff}^{sph} \sim Y \frac{h^2}{R} \tag{5.1}$$

However, in the indentation of a cylinder [23], the bending energy scales as $E_{bend} \sim Y\, h^3 (\xi/R^2)^2 R\, l$, and the stretching energy as $E_{str} \sim Y\, h\, (\xi R/l^2)^2 R\, l$, where now l represents the length along the axis of the cylinder over which the deformation takes place. After minimization of the total energy, one gets that the deformation zone extends over a distance $l \sim R\sqrt{R/h}$, and that again the force becomes linear in the deformation, but now with a spring constant

$$k_{eff}^{cyl} \sim Y \frac{h^{5/2}}{R^{3/2}} \tag{5.2}$$

Therefore, the ratio of spring constants is

$$\frac{k_{eff}^{sph}}{k_{eff}^{cyl}} \sim \sqrt{\frac{R}{h}} \tag{5.3}$$

In general, viral capsids can be considered as thin shells, i.e., the radius is much bigger than the thickness of the capsid, $R \gg h$. Therefore, the cylindrical spring constant measured on the cylindrical part is predicted to be softer than the spherical one

$$k_{eff}^{cyl} < k_{eff}^{sph} \tag{5.4}$$

When we particularize Eq. (5.3) to the geometry of $\phi 29$ ($R = 21\, nm$ and $h = 1.6\, nm$ [1], the ratio of effective constants becomes approximately

$$\left(\frac{k_{eff}^{sph}}{k_{eff}^{cyl}} \right)_{\phi 29} \sim 3.6 \tag{5.5}$$

So elastic theory predicts that the indentation of the cylindrical body should be more than 3 times softer than the spherical part, the opposite to what is found in the AFM experiments of $\phi 29$. Interestingly, this continuum elasticity prediction

has been quantitatively confirmed by a detailed Finite Element Analysis (FEA) of indentation of laid down on the side and upright virus-like shells with a spherical AFM tip of radius $20\,nm$ (see Appendix C).

To tackle this puzzling disagreement between the predictions of continuum elasticity theory and experiments we should account for extra considerations.

Interestingly, it has been shown experimentally that when indenting with an AFM tip on a spherical or cylindrical object whose deformation is controlled by the surface tension, the effective elastic constant is approximately equal to the membrane tension [24].

There are two plausible mechanisms that can originate this tension in the walls of the capsid. The simplest one is the presence of an internal pressure difference Δp. One should not forget that double-stranded DNA bacteriophages such as $\phi 29$ must support at least a pressure of $30 - 60\,atm$ arising from the DNA packaged inside [25–27]. Hence $\phi 29$ mature heads behave as a nanoscopic pressurized vessel. In that case, Laplace's law [28] predicts that a thin cylindrical tank with spherical caps supports a different membrane tension along the axial than along the circumferential direction. In the spherical caps and along the axial direction of the cylinder the membrane tension is $T_x = h\sigma_x = \Delta pR/2$, whereas in the circumferential (hoop) direction the tension is twice larger, i.e. $T_\theta = h\sigma_\theta = \Delta pR$, where σ_x and σ_θ are the corresponding stresses. Now, the average tension will be given by the average of the two principal components of the tension in each zone. In the cylindrical wall the average tension is

$$\widetilde{T}_{cyl} = \frac{T_x + T_\theta}{2} = \frac{3}{2} T_x \tag{5.6}$$

while for each spherical cap is given by

$$\widetilde{T}_{sph} = \frac{T_x + T_x}{2} = T_x \tag{5.7}$$

Therefore, in a pressurized vessel the cylindrical part develops a 50% more tension to counter act the pressure Δp compared to the spherical caps. Thus, in the limit of *high* pressures, where the tensions are big enough to dominate the elastic response of the shell, the effective elastic constants will be related by

$$\frac{\widetilde{k}_{eff}^{cyl}}{\widetilde{k}_{eff}^{sph}} \sim \frac{3}{2} \tag{5.8}$$

Accordingly, in this situation, the effective elastic constant measured when indenting on the cylindrical wall is expected to be 1.5 times higher than in the caps,

which is comparable to the experimental results. However since the experiments were done with empty viruses, for which the pressure difference is expected to be zero, we have to turn to a second mechanism that could play a similar role to pressure differences and justify the counterintuitive mechanical strength observed in the experiments: the presence of a built-in stress.

Some theoretical works have recently predicted that icosahedral viruses [29] and also geometrically similar nano-structures such as fullerenes and carbon nanotubes [30, 31], have a residual stress even at the optimal configuration which minimizes the overall energy (see also Chapter 4). This residual tension arises to compensate the moment of force required to bend a hexagonal sheet into a sphere or cylinder. Thus, an empty capsid can be under tension even in the absence of a pressure difference. The magnitude of this residual stress will depend, among other factors, on the bending stiffness, the radius of the capsid, and the spontaneous curvature (i.e. the preferred curvature that a monolayer of capsid proteins will assume in the absence of any external constraints, which is dictated by the protein-protein interactions).

To analyze this hypothesis, we have resorted to our coarse-grained model, as explained in the next section.

5.4 Coarse-grained simulation

In order to explore the distribution of stresses in the prolate shell of ϕ29 we performed coarse-grained simulations using the simple model of capsomer-capsomer interactions introduced in Chapter 2. In the discrete simulation, the 42 coarse-grained capsomers (i.e., 30 hexamers and 12 pentamers) of ϕ29 are placed on the surface of a spherocylinder made of a cylindrical body of length L closed by two hemispherical caps of radius R. The capsomers interact with an effective Lennard-Jones potential where we have added the bending energy of a spherocylinder [32]

$$\frac{E_{bend}}{\kappa} = 2\pi R^2 \left(\frac{2}{R} - C_0\right)^2 + \pi RL \left(\frac{1}{R} - C_0\right)^2 \tag{5.9}$$

where κ is the bending stiffness and C_0 the spontaneous curvature. Using Monte Carlo simulations we have obtained the optimal configuration and analyzed the distribution of local and global lateral stresses following the procedure detailed in [29] and in Chapter 4 (see also Appendix D).

More specifically, in the absence of bending energy the original model of capsomer-capsomer interactions had associated an optimal radius R_0 and length L_0 for the $T = 3$ $Q_{5F} = 5$ capsid of ϕ29 (given in Chapter 2). The introduction

of the bending energy term given by Eq. (5.9) shifts this optimal solution in a way that will depend on the values of κ and C_0. This shift in the values of (R, L) with respect to the optimal values (R_0, L_0) gives rise to a deformation in the shell that builds up a tension in the walls. The value of this tension can be assimilated to a "virtual" pressure (i.e., meaning by that the value of an internal pressure that will generate the same lateral stress in the absence of bending), which is not necessarily homogeneous, i.e., the local pressure in the cylindrical zone, p_{cyl}, can differ from the pressure in the caps, p_{sph}. Thus, in a first approximation, we select only the set of (R, L) capsids that are in (constrained) mechanical equilibrium, i.e., $p_{cyl} = p_{sph}$ or $\Delta p = p_{cyl} - p_{sph} = 0$. The set of these mechanically equilibrated shells defines the curve $(R, L)_{\Delta p=0}$ and can be related to the situation of a pressurized vessel (see Appendix D).

Alternatively, we also simulate the ϕ29-like capsid for different pairs of (R, L) (the details of the simulation are explained in the Appendix D), introducing a bending term that compensates the virtual pressure associated to the deformation, and leads to a non-pressurized vessel. The new optimal value $(R, L)_{C_0,\kappa}$ is the pair (R, L) that minimizes the total energy $E_{tot} = E(R, L) + E_{bend}(R, L)_{C_0,\kappa}$, where $E(R, L)$ is the intermolecular energy of interaction obtained in the simulation.

A typical result for the distribution of 2-dimensional local lateral stresses of a capsid $(R, L)_{\kappa,C_0}$ is shown in Fig. 5.5. One can get the average value of the stress in each region, i.e., the spherical caps and the cylindrical body, simply by averaging the local stresses of all capsomers belonging to that region. We have found that the average stress and tension developed on the cylindrical body is consistently higher than that on the spherical caps, for a wide region of values of the bending stiffness κ and the spontaneous curvature C_0 (see Appendix D). The ratio of the two averaged tensions oscillates roughly between 1 and 2, depending on the values of κ and C_0. Moreover, we have found that the ratio between the cap and cylindrical stresses is independent on the details of the interaction, since it is mainly dictated by the discrete arrangement of capsomers and the curvature of the structure[1]. For small values of the bending stiffness, $\kappa < 1$, we have found that, for a wide region of the spontaneous curvature C_0, the local stress of the hexamers in the cylindrical body is roughly $\sim 1.5 - 2$ times larger than that one of the pentamers at the apex of the caps (see Fig. 5.5). Moreover, we find that the ratio between the average stresses in the cylindrical body and the spherical caps is also roughly $\sim 1.5 - 2$. Interestingly, in the absence of

[1] To confirm this point we performed different sets of simulations using alternative intermolecular potentials, for instance, adding a screened electrostatic interaction.

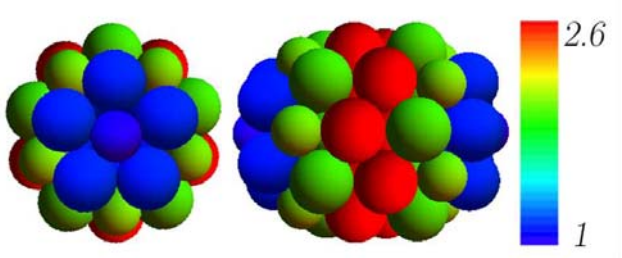


Figure 5.5: Distribution of local lateral stresses obtained in the simulation of a coarse-grained model of the $\phi 29$ capsid. The figure shows a top view (left) and a laid-down image (right) of the capsid model. Pentamers and hexamers are represented by spheres with a radius ratio ≈ 0.8. Colors indicate the relative value of the local lateral stress and the scale is normalized to the smallest value that corresponds to the tip pentamers of the caps.

bending, the isopressure line also leads to a wide range of structures with local and global ratio of stresses that are roughly $\sim 1.5-2$. Therefore, the combination of intercapsomeric interactions and bending energy associated with a preferred curvature plays a similar role as pressure, and generates a built-in lateral stress that is larger in the cylindrical body than in the spherical caps. These results corroborate the trend observed in the experimental findings[2].

5.5 Glutaraldehyde experiment

In order to further test the above mentioned hypothesis, the group of P.J. de Pablo performed further experiments using a well known cross-linking agent such as glutaraldehyde (GAD) [33] (see Appendix E).

GAD provides strength on samples, and is commonly used as a fixation agent for cells and biological tissues to tolerate the manipulations inherent for any microscopy [34]. This agent can also modify the mechanical properties of protein aggregates, because introduces covalent bonds surpassing the original inter-protein interactions. Actually, a stiffness increase introduced by GAD has been reported in microtubuli [35] and cells [36] by using AFM.

Therefore, in the case of $\phi 29$ procapsids, GAD should act as *nanostaples*

[2]Although the ratio of stresses in Fig. 5.5 matches the ratio of elastic constants measured experimentally for the $\phi 29$, our analysis is intended only to provide a qualitative description of the built-in-stress mechanism.

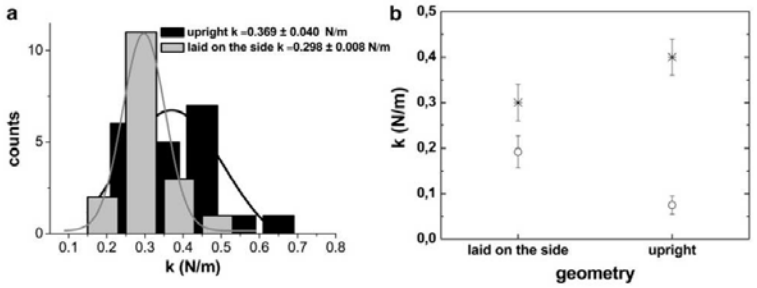


Figure 5.6: Glutaraldehyde induced mechanical reinforcement of proheads. Panel a shows the spring constant distribution of glutaraldehyde treated proheads: laid down on the side (dark) and upright (gray). Panel b presents the spring constant of the proheads with (cross) and without (circle) glutaraldehyde.

that unspecifically clamp the adjacent proteins of the shell. The new covalent bonds will reinforce and stabilize the procapsid structure, and will mask the previously preferred curvature or directionality of the interactions, which are based on weaker bonds. Assuming that the built-in stress is responsible for the anisotropy of the ϕ29 proheads, and based on the previous discussion, the reinforcement introduced by GAD should increase the intrinsic elastic constants of the shell, and hide the tension associated with the built-in stress. Consequently, this experiment should reverse the anisotropy, leading to the classical empty shell scenario.

Indentation experiments were performed on 3 upright and 3 laid down particles cross linked with GAD (see Appendix E). Interestingly, this cross-linking increases the spring constants for both upright and laid-down particles compared to the original procapsids. More importantly, the anisotropy in the elastic response has been reversed (see Fig. 5.6). In particular, the elastic constant of the cylindrical body is now ≈ 1.3 times softer than that of the caps, which is in qualitative agreement with the anisotropy predicted by the continuum elasticity theory (even though more GAD might be necessary to reach the higher values predicted by the FEA analysis).

The addition of 0.25% of GAD has been enough to mask the built-in stress, converting the pre-stressed ϕ29 procapsids into almost stress-free prolate that obey the behavior predicted by continuum elasticity theory. Hence, this experiment strongly supports the existence of built-in stress in the procapsid of ϕ29.

5.6 Conclusions

Our results reveal the strikingly anisotropic mechanical properties of ϕ29 bacteriophage prohead virus particles that seem to be due to the existence of pre-stresses in the capsid. The presence of these residual stresses might play an important biophysical role for the viability of the virus. Induced pre-stress is commonly used not only in material science, but also by many biological tissues (although not molecular aggregates so far) to increase their mechanical resistance and help them to better tolerate tension [37]. The presence of pre-stress may stiffen the capsid to prevent any damage from mechanical assaults such as osmotic shocks or DNA packing, providing protection to the viral genome during the extracellular virus cycle. On the other hand, the presence of this residual stress could play a significant role in the pressure-assisted DNA translocation (ejection) through the tail into the host [2] performed by the mature head. A simple estimation using the Laplace law[3] indicates that the measured stress of the equatorial zone is equivalent to a pressure difference of $0.19\,N/m/21\,nm \simeq 90\,atm$, the same order of magnitude than that exerted by the confined DNA [3, 25, 27]. Therefore the built-in stress could confine the internal pressure along the long axis of the prolated shell, helping to make the most of the stored elastic energy to initiate the DNA translocation process trough the tail.

Another biophysical interesting question relates to how this pre-stress is generated during the capsid formation. In the assembly process there is a competition between the tendency of proteins to aggregate at the preferred curvature and the need to minimize the rim area exposed in a partially assembled capsid by making a closed structure [38] (see Chapters 6 and 7). If the curvature of the closed capsid is different from the spontaneous one (i.e. the preferred curvature at which capsid proteins will assemble without forcing them to make a closed structure), lateral stress will develop. Interestingly, not only ϕ29, but most complex double stranded DNA bacteriophages do have a scaffolding protein which co-assembles with the main head protein to produce a prohead with the correct shape and size [39]. In the case of ϕ29, the presence of these scaffolding proteins is mandatory for a correct prohead assembly [11]. This directing role of the scaffold is transient, as once the shell is built, the scaffolding is released concomitant to the DNA packing in the shell. We can not exclude the possibility that a certain amount of scaffolding protein might be released from proheads during storage, but this is not followed by any apparent change in the structure of the proheads. Since

[3] Assuming the elastic constant k_{cyl} as the membrane tension for the cylindrical part, one gets the expression $k_{cyl}/R = \Delta p$, where R is the radius of the prolate.

protein binding interactions are relatively weak, on the order of a few kT [40], it is likely that scaffolding proteins might help to impose a curvature in the capsid significantly different from the spontaneous one, by assisting the bending of the proteins at the junctions. This will generate a much larger "membrane" stress, which will reinforce the strength of the capsid. In the absence of the scaffolding, the stress generated during the assembly might help to better tolerate the packaging, avoiding a significant expansion.

Appendices

A AFM of proheads

The nanoindentation experiments on single procapids were performed by the group of P.J. de Pablo at the Universidad Autonoma de Madrid.

Stocks of empty proheads of $\phi 29$ were stored in TMS buffer ($137\,mM$ NaCl, $2.7\,mM$ KCl, $1.5\,mM$ NaH2PO4, $8.1\,mM$ KH2PO4, pH 7.2). A single drop of 20 μl stock solution capsid was deposited on a silanized glass surface [41], which was left for 30 min on the surface and washed with buffer. The tip was pre-wetted with $20\mu l$ of buffer. The AFM (Nanotec Electrónica S.L., Madrid, Spain) was operated in jumping mode in liquid [42] using rectangular cantilevers RC800PSA, and Biolevers (BL-AC40TS) (Olympus, Tokyo, Japan) with nominal spring constants of $0.05\,N/m$ and $0.03\,N/m$, respectively. Cantilevers spring constants were routinely calibrated by using the Sader's method [43]. In order to perform nanoindentations, single proheads were deformed by the tip by carrying out single FZs (force vs. z-piezo displacement) experiments most likely right at the top of the procapsids: the particle is zoomed in continuously by reducing the x-y scanning size until the bump of the very top is under the whole piezo scan. Afterwards the FZ is executed at the top of the particle, probably with a few nm of uncertainty mainly provoked by the thermal drift, and the intrinsic non-linearity and creep of the piezo. Still, this method has been proven to be robust enough to establish electrical contact with carbon nanotubes [44], which are even smaller than viral particles. During the first stages of indentation the capsids show a linear deformation [23] which provides the spring constant of the virus k_v (if it is considered like a spring in series with the cantilever) as $k_v = k_c S_g/(S_v - S_g)$, where k_c is the spring constant of the cantilever, S_g (nm/V) the slope of the cantilever deflection on the glass (here the substrate is considered as non deformable) and S_v (nm/V) the slope of the cantilever deflection on the virus. Hence, the prohead is indented with a few separate sets of FZs of about 5 indentations in each one. Following each FZ set, an image of the prohead is taken to confirm the integrity of the prohead, as well as to know its position in order to correct for any drift if needed to perform the next FZs set. The maximum force applied during each FZ never exceeded $\sim 300\,pN$ to prevent the damage [45], collapse [23], buckling [46], or non-linear deformations [47] of the shells. The FZ speed is about $60\,nm/s$. Even if the shell integrity is maintained, in our experiments only proheads showing stable spring constants along the FZs sets were considered to avoid particle mobility effects that often occur when the

particle is loosely bound to the surface. Stiffness images were also performed [48], which roughly consist on keeping the contact slopes of all the FZ curves performed at each point of the topography image. Although this method provides a very accurate relationship of the topography and the stiffness of the virus [49], it is slower than single FZs and requires exceptional imaging conditions such as robust virus anchorage and thermal drift stability. Viruses are also very prone to destruction when using this method since, unlike a single FZ, the virus is repeatedly indented about a few thousand times in a 128×128 points image. Images were processed using the WSxM software [21]. In order to reduce the thermal drift of the AFM, it is operated under constant temperature ($21^0 C$) conditions inside an acclimatized box.

B The influence of the connector

In order to clarify the potential influence of the connector, a stiffness map of one of the two upright proheads showing the connector facing up was performed. After carrying out the usual single FZ experiments to find the prohead spring constant ($0.076 \pm 0.020\, N/m$), we succeeded in acquiring a stiffness map (see Fig. B.1b), simultaneously with the topographical image (see Fig. B.1a). The histogram of the stiffness slopes presented in Fig. B.1c shows two peaks at $0.055 \pm 0.007\, V/nm$ and at $0.075 \pm 0.005\, V/nm$ that correspond to the prohead and the substrate, respectively.

From these data, the spring constant of the prohead is calculated to be $0.074 \pm 0.020\, N/m$, very close to that obtained with the single FZ experiments. An interesting detail of the stiffness map is the existence of a soft area around the center of the prohead. Fig. B.1d shows two simultaneous profiles of the topography (gray) and the stiffness map (dark) carried out on this region, providing a direct correlation of the low stiffness area with part of the ring of the connector. Therefore, it is tempting to ascribe $0.04\, N/m$ to the connector spring constant, which is almost half of the adjacent shell area stiffness ($\sim 0.07\, N/m$, see Fig. B.1d). Nevertheless, despite the fact that the connector is locally softer, its presence in the capsid replacing one of the pentamers seems not to have any influence on the spring constant of the whole capsid.

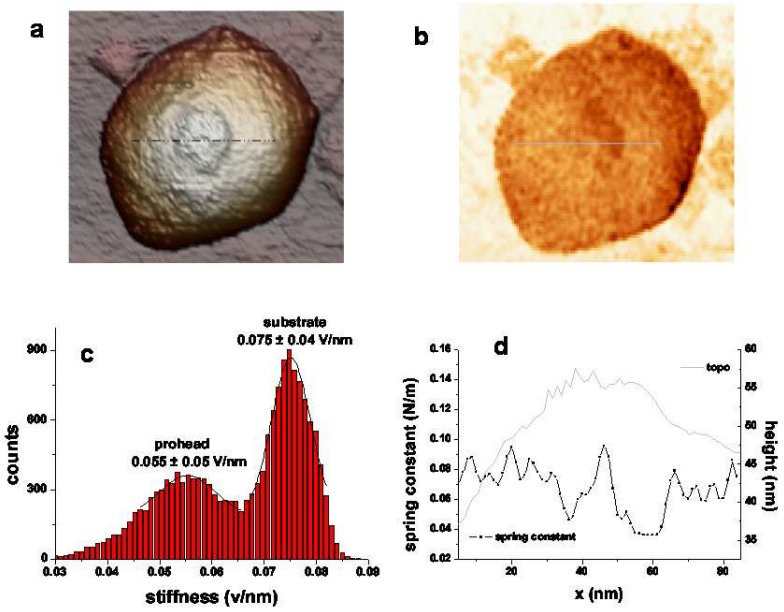


Figure B.1: Stiffness map of a prohead showing the connector. Panel a shows the topography AFM data of the connector and panel b the simultaneously acquired stiffness map. Panel c presents the histogram of the stiffness map. Panel d presents a topographical profile showing the connector (gray) and the corresponding stiffness profile (dotted dark).

C FEA analysis

In order to study the quantitative mechanical response of a continuum prolate shell, a Finite Element Analysis (FEA) was performed by A. Raman, J. Gómez Herrero and I. A. T. Schaap.

The model considers the icosahedral geometry of the prohead emphasizing the faceting of the viral particles. These simulations were performed using iANSYS. The model consists of three volumes: the capsid is modeled as a hollow icosahedron with an extra band of hexons around the equator, the tip is modeled as a rigid sphere of diameter $20\,nm$, and the substrate beneath the capsid is modeled as a flat, rigid surface. Prohead dimension are $2\,nm$ thickness, $40\,nm \times 54\,nm$ [1]. All three elements exhibit planar symmetry, so the model consists of only half the domain. Symmetry boundary conditions, which force the derivatives of the solution at the symmetry plane to zero, ensure that no features of the solution are lost. The capsid was deformed up to $20\,nm$ indentation depth in increments of $0.5\,nm$. The solution was then mirrored about the symmetry plane to obtain the complete solution. The dimensions of the prohead were taken from cryo-EM [1]. The elastic modulus was taken from [19] to be $1.8\,GPa$. The model is meshed using approximately $12,000$ SOLID92 elements, which are 10-node quadratic tetrahedra. Contact between the tip and capsid and between the capsid and substrate is captured using contact pairs with CONTAC174 and TARGE170 elements. These elements are planar and share nodes with the volume mesh at the contact surfaces. Symmetry boundary conditions were applied at the areas that are coincident with the symmetry plane and displacement boundary conditions on the tip nodes as well as the substrate nodes. The implementation of contact within ANSYS requires the specification of a friction coefficient. A detailed convergence study of friction coefficients suggested that a value of 0.2 captures buckling phenomena of individual facets of the capsid, which was a desirable component of that model. Also, based on a detailed mesh density convergence study $\sim 12,000$ elements were used. The model was computed with the parameters in Table C.1. Images of the elements and representative force-indentation curves are shown in Fig. C.1.

Element Type	SOLIDS (tetr 10-node)
Number of elements	12,832 (top) 11,582 (side)
Young's modulus	$1.8\,GPa$
Poisson's ratio	0.3
Friction coefficient	0.20
Contact pair elements	CONTACT174 and TARGE170
Basic options	Auto time stepping 25 substeps 100 max. substeps 15 min. substeps
Nonlinear options	500 max. iterations

Table C.1: Parameters of the FEA model.

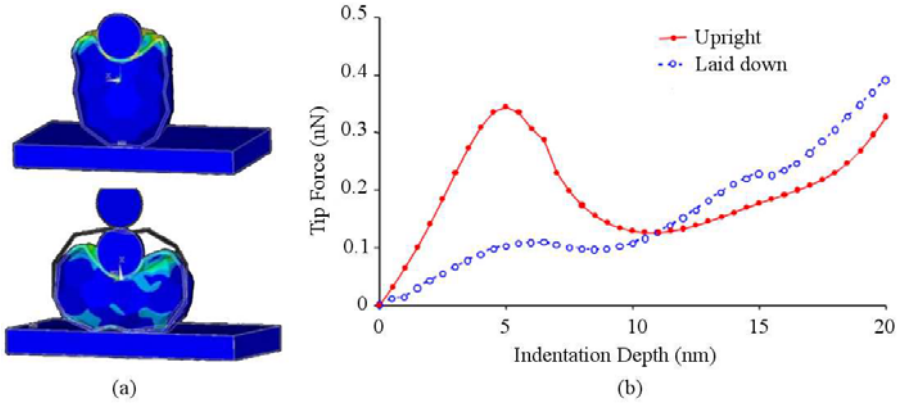


Figure C.1: FEA analysis of $\phi 29$ mechanical response. (a) The meshing for upright (top) and laid down (bottom) proheads. (c) Force indentation curves predicted by the computational model. Note that only the data in the elastic regime prior to buckling and nonlinear effects are relevant to the discussion in this chapter.

D Details of the coarse-grained simulations

The ϕ29-like capsid, $N = 42$ capsomers with $T = 3$ $Q = 5$ (5-fold) (see Chapter 1), was simulated using the Monte Carlo (MC) algorithm, introduced in Chapter 2, for different pairs of radii and body lengths (R, L) around the optimal structure (R_0, L_0). Each shell was equilibrated by $5,000$ MC steps and the magnitudes were averaged using $5,000$ MC steps more[4]. Importantly, for each deformation we computed the virtual pressure at the cylindrical body, p_{cyl}, and at the spherical caps, $p_{sph} = (p_{sph}^{left} + p_{sph}^{right})/2$. The virial pressure of a zone was computed by summing the radial force on each capsomer due the intermolecular interactions divided by the surface of the zone. Subsequently we selected the deformations that showed constrained mechanical equilibrium, where the virtual pressure was homogeneous, i.e., $\Delta p = p_{cyl} - p_{sph} = 0$. This defines the *isopressure* curve $(R, L)_{\Delta p=0}$ (see Fig. D.1).

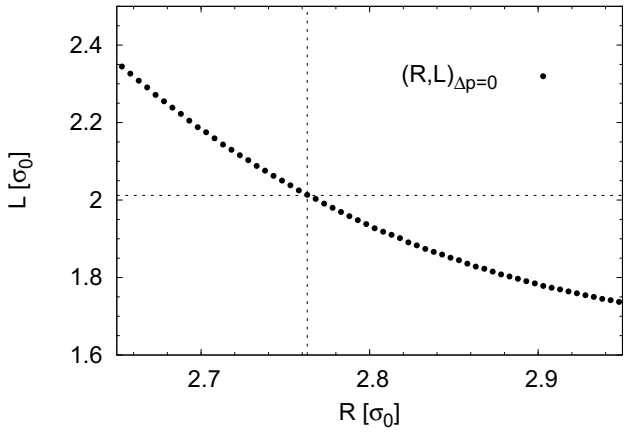
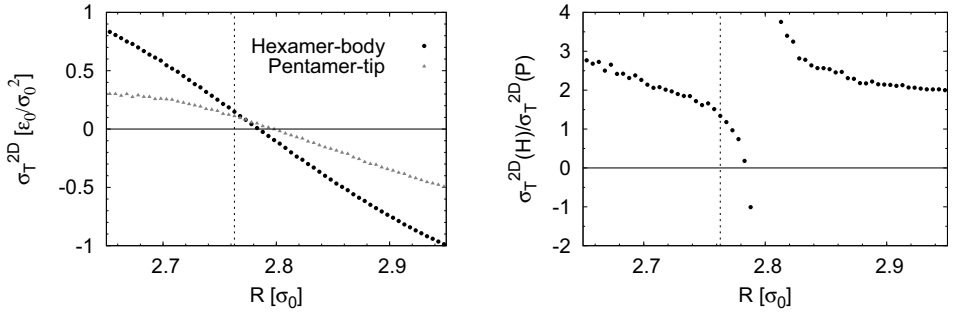


Figure D.1: Isopressure curve. The set of values (R, L) associated to the ϕ29-like capsid that are in (constrained) mechanical equilibrium $\Delta p = 0$ are plotted. Each structure was equilibrated by a MC algorithm. The dotted vertical and horizontal lines indicate intersect at the optimal structure $(R, L)_{p=0}$ or (R_0, L_0).

For every structure the local and global 2D stresses were computed (see Chapter 4). Special emphasis was placed on the study of the lateral stresses,

[4]A subset of the (R, L) map was explored using longer simulations, and agreed with the shorter ones. Note that here the initial configuration for each simulation was a (deformed) $T = 3$ $Q = 5$, rather than a random structure.

which can be related to a membrane tension [29], and that can be even measured by AFM experiments [24]. Note that for the hemi-spherical caps the local (or global) lateral stress is given by the average of the azimuthal and hoop stresses, $\sigma_T = (\sigma_{\theta\theta} + \sigma_{\phi\phi})/2$, whereas in the cylindrical zone the lateral stress corresponds to the average of the longitudinal and hoop stresses, $\sigma_T = (\sigma_{zz} + \sigma_{\phi\phi})/2$. Fig. D.2a plots the local values of lateral stresses for a hexamer at the center of the cylindrical body $\sigma_T^{2D}(H)$, and a pentamer on the tip of a spherical cap, $\sigma_T^{2D}(P)$. The stress in the hexamer is systematically bigger in magnitude than the local stress in the pentamer, except in the region close to the optimal structure R_0, due to the non-simultaneous change of sign for both capsomers. Recall that even at equilibrium, the discrete nature and curvature of the capsid has associated a residual stress (see Chapter 4 and Ref. [29]). The ratio of the hexamer and pentamer stresses is plotted in Fig. D.2b. The lateral stress in the pentamer is in general ~ 2 times larger than that of the pentamer. For compressed structures $R < R_0$ the ratio of local stresses increases and gets closer to ~ 3, and near R_0 the ratio diverges due to the fact that the stress on the pentamer becomes vanishingly small (see Fig. D.2a). We emphasize that the global lateral stresses of the spherical and cylindrical zones follow a similar behavior.



(a) (b)

Figure D.2: Local lateral stresses following the isopressure line. (a) Lateral stress of a hexamer at the center of the body of the ϕ29-like capsid, and lateral stress of the pentamer at the tip of a spherical cap. The vertical line is placed at R_0. Note that at the optimal equilibrium the residual stress is low but non-zero. Positive stress means compression, and negative means stretching. (b) Ratio of the lateral stresses plotted in panel (a). The hexamer stress is generally larger, but the ratio diverges at $R \gtrsim R_0$ because the lateral stress of the pentamer inverts the sign and becomes zero. The vertical line is placed again at R_0.

The existence of a bending energy can compensate the virtual (inhomogeneous) pressure generated in the deformed structures of a ϕ29-like capsids, and lead to highly stressed but non-pressurized vessels. In a first approximation we introduce a continuum bending term for spherocylindrical shells given by Eq. (5.9). We studied the influence of different values for the bending constant κ and the spontaneous curvature C_0 (see Fig. D.3). The total energy given by the intermolecular energy $E(R, L)$ and the bending energy $E_{bend}(R, L)_{C_0,\kappa}$, i.e., $E_{tot} = E(R, L) + E_{bend}(R, L)_{C_0,\kappa}$, was computed and minimized using the simulated structures for different values of (R, L). Interestingly, for low values of κ there is a prominent zone where the ratio of the residual stress for a hexamer (at the equator of the body) and a pentamer (at the tip of a cap) is ~ 2 (see Fig. D.3a). When one plots the stress ratio between the cylindrical body an a spherical cap, the zone for > 1 is even wider (see Fig. D.3b).

Both mechanisms, the pressurization and bending energy, lead to a similar qualitative scenario with respect to the residual stress distribution of the prolate, and are in consonance with the elastic response measured in the AFM experiments.

We emphasize that our analysis of the built-in stress is a first approximation, due to the simplicity of our intermolecular interaction, the geometrical constrain (the spherocylindrical template), and the application of a continuum bending energy term. To obtain a more accurate quantitative analysis of this reinforcement mechanism, a more realistic model, including a intermolecular term related to the orientation preference between capsomers, should be investigated.

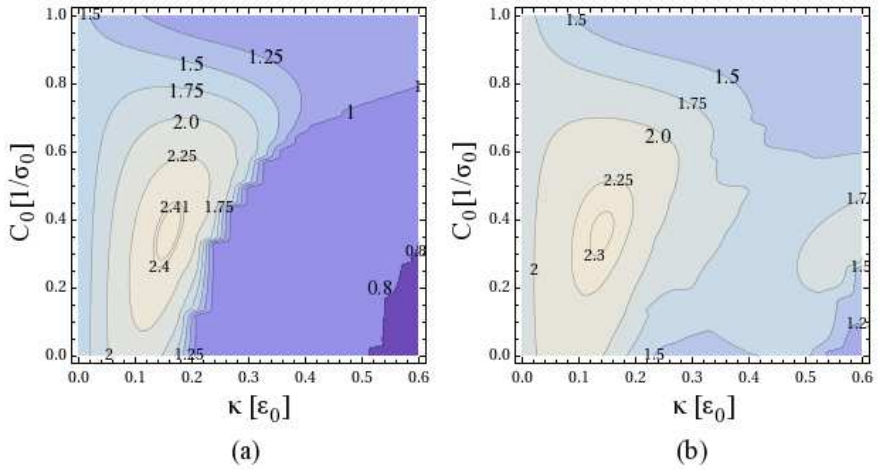


Figure D.3: Map of the ratio of lateral stresses as a function of the parameters of the bending energy. The total energy (intermolecular interactions plus bending energy) was minimized for different values of the spontaneous curvature, C_0, and bending constant, κ, using the simulated structures of the Lennard-Jones like model for different values of (R, L). (a) The local ratio of the stress in a hexamer in the body with respect to the penton at the tip of the cap for the different equilibrium structures is plotted. (b) The global ratio of the stress in the cylindrical body of the prolate with respect to the hemi-spherical cap for the different equilibrium structures is plotted.

E Glutaraldheyde experiments

Glutaraldehyde (GAD) is a fairly small molecule with two aldehyde groups, separated by a flexible chain of 3 methylene bridges. The potential for cross-linking is given by the -CHO groups and happens over variable distances since in aqueous solutions, GAD is present largely as polymers of variable size. There is a free aldehyde group sticking out of the side of each unit of the polymer molecule, as well as one at each end. All these -CHO groups will combine with any protein nitrogens of the free amino groups with which they come into contact, forming covalent bonds and cross-linking adjacent proteins. GAD was added to the prohead solution up to 0.25% and incubated for half an hour. Afterward the procapsids solution was adsorbed on silanized glass as indicated above and washed out with TMS buffer.

References

[1] Y. Tao, N. H. Olson, W. Xu, D. L. Anderson, M. G. Rossmann, and T. S. Baker, "Assembly of a tailed bacterial virus and its genome release studied in three dimensions," *Cell*, vol. 95, pp. 431–7, 1998.

[2] V. Gonzalez-Huici, M. Salas, and J. M. Hermoso, "The push-pull mechanism of bacteriophage ϕ29 DNA injection," *Mol. Microbiol.*, vol. 52, pp. 529–540, 2004.

[3] Y. R. Chemla, K. Aathavan, J. Michaelis, S. Grimes, P. J. Jardine, D. L. Anderson, and C. Bustamante, "Mechanism of force generation of a viral dna packaging motor," *Cell*, vol. 122, pp. 683–692, 2005.

[4] W. M. Gelbart and C. M. Knobler, "Pressurized viruses," *Science*, vol. 323, pp. 1682–1683, 2009.

[5] P. Grayson and I. J. Molineux, "Is phage DNA 'injected' into cells–biologists and physicists can agree," *Curr. Opin. Microbiol.*, vol. 10, pp. 401–9, 2007.

[6] M. C. Morais, K. H. Choi, J. S. Koti, P. R. Chipman, D. L. Anderson, and M. G. Rossmann, "Conservation of the capsid structure in tailed dsDNA bacteriophages: the pseudoatomic structure of phi 29," *Mol. Cell*, vol. 18, pp. 149–159, 2005.

[7] Y. Xiang and M. G. Rossmann, "Structure of bacteriophage ϕ29 head fibers has a supercoiled triple repeating helix-turn-helix motif," *Proc. Natl. Acad. Sci. USA.*, vol. 2011, pp. 1–5, 2011.

[8] B. Ibarra, J. R. Caston, O. Llorca, M. Valle, J. M. Valpuesta, and J. L. Carrascosa, "Topology of the components of the DNA packaging machinery in the phage ϕ29 prohead," *J. Mol. Biol.*, vol. 298, pp. 807–815, 2000.

[9] W. R. Wikoff and J. E. Johnson, "Virus assembly: Imaging a molecular machine," *Curr. Biol.*, vol. 9, pp. R296–R300, 1999.

[10] A. Camacho, F. Jimenez, J. Delatorre, J. L. Carrascosa, R. P. Mellado, C. Vasquez, E. Vinuela, and M. Salas, "Assembly of bacillus-subtilis phage-phi-29. 1. Mutants in cistrons coding for structural proteins," *Eur. J. Biochem.*, vol. 73, pp. 39–55, 1977.

[11] K. H. Choi, M. C. Morais, D. L. Anderson, and M. G. Rossmann, "Determinants of bacteriophage phi 29 head morphology," *Structure*, vol. 14, pp. 1723–1727, 2006.

[12] P. X. Guo, S. Erickson, W. Xu, N. Olson, T. S. Baker, and D. Anderson, "Regulation of the phage-29 prohead shape and size by the portal vertex," *Virology*, vol. 183, pp. 366–373, 1991.

[13] M. C. Morais, S. Kanamaru, M. O. Badasso, J. S. Koti, B. A. L. Owen, C. T. McMurray, D. L. Anderson, and M. G. Rossmann, "Bacteriophage phi 29 scaffolding protein gp7 before and after prohead assembly," *Nat. Struct. Biol.*, vol. 10, pp. 572–576, 2003.

[14] A. C. Steven, J. B. Heymann, N. Cheng, B. L. Trus, and J. F. Conway, "Virus maturation: dynamics and mechanism of a stabilizing structural transition that leads to infectivity," *Curr. Opin. Struc. Biol.*, vol. 15, pp. 227–36, 2005.

[15] J. E. Johnson, "Virus particle maturation: insights into elegantly programmed nanomachines," *Curr. Opin. Struc. Biol.*, vol. 20, pp. 210–216, 2010.

[16] Y. Xiang, M. C. Morais, A. J. Battisti, S. Grimes, P. J. Jardine, D. L. Anderson, and M. G. Rossmann, "Structural changes of bacteriophage phi 29 upon dna packaging and release," *EMBO J.*, vol. 25, pp. 5229–5239, 2006.

[17] D. Panja and I. J. Molineux, "Dynamics of bacteriophage genome ejection in vitro and in vivo," *Phys. Biol.*, vol. 7, p. 045006, 2010.

[18] P. Grayson, A. Evilevitch, M. M. Inamdar, P. K. Purohit, W. M. Gelbart, C. M. Knobler, and R. Phillips, "The effect of genome length on ejection forces in bacteriophage lambda," *Virology*, vol. 348, pp. 430–6, 2006.

[19] I. L. Ivanovska, P. J. de Pablo, B. Ibarra, G. Sgalari, F. C. MacKintosh, J. L. Carrascosa, C. F. Schmidt, and G. J. L. Wuite, "Bacteriophage capsids: tough nanoshells with complex elastic properties," *Proc. Natl. Acad. Sci. USA.*, vol. 101, pp. 7600–7605, 2004.

[20] I. Horcas, R. Fernandez, J. M. Gomez-Rodriguez, J. Colchero, J. Gomez-Herrero, and A. M. Baro, "Wsxm: A software for scanning probe microscopy and a tool for nanotechnology," *Rev. Sci. Instrum.*, vol. 78, p. 013705, 2007.

[21] C. Carrasco, A. Carreira, I. A. T. Schaap, P. A. Serena, J. G.-H. M. G., Mateu, and P. J. de Pablo, "DNA-mediated anisotropic mechanical reinforcement of a virus," *Proc. Natl. Acad. Sci. USA.*, vol. 103, pp. 13706–11, 2006.

[22] L. D. Landau and I. M. Lifshitz., *Theory of elasticity.* Pergamon, London, 1986.

[23] P. J. de Pablo, I. A. T. Schaap, F. C. MacKintosh, and C. F. Schmidt, "Deformation and collapse of microtubules on the nanometer scale," *Phys. Rev. Lett.*, vol. 91, p. 98101, 2003.

[24] M. Arnoldi, M. Fritz, E. Bauerlein, M. Radmacher, E. Sackmann, and A. Boulbitch, "Bacterial turgor pressure can be measured by atomic force microscopy," *Phys. Rev. E*, vol. 62, pp. 1034–1044, 2000.

[25] A. Evilevitch, L. Lavelle, C. M. Knobler, E. Raspaud, and W. M. Gelbart, "Osmotic pressure inhibition of dna ejection from phage," *Proc. Natl. Acad. Sci. USA*, vol. 100, pp. 9292–9295, 2003.

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

[27] S. Koster, A. Evilevitch, M. Jeembaeva, and D. A. Weitz, "Influence of internal capsid pressure on viral infection by phage lambda," *Biophys. J.*, vol. 97, pp. 1525–1529, 2009.

[28] P.-G. de Gennes, F. Brochard-Wyart, and D. Quere, *Capillarity and wetting phenomena: drops, bubbles, pearls, waves*. Springer, 2003.

[29] R. Zandi and D. Reguera, "Mechanical properties of viral capsids," *Phys. Rev. E*, vol. 72, p. 021917, 2005.

[30] D. H. Robertson, D. W. Brenner, and J. W. Mintmire, "Energetics of nanoscale graphitic tubules," *Phys. Rev. B*, vol. 45, pp. 12592–12595, 1992.

[31] N. Chandra, S. Namilae, and C. Shet, "Local elastic properties of carbon nanotubes in the presence of stone-wales defects," *Phys. Rev. B*, vol. 69, p. 094101, 2004.

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

[33] J. A. Kiernan, "Formaldehyde, formalin, paraformaldehyde and glutaraldehyde: What they are and what they do," *Microscopy Today*, vol. 8, pp. 1–3, 2000.

[34] B. Alberts, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular biology of the cell*. Garland, New York, 2002.

[35] A. Vinckier, I. Heyvaert, A. Dhoore, T. Mckittrick, C. Vanhaesendonck, Y. Engelborghs, and L. Hellemans, "Immobilizing and imaging microtubules by atomic-force microscopy," *Ultramicroscopy*, vol. 57, pp. 337–343, 1995.

[36] J. L. Hutter, J. Chen, W. K. Wan, S. Uniyal, M. Leabu, and B. M. C. Chan, "Atomic force microscopy investigation of the dependence of cellular elastic moduli on glutaraldehyde fixation," *J. Microsc. (Oxf.)*, vol. 219, pp. 61–68, 2005.

[37] C. Storm, J. J. Pastore, F. C. MacKintosh, T. C. Lubensky, and P. A. Janmey, "Nonlinear elasticity in biological gels," *Nature*, vol. 435, pp. 191–194, 2005.

[38] R. Zandi, P. van der Schoot, D. Reguera, W. Kegel, and H. Reiss, "Classical nucleation theory of virus capsids," *Biophys. J.*, vol. 90, pp. 1939–1948, 2006.

[39] T. Dokland, "Scaffolding proteins and their role in viral assembly," *Cell. Mol. Life Sci.*, vol. 56, pp. 580–603, 1999.

[40] A. Zlotnick, "Are weak protein-protein interactions the general rule in capsid assembly?," *Virology*, vol. 315, pp. 269–274, 2003.

[41] J. P. Michel, I. L. Ivanovska, M. M. Gibbons, W. S. Klug, C. M. Knobler, G. J. L. Wuite, and C. F. Schmidt, "Nanoindentation studies of full and empty viral capsids and the effects of capsid protein mutations on elasticity and strength," *Proc. Natl. Acad. Sci. USA.*, vol. 103, pp. 6184–6189, 2006.

[42] F. Moreno-Herrero, P. J. de Pablo, M. Alvarez, J. Colchero, J. Gomez-Hertero, and A. M. Baro, "Jumping mode scanning force microscopy: a suitable technique for imaging DNA in liquids," *Appl. Surf. Sci.*, vol. 210, pp. 22–26, 2003.

[43] J. E. Sader, J. W. M. Chon, and P. Mulvaney, "Calibration of rectangular atomic force microscope cantilevers," *Rev. Sci. Instrum.*, vol. 70, pp. 3967–3969, 1999.

[44] P. J. de Pablo, C. Gomez-Navarro, M. T. Martinez, A. M. Benito, W. K. Maser, J. Colchero, J. Gomez-Herrero, and A. M. Baro, "Performing current versus voltage measurements of single-walled carbon nanotubes using scanning force microscopy," *Appl. Phys. Lett.*, vol. 80, pp. 1462–1464, 2002.

[45] W. Klug, R. Bruinsma, J.-P. Michel, C. Knobler, I. Ivanovska, C. Schmidt, and G. Wuite, "Failure of viral shells," *Phys. Rev. Lett.*, vol. 97, p. 228101, 2006.

[46] G. A. Vliegenthart and G. Gompper, "Mechanical deformation of spherical viruses with icosahedral symmetry," *Biophys. J.*, vol. 91, pp. 834–41, 2006.

[47] M. Zink and H. Grubmuller, "Mechanical properties of the icosahedral shell of southern bean mosaic virus: A molecular dynamics study," *Biophys. J.*, vol. 96, pp. 1350–1363, 2009.

[48] E. A-Hassan, W. F. Heinz, M. D. Antonik, N. P. D'Costa, S. Nageswaran, C. A. Schoenenberger, and J. H. Hoh, "Relative microelastic mapping of living cells by atomic force microscopy," *Biophys. J.*, vol. 74, pp. 1564–1578, 1998.

[49] I. A. T. Schaap, C. Carrasco, P. J. de Pablo, F. C. MacKintosh, and C. F. Schmidt, "Elastic response, buckling, and instability of microtubules under radial indentation," *Biophys. J.*, vol. 91, pp. 1521–1531, 2006.

Part III

Self-Assembly of Viral Capsids

Chapter 6

Classical Nucleation Theory of Spherical Shells

6.1 Introduction

Viruses use several mechanisms to infect a cell, which can be very specific depending on the viral species and its host [1]. After the infection and at certain cellular conditions, the replication is triggered, and the viral genetic material takes advantage of the cell machinery to produce the required components for a new viral progeny, i.e., coat proteins, genetic material, and other products. This can be a very complex process with some similarities to an industrial factory [2], where the place and rhythm of the events must be properly orchestrated to produce a mature virus. Therefore, understanding the role of the different steps during viral assembly is a key question to identify targets that could control a virus infection [3, 4].

Depending on the virus, the capsid is successfully assembled from the coat proteins alone [5–16], or in combination with scaffolding proteins [17–20], or the genetic material [21–23]. However, regardless of the scenario, the process of capsid formation is spontaneous and does not require energy consumption (in form of ATP). This suggests that the assembly of viruses is controlled by general and basic physical principles, and thus its amenable to a thermodynamic and kinetic description [24–28]. In fact, many viruses can be reconstituted *in vitro* from their basic components, yielding virus like particles which are fully infective and indistinguishable from native virions. For instance, back in the forties, tobacco mosaic virus was the first virus to be fully reconstituted *in vitro* by mixing the coat proteins and the viral RNA [21].

Also, empty capsids have been produced even for viruses that seem to require the genetic material to assemble *in vivo*, such as CCMV or BMV. Moreover, it has been shown that by tuning the coat protein concentrations, the pH, or the ionic strength conditions the same protein is able to assemble into different capsid structures [5, 6, 9, 11, 29–31]. This has opened also the possibility to engineer the viral coat proteins for different type of applications. For instance, encapsulation of materials for biomedical or material purposes or for templating nanostructures [32].

All this has stirred the interest in studying the mechanisms of viral assembly. To elucidate the basic aspects of the formation of well ordered spherical capsids, typical experiments are based on scattering techniques, cryo-EM, and size exclusion chromatography (SEC), combined with simple models [7, 33]. The assembly process of spherical shells has been studied also from theory [26, 27, 34–38] and simulations [28, 39–54]. In particular, the work of Zlotnick and collaborators has been specially relevant in this context. Their *in vitro* experiments on HBV [13,55] and papilloma viruses [14], combined with his seminal kinetic rate assembly models [26,56,57], and complemented also with simulations [58,59] have provided very useful insights on the process of viral assembly. Nevertheless, and despite the efforts of the community, many questions regarding the formation of shells, and specially its kinetics, still remain open.

Viruses have developed specific mechanisms to succeed in the infection of a particular host. However, there are many evidences from experiments and simulations that unveil the presence of some characteristic and generic features in the assembly of viral capsids. First, there is a steep dependence of capsid production on protein concentration. Below a critical (or characteristic) concentration no capsids can be built, whereas for concentrations larger than that, the yield of capsids is high. Another distinguishing feature is the lack of intermediates: either subunits in solution or fully formed capsids are typically observed in experiments. On what concerns the kinetics, it is characterized by a marked sigmoidal kinetics and the existence of a lag time before the formation of the first capsids takes place. A final trend is that there is a pronounced hysteresis between the conditions at which capsid assembly and capsid disassembly take place. Once capsids are formed, they do not dissociate easily even if the concentration of free subunits is drastically reduced. This has a clear biological relevance, since viruses must resist a number of different environments to reach a new host, where no viral proteins are present.

These distinctive features suggest a clear mechanism that could explain all the above observations: nucleation. Hence the assembly of viruses shares some

similarities with a classical first order phase transition, like the formation of a crystal or a liquid phase from vapor. The distinguishing trend of all these phenomena is their activated nature, i.e., the presence of an energetic barrier that has to be surmounted for the formation of the new phase, e.g., a fully formed capsid in the viral case. The origin of the barrier comes from the energetic cost involved in the formation of a partial capsid. In favorable assembly conditions, the free energy of binding between proteins, due to hydrophobic, electrostatic, and other specific interactions, will drive the aggregation process overcoming the lost of configurational entropy of the free subunits when they form a capsid. However, during the process of capsid formation the subunits in the rim of a partially formed capsid will lack some of the contacts present in the final structure. This introduces an energetic cost that is responsible of the appearance of an energy barrier.

Given the activated nature of the process and its similarities with a classical transition, we will use the analogy and adapt the classical nucleation theory to the case of viral assembly. We will explicitly see how this nucleation mechanisms can explain the main features observed in the assembly and disassembly experiments, and is able to predict successfully the conditions and the rate at which capsid formation takes place.

The basic ideas of a nucleation description of viral assembly where presented in Ref. [27]. But in this chapter we will bring these ideas to a new level by extending the theory in several directions. First, the theory is rederived to facilitate its direct application to experiments by connecting its main ingredients to experimentally measurable quantities. Then, a kinetic description of assembly and disassembly is elaborated, taking into account explicitly time-dependent effects. Finally, these ideas will be used to explain the remarkable hysteresis observed between assembly and disassembly experiments [55] and simulations [46]. Interestingly, some of the basic assumptions of this analysis will be tested and verified in the next chapter by simulating the templated assembly of our coarse-grained model.

This chapter is divided as follows: in Section 6.2 a thermodynamic description to characterize the equilibrium aspects of viral assembly is introduced. Section 6.3 is devoted to discuss the origin of the energy barrier involved in the formation of intermediates. The kinetics of viral assembly is derived in Section 6.4, where a special attention is placed on explaining the general properties observed in experimental scenarios. In Section 6.5 the assembly ideas are adapted to explain the disassembly situation, and to characterize the hysteresis phenomenon. The nucleation theorem is introduced in Section 6.6, which allows a model independent

connection with experimental observations. Finally, in Section 6.7 we summarize our findings and remark some important biological implications.

6.2 Equilibrium aspects of viral self-assembly

As mentioned before, viruses follow different strategies for assembling their capsids, and, in addition to the coat proteins, in some cases the presence of scaffolding proteins or the genetic material is also required. For the sake of simplicity, we will focus on the simplest case: the *in vitro* self-assembly of empty viral capsids formed exclusively by coat proteins. This scenario has been investigated by many in vitro experiments [6, 10, 13–15, 29, 60]. Nevertheless, it is worth to mention that the model and ideas presented below can be extended to deal with more complicated situations that might happen *in vivo*. Another important remark is that viruses can use different paths and building blocks in their assembly process. Here we will use generically the term *subunits* to refer to the elementary building blocks in the assembly, which can be, for instance, single proteins in the case of penicilium chrysogenum virus (PcV) [61], dimers for cowpea chlorotic mottle virus (CCMV) [62], pentamers in polyomavirus [9], or pentamers and hexamers in solution for bacteriophage Hong Kong 97 (HK97) [10].

6.2.1 Quasichemical thermodynamic description of the assembly

During the self-assembly of a virus, the initially free subunits in solution start to join together and form successive partial structures or clusters that will eventually lead to a fully formed capsid [44, 63] (see Fig. 6.1). We will denote by q the total number of subunits in the final shell. For instance, $q = 60$ for a $T = 1$ PcV capsid, whose individual building blocks are single proteins, or $q = 72$ for $T^\star = 7$ human papillomavirus (HPV), whose subunits are pentamers in solution. Let's start our analysis by describing the general equilibrium situation where free subunits, partially formed capsids, and fully-formed shells coexist.

The thermodynamic description of a self-assembling or an aggregative system can be made in two different ways: using the quasichemical or the phase description [64]. In the quasichemical framework, clusters of different number of subunits are treated as different chemical species; whereas in the phase approach, the assembly process is interpreted as a phase transition where subunits can be either free in solution (phase α), or in a complete capsid (phase β). In this context a partial shell is treated like a liquid drop or a crystallite, i.e., a small region of the compact phase β surrounded by the diluted phase α. The phase description is the

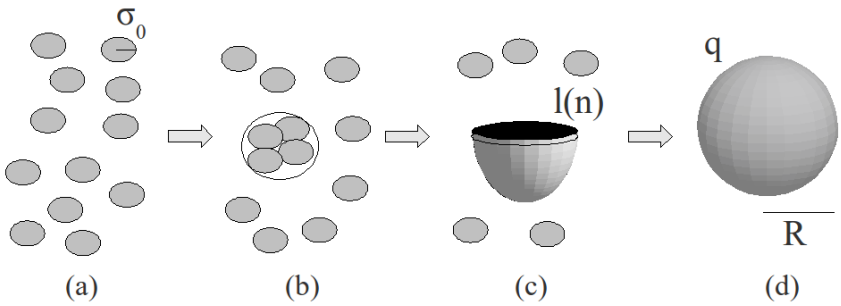


Figure 6.1: Sketch of capsid assembly. The process of assembly of a single capsid is illustrated from left to right. (a) The process is initiated by the presence of free subunits, of a typical size σ_0, in solution; (b) these subunits start to join together by thermal fluctuations, and form a cluster of subunits that acts as a capsid embryo; (c) the embryo grows, and forms a partial capsid characterized by the presence of a rim of length $l(n)$ (where n is the number of aggregated subunits), as the one shown corresponding to a half-formed shell; (d) finally, a complete capsid made of q subunits and a typical radius R is obtained.

one commonly used in classical nucleation theory, and was already introduced by Zandi *et al.* to study the assembly of viral capsids [27]. For completeness, here we will use the quasichemical description, following the ideas of Ref. [64,65], but it is worth to stress that both approaches are equivalent.

Following typical experimental conditions, let's consider a solution of subunits, intermediates, and capsids where the pressure p, the temperature T, and the total number of subunits N are kept constant. In this context the Gibbs free energy, $G(N_1, N_2, ..., N_n, p, T)$, contains all the thermodynamic information of the system[1]

$$G(N_1, N_2, ..., N_n, p, T) = \sum_{n=1}^{q} \mu_n N_n \tag{6.1}$$

where n is the number of subunits in a (partial) capsid that also acts as the label for each species; and, N_n is the number of aggregates of size n that, since the

[1]If the total volume, rather than the pressure, is the variable that it is kept constant, the appropriate thermodynamic potential will be the Helmholtz free energy $F = G - PV$. Since the pressure-volume work contributions are expected to be very small with respect to the total energy of the solution, including the solvent, the description using the Gibbs free energy is for all practical purposes equivalent.

total number of subunits is constant, is related to N by

$$N = \sum_{n=1}^{q} nN_n \,. \tag{6.2}$$

In addition, μ_n is the corresponding chemical potential, which is just the variation of free energy when a particle of a certain species is added to the system

$$\mu_n = \left(\frac{\partial G}{\partial N_n}\right)_{N_{j\neq n},p,T} \tag{6.3}$$

The index $N_{j\neq n}$ indicates that the derivative must be evaluated keeping constant the population of the rest of species.

In general, the total concentration of subunits is far below the concentration of solvent, where the assembly takes place[2]. Thus, we can consider the system as a diluted solution. Moreover, since the solvent is much more abundant than any species N_n, it is reasonable to ignore all interactions between aggregates of different sizes n. Under these circumstances, the chemical potential is given by the standard expression for an ideal mixture of non-interacting species [66, 67]

$$\mu_n = \mu_n^0 + kT \ln \rho_n \tag{6.4}$$

where

$$\rho_n = \frac{N_n}{N_s} \tag{6.5}$$

is the molecular fraction of the species n, and N_s is an arbitrary reference state, which for a diluted system is usually assigned to the molarity of the solvent. The term $\mu_n^0(p,T)$ is the standard part of the chemical potential, i.e., the mean free energy per aggregate of size n, which includes the configurational entropies, and all the internal interactions of a single (partial) capsid, such as the hydrophobic and electrostatic forces.

Indeed, in experiments is common to work with volume concentrations, $c_n = N_n/V$, instead of molecular fractions. Thus, to facilitate the connection with experiments, we derive the theory using volume concentrations. In particular, the molecular fraction ρ_n can be reexpressed as[3]

$$\rho_n = \frac{c_n}{c_s} \tag{6.6}$$

[2]The molarity of water is $\rho_s \sim 56M$, whereas for typical experimental conditions the concentration of subunits is on the order of $\rho_1 \sim 10\mu M$. Therefore $\rho_s \gg \rho_1$, and the assumption of being a dilute solution will generally hold.

[3]Technically, the volume V of the system is not constant, but in practice the fluctuations are small, due to the abundance of solvent compared to the rest of species.

and the chemical potential now reads

$$\mu_n = \mu_n^0 + kT \ln(c_n/c_s) \tag{6.7}$$

It is worth mentioning that the standard part $\mu_n^0(p,T)$ implicitly depends on the reference state given by c_s. Since the solvent is usually composed mainly by water, for simplicity and convenience, we chose the molarity of pure water $c_s = 56M$ as a reference state. But, it is worth to emphasize that the description holds irrespective of the arbitrarily chosen reference state.

The equilibrium distribution of aggregates in the system can be easily obtained by minimizing the Gibbs free energy, Eq. (6.1), subjected to the constraint that the total number of subunits N is constant, Eq (6.2). This minimization can be done, for instance, by using Lagrange multipliers, and one gets the familiar equilibrium condition

$$\mu_1^{eq} = \frac{\mu_n^{eq}}{n} \equiv \mu^{eq} \tag{6.8}$$

where μ^{eq} is the equilibrium chemical potential. Therefore, at equilibrium in the quasichemical approach, the chemical potential per subunit is the same for all partial capsids. Now, by using Eq. (6.7), this condition leads to the well known law of mass action

$$\frac{c_n^{eq}}{(c_1^{eq})^n} = \frac{e^{-\Delta W(n)/kT}}{(c_s)^{n-1}} \equiv K_n^{eq} \tag{6.9}$$

where K_n^{eq} is the equilibrium constant, which is independent on the total subunit concentration. The term

$$\Delta W(n) \equiv \mu_n^0 - n\mu_1^0 \tag{6.10}$$

is the standard free energy difference required to form a cluster of n subunits. However, we stress that its actual value depends on the choice of c_s, and such reference must be provided to properly compare the energies derived from different experiments [68, 69]. Nonetheless, the standard free energy difference vanishes for single subunits, i.e., $\Delta W(1) = 0$, by construction and independently on the choice of c_s.

The law of mass action, Eq. (6.9), can be rewritten in a standard equilibrium Boltzmann's like expression as

$$c_n^{eq} = c_s e^{-\Delta G(n)/kT}, \tag{6.11}$$

where

$$\Delta G(n) = \Delta W(n) - n\,kT \ln(c_1/c_s) \tag{6.12}$$

is the total work of formation of a partially formed cluster containing n subunits. At this stage, it is worth to emphasize the differences between $\Delta G(n)$ and $\Delta W(n)$: $\Delta W(n)$ represents an standard (or intrinsic) free energy difference, which is independent on the subunit concentration; on the contrary, $\Delta G(n)$ is the full free energy difference required to build a (partial) capsid of n subunits, and consists of the intrinsic energy difference between having the subunits in a capsid or in solution, $\Delta W(n)$, plus an entropic penalty for removing the n free subunits from the solution to make the cluster, $-nkT\ln(c_1/c_s)$, which obviously depends on the concentration of free subunits c_1. Another important remark is that $\Delta G(n)$ does not depend on any other concentration c_n with $n \geq 2$, and its valid irrespective of whether the concentration of aggregates has reached the aggregative equilibrium or not.

Therefore, from the equilibrium distribution of cluster sizes c_n^{eq} or, equivalently, by measuring the equilibrium constants K_n^{eq}, one can extract the work of formation for any (partial) capsid. However, as we mentioned in the introduction, in typical viral assembly experiments, free subunits and fully-formed capsids are essentially the only species detectable in solution. So, with the current experimental resolution, it is not possible to extract the equilibrium constant for an arbitrary intermediate of size n [63]. In Section 6.3, we will introduce a simple model for $W(n)$, inspired by the classical nucleation theory [27], that justifies the absence or scarcity of intermediates. But before that, it is worth to discuss the equilibrium when only the two dominant species, i.e., free subunits and fully formed capsids, are present in the solution.

6.2.2 Equilibrium between free subunits and fully-formed capsids

Experiments and simulations confirm that generally at the end of the assembly proteins are mainly distributed in two populations: free subunits and complete capsids [13, 15, 42, 44], as illustrated in Fig. 6.2. Therefore, one can simplify the general expressions derived above by considering only the coexistence of N_1 free subunits and N_q complete shells, made of q subunits. Then, the conservation of the total number of subunits, Eq. (6.2), implies

$$N = N_1 + qN_q \tag{6.13}$$

and the equilibrium condition, Eq. (6.8), becomes

$$\mu_q^{eq} = q\mu_1^{eq} \equiv \mu^{eq} \tag{6.14}$$

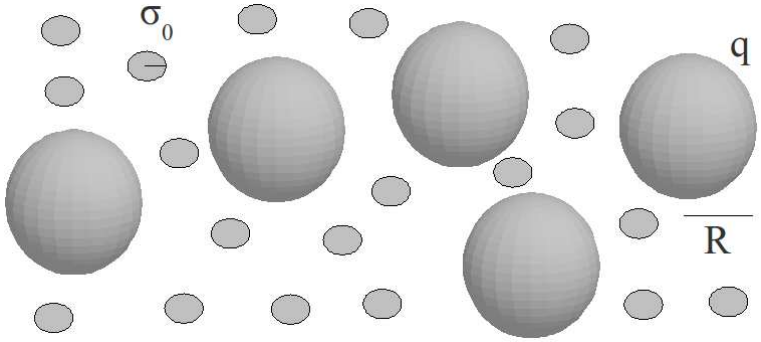


Figure 6.2: Equilibrium between free subunits and complete capsids. The ideal mixing of N subunits distributed in N_1 free subunits, of typical size σ_0, and N_q capsids, of radius R and made of q subunits, is illustrated.

Now, taking into account Eq. (6.7), the equilibrium condition leads again to the law of mass action

$$\frac{c_q^{eq}}{(c_1^{eq})^q} = \frac{e^{-\Delta W(q)/kT}}{(c_s)^{q-1}} \equiv K_q^{eq} \tag{6.15}$$

with

$$\Delta W(q) = q(\mu_q^0/q - \mu_1^0) \equiv q\Delta g \tag{6.16}$$

The last expression defines an effective binding energy per subunit Δg, which is the difference between standard chemical potentials of a subunit in a complete capsid or free in solution.

Let's define

$$f = \frac{qc_q}{c}, \tag{6.17}$$

as the fraction of subunits in fully-formed capsids, where

$$c = c_1 + qc_q, \tag{6.18}$$

represents the total concentration of subunits in the solution. Using Eqs. (6.17) and (6.18), we can rewrite the concentration of free subunits and complete capsids as

$$c_q = \frac{fc}{q}, \tag{6.19}$$

and

$$c_1 = (1 - f)c. \tag{6.20}$$

By substituting both expressions in the law of mass action, Eq. (6.15), after a simple rearrangement, one gets

$$\frac{f}{(1-f)^q} = K_q^{eq}\, q\, c^{q-1}. \tag{6.21}$$

Let's define the critical micelle concentration c_{cmc} as the value of the total concentration at which $f = 1/2$, i.e., half of the subunits are in complete capsids at equilibrium. By imposing this requirement in the previous equation we get

$$\frac{c_{cmc}}{2} = \left(qK_q^{eq}\right)^{\frac{-1}{q-1}}. \tag{6.22}$$

so we can rewrite Eq. (6.21) as

$$\frac{f}{(1-f)^q} = \left(\frac{2c}{c_{cmc}}\right)^{q-1}. \tag{6.23}$$

We can now analyze how the subunits are distributed between being free in solution or in complete capsids as a function of the total concentration. For $q \gg 1$, which is typically the case for viral capsids, if $2c/c_{cmc} < 1$, then $f \simeq 0$ and basically all subunits remain in solution. To analyze the behavior for $2c/c_{cmc} > 1$ it is better to rewrite Eq. (6.23) as

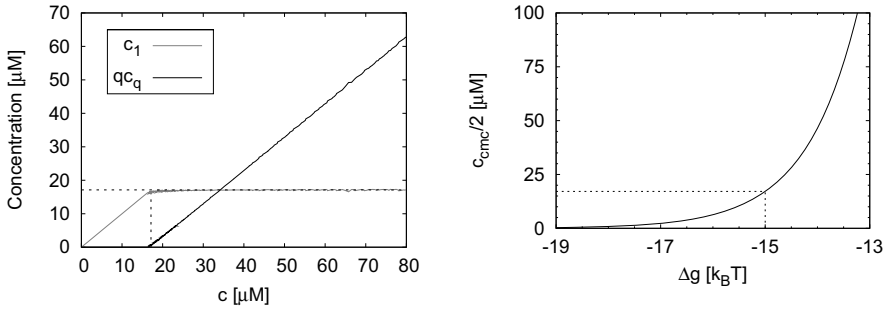
$$1 - f = f^{1/q}\left(\frac{c_{cmc}}{2c}\right)^{1-1/q}. \tag{6.24}$$

which for $q \gg 1$ can be approximated as

$$1 - f \simeq \left(\frac{c_{cmc}}{2c}\right). \tag{6.25}$$

So for $2c/c_{cmc} > 1$, the concentration of free subunits remains constant $c_1 = c(1-f) \simeq c_{cmc}/2$, whereas the concentration of subunits in capsids grows linearly with the total concentration as $qc_q = fc \simeq c - c_{cmc}/2$. This behavior is represented in Fig. 6.3. One can clearly see that the $c_{cmc}/2$ is roughly the concentration at which aggregates start to appear.

It is remarkable that *in vitro* self-assembly experiments performed at different conditions, where the binding energy has been modified, comply with the predicted universal aggregation curve, Eq. (6.24). In particular, Kegel and van der Schoot [68] showed this universality for HBV, plotting the fraction of subunits assembled into capsids, f, versus the ratio of total and critical concentrations of subunits (dimers in the case of HBV), c/c_{cmc} (see Fig. 6.4). The *in vitro*



(a) (b)

Figure 6.3: Law of mass action. (a) The concentration of subunits free in solution, c_1, and in capsids, qc_q, is illustrated for different total concentrations c, according the law of mass action, Eq. (6.15). The reference concentration is $c_s = 56\,M$, the binding energy per subunit is $\Delta g = -15\,k_BT$, and capsids are made of $q = 120$ subunits. Both horizontal and vertical dashed lines correspond to the aggregation concentration $c_{cmc}/2$, Eq. (6.22). (b) The exponential dependence of the critical micelle concentration c_{cmc} with respect to the binding energy is illustrated. The dashed lines correspond to the Δg used in (a).

HBV assembly experiments where done at room temperature for different salt concentrations, and all data fall on top of the prediction of Eq. (6.24).

Finally, let's relate c_{cmc} with the binding energy per capsomer. Using Eqs. (6.15) and (6.22), one gets

$$\frac{c_{cmc}}{2} = c_s e^{\frac{\Delta g}{kT(1-1/q)}} q^{\frac{-1}{q-1}}. \tag{6.26}$$

that in the limit $q \gg 1$ yields

$$\frac{c_{cmc}}{2} \simeq c_s\, e^{\,\Delta g/kT}. \tag{6.27}$$

Therefore, the concentration at which the capsid formation starts to become favorable depends exponentially on the binding energy per subunit in a fully formed capsid.

Apparently, all these results seem to suggest that capsid self-assembly could be considered as an equilibrium aggregation process. However, this is not completely true, since, surprisingly, shell dissociation does not occur at the same concentrations at which the assembly is triggered. This hysteresis phenomenon, together with the scarcity of intermediates observed, point to the existence of an

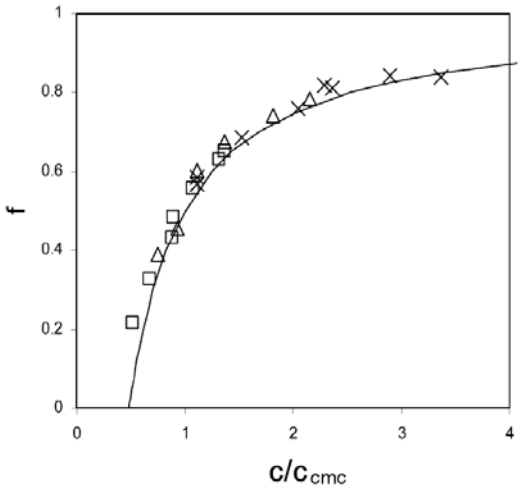


Figure 6.4: Universality of capsid assembly. Fraction of subunits in fully-formed capsids, f, versus the overall concentration of subunits c (dimers in the case of HBV), scaled to the critical micelle concentration c_{cmc}. The symbols are data from Ceres and Zlotnick [13] in the study of HBV capsid assembly under different salt concentrations. All samples are at a temperature of 25^oC. Crosses, $c_{salt} = 0.7\,M$; triangles, $c_{salt} = 0.5\,M$; and squares, $c_{salt} = 0.3\,M$. The plotted line is the universal aggregation curve, given by Eq. (6.24). Figure adapted from Ref. [68].

energy barrier to overcome during capsid formation [27]. As we will see along the chapter, the experimental results discussed above do not correspond in general to a true equilibrium aggregation, but rather to a kinetically stopped activated process whose final state resembles an apparent equilibrium.

We devote the next subsection to explain the origin of this energy barrier and to model the free energy of formation of a partially formed capsid, following the standard ideas of Classical Nucleation Theory (CNT).

6.3 Free energy barrier for capsid formation

As we have just seen, the aggregation of a capsid is driven by the binding energy, i.e., by the fact that a subunit in a fully formed capsid has a lower energy that in solution. For $c > c_{cmc}/2$, this energetic advantage overcomes the entropic penalty of removing that free subunit from the solution, and thus the formation of a *com-*

plete capsid becomes thermodynamically favorable. Nevertheless, the assembly of capsids proceeds through the successive addition of subunits to intermediate structures, and those subunits placed at the rim of a partial capsid will obviously miss some neighbors in the exposed part. Thus, the formation of an intermediate shell will intrinsically entail an energetic cost that originates an energy barrier.

The situation is analogous to a vapor-liquid phase transition, where the formation of a liquid drop in the vapor phase has an energetic penalty associated to the surface tension. Following the analogy with a drop, the simplest way to model this energetic cost is by introducing a rim energy penalty associated to a line tension λ, which in the formation of this 2D shell plays the same role as the surface tension for a drop.

Accordingly, the standard or intrinsic energy of formation of a partial capsid containing n subunits can be modeled as

$$\Delta W(n) = n\Delta g + \lambda\, l(n). \tag{6.28}$$

The first term represents the gain related to the binding energy Δg, and the second is the total line energy for a partial capsid n, where the line tension λ is the energetic cost per unit length, and $l(n)$ is the length of the rim. We stress that the previous equation applies for $n \geq 2$, since, as mentioned before, the formation energy of a free subunit is zero by construction, $\Delta W(1) = 0$, and the line tension would be meaningless in this case. It is also worth to mention that a complete capsid, $n = q$, has no open rim, $l(q) = 0$, so that the line energy will vanish. Accordingly, we recover $\Delta W(q) = q\Delta g$, Eq. (6.16). The average binding energy Δg will depend, in general, on the pH, salt concentration, temperature, or the presence of small molecules, like ions or chaotropic agents. As a first approximation, this binding energy will be taken as constant, even though subunits in a capsid can occupy different local environments (see Chapter 1), or be engaged in different interactions. Surprisingly, we will show in the next chapter that this crude approximation works in fact rather nicely.

Evidently, the growth of a shell is produced by discrete events when new subunits are incorporated to the structure. However, following CNT ideas, it seems convenient to approximate a partial capsid by a continuum model, which will simplify the description of the rim. In particular, let's consider an intermediate shell as a spherical cap of radius R and angle θ, which indicates the degree of completion (see Fig. 6.5). In this case, the length of the rim will be $l(\theta) = 2\pi R \sin\theta$. Now, to relate θ and n we introduce two simplifying assumptions: first, we consider that the radius R of the shell remains constant during the growth; and second, the effective area occupied by a subunit in the shell, a_1,

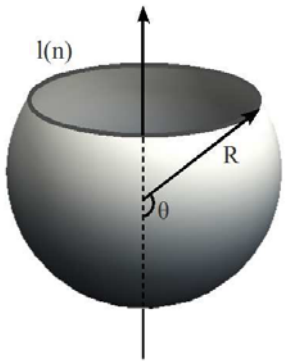


Figure 6.5: Partial capsid model. Schematic representation of a partially formed capsid in the continuum approximation. R is the radius of the capsid and the angle θ characterizes its degree of completion. The circular rim has a length $l(n)$ that depends on the number of subunits n in the partial shell.

is the same as in the complete capsid made of q subunits, so

$$a_1 = \frac{4\pi R^2}{q}. \tag{6.29}$$

The surface of this partial shell is $A(\theta) = 2\pi R\, h(\theta)$, with $h = R\,(1 - \cos\theta)$ being the distance from the pole of the partial capsid to the center of the hole defined by the rim. Then, dividing this surface by the area occupied by a single subunit a_1 we obtain the number of subunits n in an incomplete capsid of angle θ, namely, $n(\theta) = A(\theta)/a_1 = q(1 - \cos\theta)/2$. Using this last result, we can rewrite the boundary length $l(\theta)$ as a function of n as

$$l(n) = 4\pi \frac{R}{q}\sqrt{n(q-n)}. \tag{6.30}$$

At first sight, this circular rim description could seem too simplistic to describe accurately the assembly of capsids. However, in Appendix A we show that the dependency obtained for $l(n)$ on R, q, and n holds even for non-circular rims scenarios, and in general

$$l(n) = a\frac{R}{q}\sqrt{n(q-n)} \tag{6.31}$$

where a is a constant geometrical prefactor independent of n, R or q. In particular, for a circular boundary its value becomes $a = 4\pi$. In the next chapter we

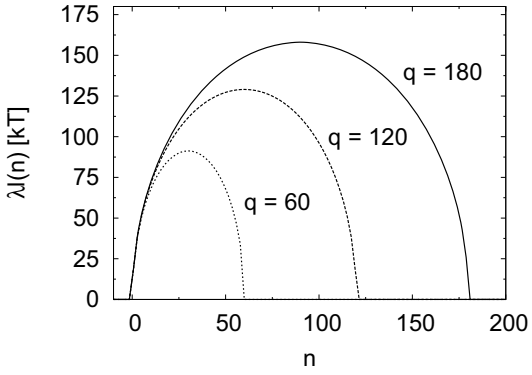


Figure 6.6: Rim energy for capsid formation. The line energy $\lambda\, l(n)$, using a circular rim approximation, Eq. (6.30), is plotted for three different capsids sizes q. The value used for the line tension is associated to a binding energy $\Delta g = -15\, kT$, Eq. (6.35).

will see that both the line tension concept and the scaling of $l(n)$ are in perfect agreement with the simulation results on the growth of shells, using the physical model introduced in Part I of the thesis.

We can now express Eq. (6.28) for $n \geq 2$ making explicit the dependence on n

$$\Delta W(n) = n\Delta g + \alpha\sqrt{n(q-n)} \tag{6.32}$$

where

$$\alpha = a\,\frac{\lambda R}{q} \tag{6.33}$$

is an energy constant absorbing the line tension λ and the geometrical prefactor in Eq. (6.31). In Fig. 6.6 we plot the line energy term $\lambda\, l(n) = \Delta W(n) - n\Delta g$, as a function of the number of subunits n of a partially formed shell. This rim energy is always positive, has its maximum when the capsid is half-formed, $n = q/2$, and vanishes for a complete capsid, $n = q$.

6.3.1 A simple estimate of the line tension

Although it is not necessary for the development of the theory, it is convenient to provide a simple estimate of the line tension λ. In the next chapter we will evaluate more precisely its value in a particular case.

Let's consider a 2D lattice that is separated in two halves by a cut of length L. The increment in boundary energy associated to this process is $\Delta E_b = 2\lambda L$. The line tension can then be calculated as

$$\lambda = \frac{1}{2}\frac{\Delta E_b}{L} \tag{6.34}$$

where the energy ΔE_b is approximately given by the total of broken contacts. Assuming that in the cut each subunit of diameter $2\sigma_0$ losses half of its binding energy, then a rough estimate of the line tension will be

$$\lambda \approx \frac{1}{2}\frac{|\Delta g|}{2\sigma_0} \tag{6.35}$$

where Δg is the energy per subunit in the original lattice.

It is then possible to give a simple estimation for the typical values of α. As we saw in Chapter 3, the radius of a shell is intimately related to the effective area of the coat subunits, i.e., $4\pi R^2 \approx q\pi(\sigma_0)^2$ for a spherical shell. This leads to $R \approx \sqrt{q}\,\sigma_0/2$, and using Eq. (6.33) for the circular rim model one obtains

$$\alpha \approx 2\pi\frac{\lambda}{\sqrt{q}}\sigma_0. \tag{6.36}$$

Now, plugging the estimation of λ derived above, one gets

$$\alpha \approx \frac{\pi}{2}\frac{|\Delta g|}{\sqrt{q}}. \tag{6.37}$$

Therefore, for a capsid made of $q = 120$ subunits with an average binding energy per subunit $\Delta g = -15\,kT$, one gets $\alpha \approx 2.15\,kT$. We remark that α is inversely proportional to $\sqrt{q}$, so this value can vary significantly for different capsid sizes.

In any case, in the following we will continue our derivation using a generic value of α.

6.3.2 Free energy of capsid formation

Using this line energy concept for the intrinsic part, $\Delta W(n)$, the full work of formation $\Delta G(n)$, Eq. (6.12), becomes

$$\Delta G(n) = n\Delta g - nkT\ln(c_1/c_s) + \alpha\sqrt{n(q-n)} \tag{6.38}$$

It is convenient to rewrite the total free energy of capsid formation as

$$\Delta G(n) = n\Delta\mu + \alpha\sqrt{n(q-n)} \tag{6.39}$$

with

$$\Delta\mu = -kT\ln(c_1/c^*) \tag{6.40}$$

being the chemical potential difference and

$$c^* = c_s e^{\Delta g/kT} \tag{6.41}$$

a threshold concentration, which plays the same role as the saturation or equilibrium density in a vapor to liquid transition. Notice also that this concentration roughly coincides with what would be the critical micelle concentration in an equilibrium assembly, i.e., $c_{cmc}/2$, defined in the previous section, in the limit $q \gg 1$, Eq. (6.27), so

$$c^\star \approx \frac{c_{cmc}}{2}. \tag{6.42}$$

Since the binding energy Δg is, by definition, negative, this threshold concentration gets smaller as the binding gets stronger (see Fig. 6.3b).

The chemical potential difference $\Delta\mu$ and the state of aggregation of capsid subunits depends crucially on the concentration of free subunits c_1, Eq. (6.40). For $c_1 < c^*$, the chemical potential is positive $\Delta\mu > 0$, indicating that the binding energy gained by a subunit in getting incorporated to a capsid is not able to overcome the entropic penalty of removing it from solution, so that the formation of a capsid is unfavorable. For $c_1 > c^*$ the formation of a capsid is favored, but in that process the system has still to overcome a significant energetic barrier, whose characteristics are evaluated below.

6.3.3 Nucleation barrier and critical nucleus

The free energy of capsid formation $\Delta G(n)$, Eq. (6.39), is made of two clear contributions: a term proportional to n that in general will drive the assembly, plus an energetic penalty proportional to $\sqrt{n(q-n)}$ related to the rim of partial capsids. The competition between these two opposite contributions gives rise to a barrier (see Fig. 6.7), which plays a very important role in the kinetics of capsid formation (see Section 6.4). The height of this barrier $\Delta G^\star$ and the location of the maximum $n^\star$, can be easily computed by applying the condition $\partial\Delta G/\partial n = 0$, which leads to

$$\Delta G^\star = \frac{q}{2}\left(\sqrt{\Delta\mu^2+\alpha^2}+\Delta\mu\right) \tag{6.43}$$

and

$$n^\star = \frac{q}{2}\left(1+\frac{\Delta\mu}{\sqrt{\Delta\mu^2+\alpha^2}}\right), \tag{6.44}$$

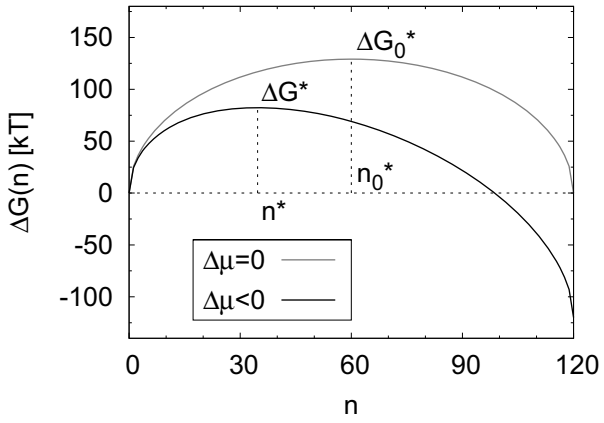


Figure 6.7: The total free energy of formation for a partial capsid. $\Delta G(n)$ is plotted according to Eq. (6.38). The total capsid has $q = 120$ subunits and the energy scale for the line energy is $\alpha \approx 2.15\, kT$. The two lines correspond, respectively, to phase equilibrium, $\Delta\mu = 0$, and a supersaturated scenario, $\Delta\mu < 0$, Eq. (6.40).

which are clearly related by

$$\frac{\Delta G^\star}{n^\star} = \sqrt{\Delta\mu^2 + \alpha^2}. \tag{6.45}$$

Therefore, there are only two effective parameters that control the energy barrier $\Delta G^\star$ and the critical cluster $n^\star$: α, which is a constant parameter associated to the line tension λ, Eq. (6.33); and $\Delta\mu$, which is a variable term that depends on the concentration of free subunits c_1, Eq. (6.40).

As we see in Fig. 6.7, for $n < n^\star$ the free energy of formation grows as we add a subunit ($\partial\Delta G/\partial n > 0$), so intermediates of these sizes will tend to disassemble back into free subunits. On the other hand, partial capsids of size $n > n^\star$, reduce their energy upon the addition of another subunit, ($\partial\Delta G/\partial n < 0$), and they will tend to grow spontaneously until complete a capsid. Thus $n^\star$ is the *critical size* that partial capsids have to reach to trigger the formation of complete shells, and can be considered as the *nucleus* or the *embryo* of the assembly process.

It is worth to emphasize that it is precisely the existence of this energy barrier associated to the formation of partial capsids what makes the population of intermediates very scarce, in general. As was shown in Section 6.2.1, the equilibrium population of intermediates of size n is proportional to $e^{-\Delta G(n)/kT}$. Thus the bar-

rier, or more precisely the line tension term, penalizes this intermediate sizes in such a way that typically only free subunits or fully-formed capsids are observed. Remarkably, the critical nucleus is a maximum of the energy of formation $\Delta G(n)$, and the associated critical partial capsid $n^\star$ will be always thermodynamically unstable. Therefore, if a experiment aims to reconstruct the sequence of partial capsids during the assembly process, the critical capsid will be precisely the most challenging species to be observed, rather than the most abundant intermediate detected. This is an important clarification in the context of the nucleation mechanism of self-assembly, and it is a common misunderstanding in the interpretation of experiments and simulations [26, 44].

It is important to remark that the height of the nucleation barrier $\Delta G^\star$ depends on the value of $\Delta\mu$, or equivalently, on the concentration c_1 of free subunits (see Fig. 6.7). As we discussed previously, at phase equilibrium, $\Delta\mu = 0$, the free energy of formation for intermediates of size n is determined exclusively by the intrinsic line energy term. In that case the critical nucleus will correspond to a half-formed capsid

$$n_0^\star = \frac{q}{2} \tag{6.46}$$

and the nucleation barrier will be just the line energy of this hemi-shell

$$\Delta G_0^\star = \frac{q}{2}\alpha \tag{6.47}$$

For typical values of α and q this barrier is much bigger than the thermal energy: $\Delta G^\star \gg k_B T$ (see Fig. 6.8a). But for supersaturated conditions, i.e., when $c_1 > c*$ or $\Delta\mu < 0$, the barrier can be low enough to promote spontaneous self-assembly at reasonable time scales[4].

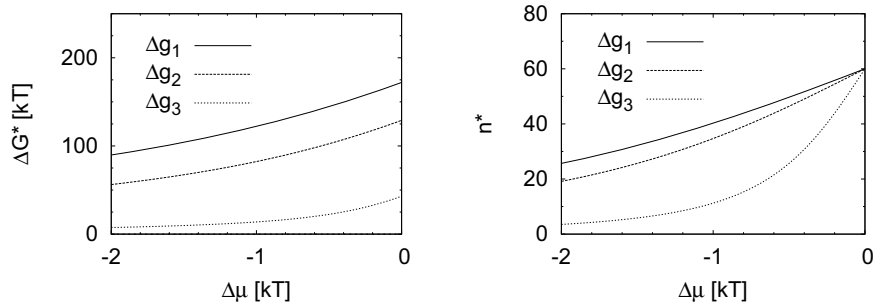
Asymptotic cases

It is worth to study two limit situations that simplify the expressions for $\Delta G^\star$ and $n^\star$, providing a clearer insight of the influence of α and $\Delta\mu$ in the formation of capsids.

The first one corresponds to a situation near phase equilibrium, where we have the condition

$$\left(\frac{\Delta\mu}{\alpha}\right)^2 \ll 1 \tag{6.48}$$

[4]The limiting case where the barrier is suppressed and nucleation does not control the rate of assembly has been studied by van der Schoot and Zandi in Ref. [34].



(a) (b)

Figure 6.8: Energy barrier and critical nucleus of assembly. (a) Height of the energy barrier, Eq. (6.39), for different supersaturation conditions, $\Delta\mu < 0$. The nucleation energies curves are associated to capsids with $q = 120$ and binding energies $|\Delta g|$ of $20\,kT$, $15\,kT$, and $10\,kT$. The values of α used were estimated from Eq. (6.37). (b) Critical nuclei, Eq. (6.44), associated to the barrier energies in (a).

so there is not a significant advantage for a subunit to be in capsids rather than in solution. By expanding the activation energy, Eq. (6.43), and critical nucleus, Eq. (6.44), up to first order one obtains

$$\Delta G^\star \approx \frac{q}{2}\alpha\left(1+\frac{\Delta\mu}{\alpha}\right) \tag{6.49}$$

and

$$n^\star \approx \frac{q}{2}\left(1+\frac{\Delta\mu}{\alpha}\right) \tag{6.50}$$

that are related by

$$\frac{\Delta G^\star}{n^\star} \approx \alpha \tag{6.51}$$

Therefore, one gets a linear reduction with $\Delta\mu$ of the activation energy and critical nucleus with respect the equilibrium values $\Delta G_0^\star$, Eq. (6.47), and $n_0^\star$, Eq. (6.46).

On the opposite limit, if the system is subjected to *strong* supersaturation conditions, i.e.,

$$\left(\frac{\alpha}{\Delta\mu}\right)^2 \ll 1 \tag{6.52}$$

then the first order expansion of the energy barrier and the critical nucleus leads to

$$\Delta G^\star \approx \frac{1}{4}\frac{q\alpha^2}{|\Delta\mu|} \tag{6.53}$$

and

$$n^\star \approx \frac{1}{4}\frac{q\alpha^2}{\Delta\mu^2} \tag{6.54}$$

that in this case are related by

$$\frac{\Delta G^\star}{n^\star} \approx |\Delta\mu| \tag{6.55}$$

So both $\Delta G^\star$ and $n^\star$ are strongly reduced.

6.4 Kinetics of viral self-assembly

The self-assembly of viral capsids is in its essence a kinetic process. In this section we will describe the kinetics of viral assembly, and connect it with the previous thermodynamic description. In this way we will be able to estimate the rate of assembly.

The assembly of a viral capsid can be thought as a reaction cascade where q free subunits end up forming a single shell. However, to determine the kinetics of the problem, it is necessary to be more specific about the intermediate steps of this reaction. In principle, a partial capsid could grow by different mechanisms, for instance, by the addition of a single subunit, multiple independent subunits simultaneously, or even by merging different partial capsids. Fortunately, simulations have shown that predominantly only one subunit is added at a time [42,44], which could be justified by comparing the *a priori* probabilities of all the different possible events. Therefore, it is reasonable to start the kinetic study considering the formation of a capsid as a series of unimolecular reactions, where the attachment or detachment of just a single unit promotes the growth or decay of a partial capsid. This can be formally expressed as

$$c_n(t) \underset{\alpha_{n+1}(t)}{\overset{\beta_n(t)}{\rightleftharpoons}} c_{n+1}(t) \quad , \quad n = 1 : q-1 \tag{6.56}$$

where $c_n(t)$ is the concentration of partial capsids made of n subunits at time t, and $\beta_n(t)$ and $\alpha_n(t)$ are the associated binding and unbinding rates, respectively, which in general may depend on the size n and time t.

Accordingly, the change in time of the population, or concentration, of partial capsids of size n is controlled by the following master equation

$$\begin{aligned}\frac{\partial c_n(t)}{\partial t} = \quad & \beta_{n-1}(t)\, c_{n-1}(t) - \alpha_n(t)\, c_n(t) \\ & -\beta_n(t)\, c_n(t) + \alpha_{n+1}(t)\, c_{n+1}(t)\end{aligned} \tag{6.57}$$

In words, this means that a partial capsid of size n can disappear and become a $n-1$ or $n+1$ intermediate by the loss or gain of a single subunit, respectively, which corresponds to the second and third terms on the right of Eq. (6.57); besides, a new intermediate of size n can be obtained by adding or subtracting a subunit to a partial capsid of size $n-1$ (the first term) or $n+1$ (the last term), respectively. The master equation applies from $n = 1$ up to $n = q$, taking the special conditions $\beta_q(t) = 0$ and $\alpha_1(t) = 0$, i.e., a fully formed capsid is the last step of the reaction and cannot grow further, while a free subunit cannot *loose* any more subunits.

It is convenient to rewrite the master equation as

$$\frac{\partial c_n(t)}{\partial t} = J_{n-1}(t) - J_n(t) \tag{6.58}$$

where we have defined

$$J_n(t) \equiv \beta_n(t)\, c_n(t) - \alpha_{n+1}(t)\, c_{n+1}(t) \tag{6.59}$$

as the net current of partial capsids of size n growing to size $n+1$.

By solving the master equation one obtains the populations of free subunits, intermediates, and complete capsids during the assembly process. Nevertheless, to do so one needs first to know the rates of attachment, $\beta_n(t)$, and detachment, $\alpha_n(t)$.

The rate of attachment $\beta_n(t)$ can be reasonably modeled using standard's Smoluchowski theory of aggregation, see, for instance, Ref. [70]. At normal conditions, the intermediates in the assembly may have enough time to thermalize, before any consecutive attachment or dissociation of a subunit. Hence, it is plausible to assume that any binding or unbinding event is completely uncorrelated from a previous event[5]. Under such circumstances, and without considering crowding effects, the binding rate of subunits to a partial shell n will be proportional to the concentration of free subunits, $c_1(t)$, their diffusion constant, D_1,

[5]From a statistical physics standpoint, this approximation means that the assembly of a viral capsid can be regarded as a Markovian process

and the size of the rim of the partial shell

$$\beta_n(t) = b\, c_1(t) D_1 r_n \tag{6.60}$$

where b is a geometric correction, D_1 is the diffusion coefficient of the free subunits, and r_n is the radius of the rim of the partial capsid[6]. We stress that the concentration of free subunits $c_1(t)$ will generally vary during the assembly process, so the rate of attachment $\beta_n(t)$ will be clearly time-dependent.

On the other hand, the unbinding rate $\alpha_n(t)$ is an intrinsic property of each cluster n, and depends on thermal fluctuations and the interactions holding the subunits in the rim of the partial shell. In general, it will be difficult to estimate, but this problem can be circumvent by considering the master equation at *constrained equilibrium* conditions, dictated by a fixed concentration of free-subunits[7] c_1.

At constrained equilibrium there will be no change in the capsid size distribution, and the net flux in the system will vanish, so

$$J_n = 0 \tag{6.61}$$

for all n. Thus, from Eq. (6.59) one has

$$\beta_n\, c_n^{eq} = \alpha_{n+1}\, c_{n+1}^{eq} \tag{6.62}$$

which is known as the *detailed balance* condition. Here is the point where the thermodynamics of capsid formation enters into the kinetic description of the assembly. The constrained equilibrium concentrations, c_n^{eq}, must follow a Boltzmann's like distribution, Eq. (6.11), so we obtain

$$\alpha_{n+1} = \beta_n\, e^{\,(\Delta G(n+1) - \Delta G(n))/kT} \tag{6.63}$$

where $\Delta G(n)$ is the free energy of formation of a partial capsid n for a given c_1 concentration of free subunits, as determined by Eq. (6.38). Notice that the free

[6]This expression is derived by calculating the flux of free subunits that collides with a sphere of radius r_n. Since subunits attach to the rim of the partial shell rather than on the exposed surface of the "hole" this rate is probably an overestimation, but provides at least a fair qualitative picture of the attachment rate. The precise calculation of Smoluchowski's aggregation rate for partial shells will be pursued in future investigations.

[7]The constrained equilibrium state is defined as the aggregative equilibrium state for which the equilibrium concentration of subunits will be c_1. Note that this will not necessarily coincide with the true aggregative equilibrium that is fixed by the total concentration c, rather than by c_1. It is worth to remark that the same final results can be obtained taking as a reference the full aggregative equilibrium state. The derivation of the rate using this reference is, however, a bit more indirect.

energy difference in the exponential leads to a $1/c_1$ term that cancels with the c_1 factor coming from β_n, Eq. (6.60). Therefore, the rate of detachment is independent of the concentration of free subunits or intermediates, as expected, and depends on the intrinsic properties of the partial capsid. Importantly, Eq. (6.63) will be valid for any assembly conditions (as far as the assumptions introduced so far hold).

The master equation can be solved numerically if one knows the initial conditions, and either the attachment α_n and evaporation rates β_n, or the equilibrium free energy of formation $\Delta G(n)$. However, this could be a daunting task depending on the assembly conditions and, specially, for large capsids q. Furthermore, the master equation does not provide any explicit insight of the influence of the free subunit concentration c_1, or the binding energy per subunit Δg, in the assembly kinetics. To properly characterize the assembly process, it is convenient to derive an approximate analytical solution for Eq. (6.57).

Using the detailed balance condition, Eq. (6.62), we can now eliminate α_{n+1} from the expressions of the current, Eq. (6.59), yielding

$$J_n(t) = \beta_n(t)\, c_n^{eq} \left(\frac{c_n(t)}{c_n^{eq}} - \frac{c_{n+1}(t)}{c_{n+1}^{eq}} \right) \tag{6.64}$$

or

$$\frac{J_n(t)}{\beta_n(t) c_n^{eq}} = \frac{c_n(t)}{c_n^{eq}} - \frac{c_{n+1}(t)}{c_{n+1}^{eq}} \tag{6.65}$$

By adding up all the terms in the last expression from $n = 1$ to $n = q - 1$, only the first and last terms survive, obtaining

$$\sum_{n=1}^{q-1} \frac{J_n(t)}{\beta_n(t) c_n^{eq}} = \frac{c_1(t)}{c_1^{eq}} - \frac{c_q(t)}{c_q^{eq}} \tag{6.66}$$

Therefore, the rate of capsid formation depends on the boundary conditions related to the concentration of free subunits and fully formed capsids. Let's now discuss two different practical situations.

6.4.1 Steady-state rate of capsid formation

Let's first study the solution in the steady-state scenario, which is characterized by the condition $\partial c_n(t)/\partial t = 0$ that applies for all n. This means that the current is the same for all capsid sizes, i.e.,

$$J_n = J_{n+1} \equiv J \tag{6.67}$$

which defines the steady state rate of capsid formation J. As we saw above, the value of this rate depends on the boundary conditions for the concentrations of free subunits and fully-formed capsids. In particular,

$$J = \left(\frac{c_1(t)}{c_1^{eq}} - \frac{c_q(t)}{c_q^{eq}} \right) / \sum_{n=1}^{q-1} \frac{1}{\beta_n(t) c_n^{eq}} \tag{6.68}$$

It is thus clear that the equilibrium state is just a particular case of steady-state where the assembly rate vanishes $J^{eq} = 0$.

To get a steady state with a finite current J, the proper boundary conditions are that the concentration of free subunits is fixed at the *constrained* equilibrium concentration, $c_1(t)/c_1^{eq} = 1$, and that of fully-formed capsids is vanishing small, $c_q(t)/c_q^{eq} \simeq 0$. Strictly, to maintain this situation one should remove the newly produced capsids and return their subunits back to the solution. However, these conditions will be a reasonable approximation at the beginning of the assembly, provided that the production of capsids is low enough to not alter much the concentration of free subunits. Then, plugging these conditions into Eq. (6.68), one finds a general expression for the steady-state rate of assembly

$$J = 1/ \sum_{n=1}^{q-1} \frac{1}{\beta_n \, c_n^{eq}} \tag{6.69}$$

Using the Boltzmann distribution for the equilibrium concentration, Eq. (6.11), the expression for the nucleation rate becomes

$$J = c_s \left[\sum_{n=1}^{q-1} \frac{1}{\beta_n e^{\,-\Delta G(n)/kT}} \right]^{-1} \tag{6.70}$$

The next step is to evaluate the summatory within the brackets, which is not a trivial task in general. Fortunately, the existence of a pronounced maximum in the formation energy landscape $\Delta G(n)$ at $n^\star$ facilitates the procedure to obtain an accurate analytical approximation.

In order to do that, it is convenient first to pass from the discrete summatory to a continuum integral

$$J = c_s \left[\int_1^{q-1} \frac{\mathrm{d}n}{\beta_n \, e^{\,-\Delta G(n)/kT}} \right]^{-1} \tag{6.71}$$

The presence of a maximum allows the evaluation of the integral using the method of steepest descent. Essentially, the overwhelming contribution to the integral

comes, indeed, from the exponential around the maximum at the critical size $n^\star$, so we first expand $\Delta G(n)$ around $n^\star$:

$$\Delta G(n) \approx \Delta G(n^\star) + \frac{1}{2}\frac{\partial^2 \Delta G(n^\star)}{\partial n^2}(n - n^\star)^2 \tag{6.72}$$

noting that the contribution from the first derivative drops out because $\Delta G(n)$ has a maximum at $n^\star$. Using this expression in the integral we finally get

$$J = c_s\, \beta^\star\, Z\, e^{-\Delta G^\star / kT}, \tag{6.73}$$

where $\beta^\star = \beta_{n^\star}$ is the rate of attachment to the critical nucleus, $\Delta G^\star$ is the activation energy given by Eq. (6.43), and

$$Z = \sqrt{\left| \frac{\Delta G^{''}(n^\star)}{2\pi kT} \right|} \tag{6.74}$$

is the so-called Zeldovich factor, where $\Delta G^{''}(n^\star)$ is the second derivative of $\Delta G(n)$ evaluated at the critical nucleus. The Zeldovich factor is a relatively small correction that accounts for the ability of partial capsids around $n^\star$ to recross the barrier and dissolve back into the solution[8].

Therefore, we have arrived at the desired result that quantifies the rate of capsid formation, J, i.e., the number of capsids formed per unit volume and unit time. In essence, this rate depends on the rate of attachment of individual subunits to the critical size cluster, $\beta^\star$, and exponentially on the height of the nucleation barrier $\Delta G^\star$. Thus viral assembly should be regarded as an activated process where the rate of capsid formation, J, is given by an Arrhenius-like equation. Accordingly, the height of the energy barrier controls exponentially the velocity of the process, and could even abort the assembly at reasonable experimental (or host) time scales. Since the energy barrier depends on $\Delta\mu$, one can effectively reduce its size and trigger the formation of capsids by increasing the concentration of free subunits (see Figs. 6.7 and 6.8a).

It is important to mention that for typical *in vitro* experimental conditions, the total concentration of subunits c is constant. In this situation, the only real steady-state will correspond to the final aggregative equilibrium situation. In other words, the concentration of free subunits c_1 will constantly vary and the nucleation will be a time-dependent problem, whose particularities will be discussed in the next subsection.

[8]Indeed, this property is controlled by the curvature $\Delta G^{''}(n^\star)$. For shallow maxima the curvature is lower and more clusters will be able to cross back the barrier, which will reduce the rate J. Instead, for sharper maxima one will have the opposite behavior.

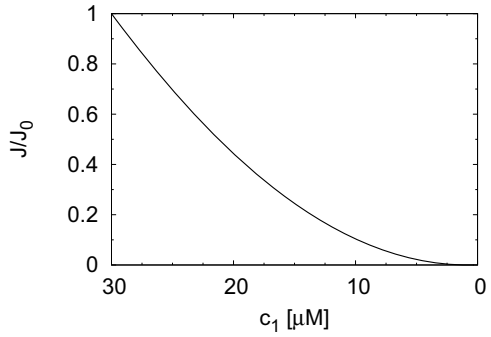


Figure 6.9: Rate of assembly at different free subunit concentrations. The rate of assembly, given by Eq. (6.73), is plotted for different concentrations of free subunits, starting at $c_0 = 30\mu M$. The curve corresponds to the assembly of capsids made of $q = 120$ subunits, with a line tension λ associated to a binding energy $|\Delta g| = 15\, kT$, given by Eq. (6.35). The values of the rate are normalized by the initial rate $J_0 = J(c_0)$.

6.4.2 Time-dependent self-assembly

When the concentration c is constant, the formation of new capsids reduces the concentration of free subunits c_1, which modifies in turn both the nucleation barrier as well as the subsequent rate of formation of capsids Fig. 6.9. To deal with this common time-dependent scenario, one could solve numerically the set of q master equations given by Eq. (6.57). Nevertheless, there is a simpler strategy that involves the use of an adiabatic approximation. In this approximation the assembly reaction is considered to be *slow* enough to be characterized by a *quasi-steady* state rate, $J(c_1(t))$, which is given by Eq. (6.73), and that will change depending on the instantaneous concentration of free subunits $c_1(t)$.

The total concentration of subunits c is conserved during the assembly process, and this constraint can be expressed as

$$c = \sum_{n=1}^{q} n\, c_n(t). \tag{6.75}$$

As discussed earlier, intermediates are in general rare because of the energetic penalty associated to the presence of a rim. Therefore, it is reasonable to consider that subunits are essentially either free in solution, $c_1(t)$, or in complete capsids,

$c_q(t)$. Then the constrain for c reduces to

$$c \simeq c_1(t) + q\, c_q(t). \tag{6.76}$$

so that, at any time during the assembly, the concentration of free subunits can be estimated by

$$c_1(t) = c - q\, c_q(t) \tag{6.77}$$

By assuming that the system is locally in a steady-state at a given concentration $c_1(t)$, then the rate of formation of new complete capsids will depend on $c_q(t)$ and will be given by

$$J(c_q(t)) = \frac{\partial c_q(t)}{\partial t} \tag{6.78}$$

Thus the concentration of capsids as a function of time can be evaluated by solving the following implicit equation

$$t - t_0 = \int_0^{c_q(t)} \frac{\mathrm{d}c_q(t)}{J(c_q(t))} \tag{6.79}$$

which is obtained by integrating Eq. (6.78). In the integration we have considered that no complete shells are present in the solution at the beginning of the experiment, $c_q(t_0) = 0$.

Fig. 6.10 plots the evolution in time of the concentration of complete capsids for $q = 72$, obtained from the numerical solution of the set of master equations, Eq. (6.57), and by the "adiabatic approximation", Eq. (6.79). We can observe a clear sigmoidal kinetics and that the approximation reproduces accurately the expected production of complete capsids.

When does viral assembly start?

It is very important to emphasize that self-assembly does not initiate at phase equilibrium conditions, i.e., when the chemical potentials of a subunit in the capsid and free in solution are equal, $\Delta\mu = 0$. At those conditions, although the formation of a complete capsid is energetically favorable, there is still a very high energetic barrier to overcome that will make the process unfeasible (see Fig. 6.7). For higher free subunit concentrations, the energy of formation of a partial capsid is reduced, which lowers and enables to initiate the process. Therefore, the self-assembly will start at a concentration c_1^{start} larger than the critical equilibrium or threshold concentration c^*. This might have important biological implications, since this mechanism may warrant that enough capsid

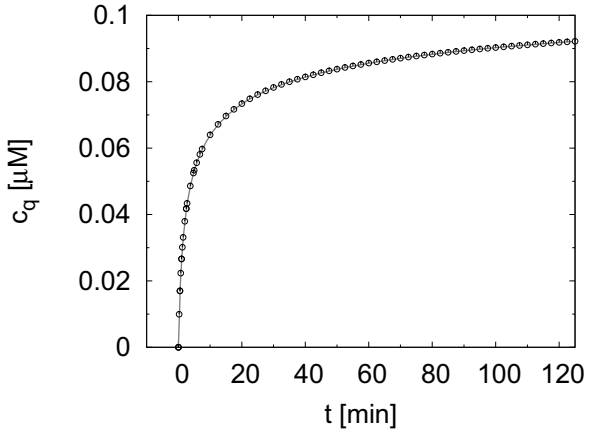


Figure 6.10: Time dependent production of capsids. The concentration of capsids increases with time along the assembly. The dots correspond to the solution of the master equations, Eq. 6.57, while the curve is the theoretical prediction by the "adiabatic approximation" given by Eq. (6.79). The plot has been calculated for typical assembly conditions for capsids made of $q = 72$ subunits, with a critical concentration of $c_{cmc} = 0.2\mu M$, a line energy $\alpha \approx 2\,kT$, and a rate of attachment $\approx 70 Hz/\mu M$. The initial concentration was $c_1(0) = 10\mu M$.

proteins are produced before the assembly starts, thus preventing, in general, the malformation of incomplete capsids due to the depletion or the lack of enough free subunits. In addition, it could facilitate the regulation of the process by a seed or a signal related to a change in the conditions.

Remarkably, even at supersaturation conditions where the formation of capsids becomes feasible, the assembly process will not start instantaneously. At the beginning, the system will require a certain time to overcome the nucleation barrier $\Delta G(n^\star)$ from scratch, and to reach a quasi-steady state rate. During this time lapse no complete capsids will be produced. This *lag time* can be estimated as [71]

$$\tau \sim \frac{1}{4\pi\beta^\star Z^2} \tag{6.80}$$

Therefore, it is inversely proportional to the square of the Zeldovich factor Z^2 and to the rate of attachment of subunits to the critical nucleus $\beta^\star$, which scales linearly with the concentration of free subunits c_1. The existence of lapse of time has been reported in experiments and simulations [15, 16].

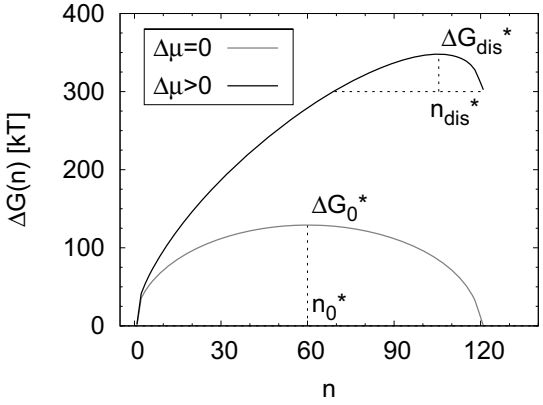


Figure 6.11: Free energy of formation of a partial capsid (disassembly). The two lines correspond, respectively, to phase equilibrium, $\Delta\mu = 0$, and to an undersaturated scenario, $\Delta\mu > 0$. The total capsid has $q = 120$ subunits and the energy scale for the line energy is $\alpha \approx 2.15\, kT$, Eq. (6.37.)

When does viral assembly stop?

Once the assembly has started, the successive production of capsids will reduce the concentration of free subunits. This will progressively increase the size of the barrier to overcome, and will affect the rate of formation of new capsids (see Fig. 6.9). Eventually, the rate will be so slow that the number of capsids will appear stable over the experimental time scale, and this will constitute an apparent end of the assembly.

So in practice, the end of the assembly will take place when the concentration of free subunits drops below a critical value c_1^{stop}. This concentration will be smaller than the one triggering the assembly, but generally larger than the critical equilibrium concentration, i.e., $c^\star < c^{stop} < c^{start}$. Therefore, in the nucleation framework the self-assembly of capsids does not stop necessarily when the reaction reaches the *phase equilibrium*, $\Delta\mu = 0$, but rather when the rate of formation becomes experimentally unobservable, which in general will still occur at supersaturated conditions, $\Delta\mu^{stop} < 0$.

With these ideas in mind, let's now discuss the inverse process of disassembly, and the associated hysteresis.

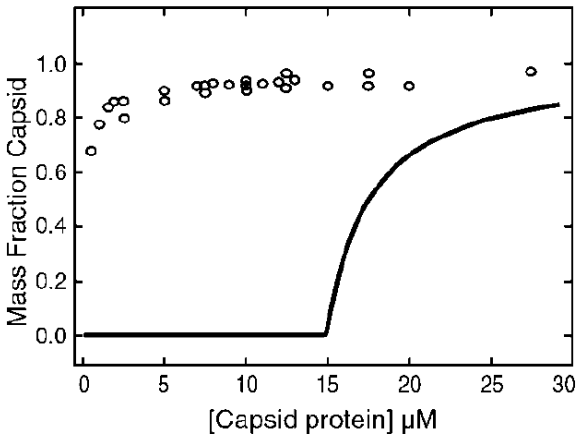


Figure 6.12: Hysteresis of HBV assembly. Mass fraction of proteins in capsids as a function of the total capsid protein concentration in a disassembly experiment, where purified capsids of HBV were diluted at controlled conditions, and the dissociation was assayed by size exclusion chromatography (SEC) (open circles). The assembly isotherm (solid line) was predicted for the same conditions, based on previous assembly experimental studies [13]. Taken from Ref. [55]

6.5 Disassembly of capsids and hysteresis

In an assembly experiment, one starts from only free subunits in solution and monitors the formation of capsids. Conversely, in a disassembly experiment, initially one only has complete capsids and monitors how they disaggregate back into free subunits. If the assembly of capsids would be a truly equilibrium process with no energy barrier, rather than an activated process, the assembly and disassembly would occur at the same concentrations, and the same master curve would describe both processes. However, this is usually not the case. For instance, Fig. 6.12, extracted from Ref. [55], represents the final concentrations of capsids as a function of the total initial concentration of proteins for assembly and disassembly experiments of HBV capsids, where all other conditions are the same. The different behavior in both situations is what is referred to as hysteresis in capsid assembly, and it can be explained in the context of the nucleation theory of viral capsids.

6.5.1 The rate of capsid disassembly

The thermodynamic and kinetic theory introduced in the previous sections is also able to account for capsid disassembly. To do so, one just has to interpret the dissociation of a shell as the inverse process of capsid formation. Therefore, let's consider a system with a total concentration of subunits c, where only fully-formed capsids are initially present, so $c_q(t=0) = c/q$.

Now, the dissociation of capsids requires to jump the free energy barrier in the opposite direction of assembly, i.e., from right to left in Fig. 6.11. The height of this activation barrier for disassembly will be now given by

$$\Delta G^{\star}_{dis} = \Delta G(n^{\star}) - \Delta G(q) \tag{6.81}$$

where $n^{\star}$ is the critical cluster size, whose expression is the same derived in the assembly scenario, Eq. (6.44). Thus the activation energy for dissociation will be just given by

$$\Delta G^{\star}_{dis} = \frac{q}{2}\left(\sqrt{\Delta\mu^2 + \alpha^2} - \Delta\mu\right) , \tag{6.82}$$

Notice that, to lower the barrier towards disassembly, undersaturated conditions, $\Delta\mu > 0$, will be required, rather than the supersaturated scenario, corresponding to $\Delta\mu < 0$, discussed in the assembly case.

Now, having characterized the thermodynamics of disassembly, let's turn to the kinetics. The set of master equations given by Eq. (6.57) can also describe the dissociation process, using the proper boundary conditions. Again, to get some analytic insight, let's focus first on the steady-state limit.

For disassembly, the right boundary condition for the rate in Eq. (6.68) will be $c_1/c^{eq} = 1$ and $c_q/c_q^{eq} \gg 1$, since disassembly will only be possible when the concentration of capsids in solution is much bigger than the corresponding at equilibrium[9].

If during the disassembly we assume that the boundary conditions are kept constant, one can derive from Eq. (6.68) a good approximation for the steady-state rate of disassembly as

$$J_{dis} \simeq -\left(\frac{c_q}{c_q^{eq}}\right) \Big/ \sum_{n=1}^{q-1} \frac{1}{\beta_n c_n^{eq}} \tag{6.83}$$

[9]The concentration of free subunits c_1 determines the total work of formation of a partial capsid, $\Delta G(n)$, or equivalently, the constrained equilibrium concentrations of species n, which were used to derive the rate of detachment α_n, Eq. (6.63). Therefore, even for negligible c_1 concentrations the left boundary condition of the rate in Eq. (6.68) will be by construction $c_1/c^{eq} = 1$.

By using the Boltzmann distribution for c_n^{eq}, Eq. (6.11), we get

$$J_{dis} = -c_q \left[\sum_{n=1}^{q-1} \frac{1}{\beta_n e^{-\Delta G_{dis}(n)/kT}} \right]^{-1} \tag{6.84}$$

where $\Delta G_{dis}(n) = \Delta G(n) - \Delta G(q)$. Finally, applying the steepest-descent approximation around the maximum at $n^\star$ in the integral representation of the summatory, one obtains

$$J_{dis} \simeq -c_q \, \beta^\star Z \, e^{-\Delta G^\star_{dis}/kT} \tag{6.85}$$

where Z is the Zeldovich factor given again by Eq. (6.74), since $\Delta G''(n^\star) = \Delta G''_{dis}(n^\star)$ because both energies of formation differ only by an additive constant.

Therefore, we have arrived at the desired result that quantifies the rate of capsid disassembly J_{dis}, i.e., the number of capsids disassembled per unit volume and unit time. Interestingly this dissociation rate is completely analogous to the assembly rate J_{as} given by, Eq. (6.73), except for the negative sign (indicating the disappearance of complete capsids) and the prefactor. So both, the assembly and disassembly scenarios, are characterized by Arrhenius-like laws, where the activation energy plays a crucial role.

6.5.2 The origin of hysteresis

The existence of an activation energy controlling the rate of both assembly and disassembly processes is the fundamental origin for the hysteresis observed in capsids.

More precisely, we have seen that capsid assembly cannot take place in general at phase equilibrium conditions, i.e., $\Delta\mu = 0$, since in that case the activation barrier is so high that becomes insurmountable. Thus, to make capsid formation feasible, we had to start from a concentration of subunits much larger than the threshold c^*, corresponding to $\Delta\mu < 0$, to lower the nucleation barrier and facilitate its crossing. Similarly, to induce capsid disassembly, we now have to lower the *disassembly barrier* $\Delta G^\star_{dis}$, and this can only be achieved by having undersaturated conditions corresponding to $\Delta\mu > 0$ (see Fig. 6.13).

Therefore, a successful assembly requires $\Delta\mu < 0$, but to have dissociation one needs $\Delta\mu > 0$. Since the chemical potential difference depends on the concentration of free subunits, and on the total concentration, that means that the concentrations at which assembly and disassembly will start (or stop) will be different. And that is in essence the origin of the hysteretic behavior observed experimentally.

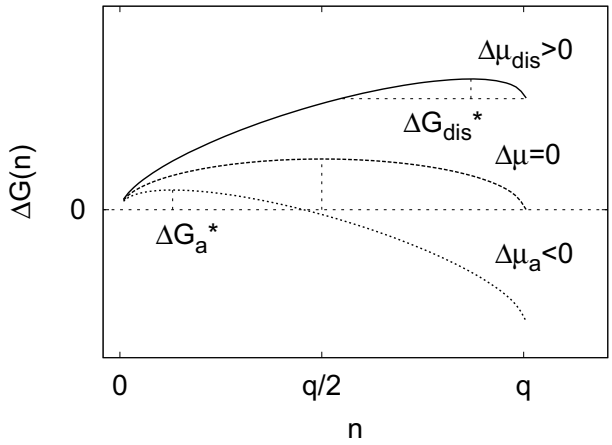


Figure 6.13: Free energy landscapes associated to the hysteresis. The energy of formation of a partial capsid at phase equilibrium, $\Delta\mu = 0$, and in the kinetically stopped scenarios of assembly, $\Delta\mu_a < 0$, and disassembly $\Delta\mu_{dis} > 0$, are illustrated. The vertical dashed lines indicate the barrier energies that have to be surmounted in each case.

In an assembly experiment, if the total concentration of proteins is fixed, the barrier associated to the formation of capsids will increase in time due to the progressive depletion of free subunits. This will in turn slow down the rate of assembly, until the rate is at some point so low that the formation of new capsids becomes undetectable at a reasonable experimental time scale. The experiment will have then reached a state of *apparent equilibrium*, even though the conditions could still be far from the true equilibrium, determined by the law of mass action, Eq. (6.15).

Let's call $\Delta\mu_a^{stop}$ to the value of the chemical potential difference corresponding to this apparent equilibrium situation, for which the barrier ΔG_a^* is so high that the assembly rate J_a is negligibly small.

Now, if we perform a disassembly experiment, it is reasonable to assume that the process will stop and the new apparent equilibrium will be reached when the final rate for disassembly is similar to the same threshold value for the assembly, namely

$$J_a \simeq J_{dis} \tag{6.86}$$

Since the most important factor controlling both rates is the height of their

respective barriers (except for some logarithmic corrections coming from the prefactors), the rates will be similar when the heights of the activation barriers will be the same

$$\Delta G_a^\star \approx \Delta G_{dis}^\star \tag{6.87}$$

In addition, using Eqs. (6.43) and (6.82), this relation implies that

$$\Delta \mu_a \approx -\Delta \mu_{dis} \tag{6.88}$$

Finally, using the definition of the chemical potential in terms of the concentrations, Eq. (6.40), we get

$$(c^\star)^2 \simeq c_1^a \, c_1^{dis} \tag{6.89}$$

where c_1^a and c_1^{dis} are the apparent equilibrium concentrations of free subunits obtained at the end of the assembly and disassembly experiments, respectively. This relation is specially relevant, because c_1^a and c_1^{dis} are experimentally accessible quantities [13,55], so the critical concentration $c^\star$, or equivalently the average binding energy, given by Eq. (6.41) can be obtained. Thus combining assembly and disassembly experiments is an excellent strategy to obtain the thermodynamic properties of capsids.

To end this analysis, let's clarify why, despite the fact that both assembly and disassembly are out-of-equilibrium processes, they seem to verify a kind of law of mass action, which has led in some cases to misinterpret these experiments as an equilibrium situation.

6.5.3 The generalized law of mass action

As we have discussed along this chapter, *in vitro* experiments show that the assembly of viruses starts above a critical concentration of free subunits, which seems to follow a universal curve for different experiments [68] (see Fig. 6.4) similar to the law of mass action [13].

Interestingly, disassembly experiments of fully formed capsids lead to final concentrations of subunits that also seems to obey a law of mass action, but surprisingly with a much stronger dissociation constant than the expected from assembly experiments [55] (see Fig. 6.12). The emergence of these two different equilibrium-like scenarios has led to consider that viruses follow a *pseudo* law of mass action [35], and it has been proposed that conformational changes in the coat proteins that modify their effective binding interactions are responsible for this hysteresis between assembly and disassembly [55]. Here, we demonstrate that the apparent final concentrations of free subunits and capsids, discussed

above, obey, indeed, a sort of generalized law of mass action when the formation of a capsid is considered as an activated process. This means that no structural modifications are in principle required to explain the puzzling results of assembly and disassembly, even though their presence can also have repercussions in the dissociation and stability of viruses [72].

As discussed before, under favorable assembly conditions, the formation of new capsids stops when the concentration of free subunits reaches a particular value, which has an associated nucleation barrier that is too high to be overcome. We called the corresponding concentration c_1^a, and the chemical potential difference $\Delta\mu_a$. Obviously, in an assembly experiment, if the initial total concentration is $c < c_1^a$, the barrier will be even higher, and no capsid will form, so all proteins will remain free in solution. Instead, when $c > c_1^a$, the assembly will proceed until $c_1 \simeq c_1^a$. In addition, since we consider experiments with constant total number of subunits, the concentration of subunits in capsids at the end of the reaction will be

$$qc_q = c - c_1^a \tag{6.90}$$

As in Section 6.2.2, we can reexpress this relation as a function of the fraction of subunits in complete capsids f, Eq. (6.19), and rearranging some terms we get

$$1 - f = (c_1^a/c) \tag{6.91}$$

or equivalently

$$1 - f = (c_{stop}^\star/c) \tag{6.92}$$

where, for convenience, we have identified $c_{stop}^\star = c_1^a$. This last expression is completely analog to the universal law derived for true equilibrium in Eq. (6.25). Thus, $c_{stop}^\star$ plays the same role as $c^\star = c_{cmc}/2$ in an equilibrium aggregation process. More precisely, if we plot the concentrations of free subunits and complete capsids as a function of the total concentration, we will get exactly the same behavior as that predicted for the equilibrium micellization (see Fig. 6.4).

The important point here is that, by interpreting those experiments as equilibrium states, one only gets an apparent equilibrium constant, or, equivalently, and apparent binding energy, which is not the true thermodynamic equilibrium value for the capsid. In particular, $c_{stop}^\star$ and μ_a are intimately related by Eq. (6.40), and taking into account also Eq. (6.41) we have

$$c_{stop}^\star \simeq c_s e^{(-\Delta\mu_a + \Delta g)/kT} \tag{6.93}$$

or

$$c_{stop}^\star \simeq c_s e^{\Delta g_{as}^{app}/kT} \tag{6.94}$$

where we have identified the apparent binding energy of assembly as

$$\Delta g_{as}^{app} = \Delta g - \Delta \mu_a \tag{6.95}$$

Hence, the difference between the true equilibrium and the pseudo-law of mass action, respectively, Eqs. (6.27) and (6.93), is the shift in the free energy associated to the chemical potential excess $\Delta\mu_a$.

Evidently, a disassembly experiment will lead to an analogous situation. Thus one will recover again a law of mass action kind of behavior, controlled by another apparent energy

$$\Delta g_{dis}^{app} = \Delta g - \Delta \mu_{dis} \tag{6.96}$$

In the previous section, using a simple time scale argument, we concluded that the condition that leads to similar final rates was $\Delta\mu_{as} \simeq -\Delta\mu_{dis}$, Eq. (6.88). Therefore, comparing both apparent energies we finally obtain

$$\Delta g \simeq \frac{\Delta g_{dis}^{app} + \Delta g_{a}^{app}}{2}. \tag{6.97}$$

which corresponds to the true effective binding energy.

That means that we have to correct the interpretation of those experiments on virus assembly and disassembly that are kinetically stopped by the presence of an insurmountable activation barrier, rather than strictly equilibrated. In particular, according to our thermodynamic-kinetical model, the apparent energies per contact measured from association and dissociation experiments are both different from the true equilibrium binding energy. Nevertheless, a good estimate of the right binding energy can be obtained by simply averaging these two apparent binding energies, as far as both type of experiments are performed at the same physicochemical conditions.

6.6 Nucleation theorem: the size of the critical nucleus

In the context of nucleation, there is a very useful relation between the size of the critical cluster, the nucleation barrier, and the nucleation rate. This relation is known as the *nucleation theorem*, and was originally derived by Hill in the context of his small systems thermodynamics [73, 74], and then applied to experiments by Kaschiev [75].

In essence, the nucleation theorem establishes that

$$\left(\frac{\partial \Delta G(n)}{\partial \Delta \mu}\right)_{T,p} = n \tag{6.98}$$

and it holds in general for any n. This result is specially useful applied to the critical size n^*, due to its significance in the kinetics of assembly In that case we obtain

$$\left(\frac{\partial \Delta G^\star}{\partial \Delta \mu}\right)_{T,p} = n^\star \tag{6.99}$$

A remarkable aspect of this relation is that it is model independent, i.e., it does not depend on the details of the explicit model of line tension introduced in Section 6.3.

More importantly, the nucleation theorem, combined with the kinetics derived from classical nucleation theory, establishes a route to extract the critical size directly from the experimental measures of the capsid formation rate. In particular, from the expression of the steady state rate, Eq. (6.73), one gets

$$\frac{\partial \ln J}{\partial \Delta \mu} = \frac{\partial \ln \beta^\star}{\partial \Delta \mu} - \frac{n^\star}{kT} \tag{6.100}$$

In addition, from the definition of the chemical potential difference $\Delta\mu$, Eq. (6.40), one has

$$\frac{\partial \Delta \mu}{\partial \ln c_1} = -kT \tag{6.101}$$

and considering also the dependency of $\beta^\star$ on c_1 given by Eq. (6.60), we finally obtain

$$\frac{\partial \ln J}{\partial \ln c_1} = n^\star + 1 \tag{6.102}$$

Hence, in an assembly experiment, the slope of the rate of capsid formation with respect to the concentration of free subunits will be directly related to the size of the critical nucleus.

A very important remark is that, in the context of nucleation, the size of the critical nucleus is not unique, but depends on the supersaturation or, equivalently, on the concentration of free subunits. Therefore, this might justify the controversy regarding the distinct critical cluster sizes that have been reported for the same virus at different experimental conditions [15].

6.7 Conclusions

There are two main conclusions that can be drawn from the results of this chapter. The first one is that the formation of viral capsids is an activated process, rather than an equilibrium aggregation. And the second one, is that the classical

nucleation theory of capsids (CNTC) provides a solid framework for the understanding of the thermodynamics and kinetics of viral self-assembly. The CNTC description was already introduced in the seminal paper of Zandi *et al.* [27] but in the present chapter we have extended these ideas in different directions and used them to get new and interesting results. In particular, we have rederived the thermodynamics of viral assembly using a quasichemical approach, that complements the phase route used in Ref. [27], and facilitates its understanding. We have also worked out the kinetic description of the assembly, computing explicitly the rate of assembly for both the steady-state and time-dependent scenarios.

In this framework, we have shown that CNTC consistently explains the main phenomenology observed in assembly experiments. The most important feature of this description is the existence of an energy barrier for the formation of a capsid that originates from the energetic penalty associated to the missing binding interactions at the rim of a partially-formed capsid. It is precisely this energetic cost, that can be modeled as a line energy, that justifies the scarcity of intermediates observed in experiments. Moreover, the existence of a lag time at the beginning of the assembly process corresponds to the lapse required for the free subunits to overcome the activation energy and start the production of capsids. Both the height of this activation barrier and the size of the critical nucleus depend strongly on the concentration of free subunits. As the assembly proceeds, the height of the barrier and the associated assembly rate progressively decrease due to the consumption of free subunits and the fact that the total concentration of coat proteins remains constant. At a certain point, the rate becomes so low, that in practice the reaction stops and the systems reaches an apparent equilibrium, different from the aggregative equilibrium, but that can also be characterized by a pseudo law of mass action.

We have also extended the CNTC description to account for capsid disassembly. That has allowed us to explain the origin of the hysteresis between the conditions at which capsid assembly and disassembly takes place in experiments. Moreover, we have shown that a proper interpretation of the thermodynamics of these assembly and disassembly experiments can be used to extract the true binding energies for capsids.

Interestingly, the CNT model of viral capsids clarifies some controversies reported in the literature. For instance, the existence of a pseudo-law of mass action is fully compatible with a nucleation mechanism and it is a kinetic consequence, rather than an equilibrium trait; the hysteresis between the association and dissociation conditions of viral shells is a natural result of the presence of an energy barrier, and it is not necessarily related to conformational changes of coat

proteins; or, remarkably, the critical nucleus is the most unstable intermediate, and should be the most challenging specie to capture or monitor during an experiment or a simulation. In addition, its size depends on the concentration of free subunits, and it is neither an intrinsic property of the coat proteins interactions, nor a precise geometrical arrangement of them. Thus, it is natural to obtain different critical nuclei depending on experimental conditions.

The assembly of viral capsids, based on the nucleation mechanism described by CNTC, has associated general features that might have important biological consequences during different steps of the virus life cycle. In particular, in the absence of an energy barrier, the formation of intermediates could deplete the stock of coat proteins before even producing a single complete capsid. Instead, the existence of the barrier requires a higher concentration of subunits to activate the process, and destabilizes the partial capsids, which facilitates the formation of fully-formed capsids. Moreover, once a capsid is formed, the presence of the barrier will prevent the virus to disassemble even at conditions where the assembly would not be possible. This hysteresis phenomenon is of vital importance for viruses, since they usually have to resist different physiological environments before infecting a new host.

Beyond its descriptive power, the CNTC can also be used to predict, at least qualitatively, the thermodynamics and kinetics of the assembly of specific viruses. In particular, this might be specially useful to design experiments at conditions that will ensure the efficient production of capsids at reasonable time scales. Conversely, the understanding of the general principles that rule viral assembly could also facilitate the design of novel widespectrum anti-viral strategies, targeted to interfere or impede a proper assembly of the virus.

Appendices

A Boundary length $l(n)$

We derive here a general expression for the boundary length of a partial shell. The rim of a capsid vanishes only when the structure is closed, $l(q) = 0$, or in the virtual situation of a *partial* shell made of no subunits, $l(0) = 0$. Then, if we assume that the density of subunits in the partial capsid is constant during the growing process, by symmetry, the maximum length will take place for a half-formed capsid, $l_{max} = l(q/2)$. The simplest formulation for these constraints will be $l(n) = g[n(q-n)]^z$, where $z > 0$ and g is a prefactor including the boundary length scale. Now, let's consider the two characteristic length scales of the problem, i.e., the subunit and capsid radii, σ_0 and R, respectively. For a single subunit the length is related to its size $l(1) \sim \sigma_0$, which leads to $g \sim \sigma_0/q^z$ assuming $q \gg 1$ (a standard condition in capsids[10]). For a half capsid the length must be related to its radius $l(q/2) \sim R$ and one gets $g \sim R/q^{2z}$. It is also reasonable to assume that the density of subunits is constant during the assembly, and is fixed by the covering of subunits in the completed capsid $q\sigma_0^2 \sim R^2$. Applying this relation to the last g expression one obtains $g \sim \sigma_0/q^{2z-1/2}$, and by consistency with the first g we get $z = 1/2$. Consequently $g \sim R/q$ and the length of a partial capsid will be

$$l(n) = a\frac{R}{q}\sqrt{n(q-n)} \tag{A-1}$$

where a is a constant geometrical prefactor independent of n and the length scales. As we have shown in Section 6.3 for the limit of circular rims, $a = 4\pi$.

[10]The smallest number of subunits in a capsid is $q = 12$, which corresponds to $T = 1$ structures made of pentameric capsomers. However, if the fundamental free subunits are monomers, this small shell will be composed of $q = 60$ coat proteins.

References

[1] S. J. Flint, L. W. Enquist, V. R. Racaniello, and A. M. Skalka, *Principles of virology.* Washington, D.C. ASM Press cop., 2004.

[2] P. Forterre, "Defining life: the virus viewpoint," *Orig. Life Evol. Biosph.*, vol. 40, pp. 151–160, 2010.

[3] P. E. Prevelige, "Inhibiting virus-capsid assembly by altering the polymerisation pathway," *Trends Biotechnol.*, vol. 16, pp. 61–65, 1998.

[4] A. Zlotnick and S. J. Stray, "How does your virus grow? Understanding and interfering with virus assembly," *Trends Biotechnol.*, vol. 21, pp. 536–542, 2003.

[5] J. B. Bancroft, C. E. Bracker, and G. W. Wagner, "Structures derived from cowpea chlorotic mottle and brome mosaic virus protein," *Virology*, vol. 38, pp. 324–335, 1969.

[6] K. W. Adolph and P. J. G. Butler, "Studies on the assembly of a spherical plant virus (I. States of aggregation of the isolated protein)," *J. Mol. Biol.*, vol. 88, pp. 327–338, 1974.

[7] M. Cuillel, C. Berthet-Colominas, B. Krop, A. Tardieu, P. Vachette, and B. Jacrot, "Self-assembly of brome mosaic virus capsids. Kinetic study using neutron and X-ray solution scattering," *J. Mol. Biol.*, vol. 164, pp. 645–50, 1983.

[8] C. Berthet-Colominas, M. Cuillel, M. H. J. Koch, P. Vachette, and B. Jacrot, "Kinetic study of the self-assembly of brome mosaic virus capsid," *Eur. Biophys. J.*, vol. 15, pp. 159–168, 1987.

[9] D. M. Salunke, D. L. D. Caspar, and R. L. Garcea, "Polymorphism in the assembly of polyomavirus capsid protein VP1," *Biophys. J.*, vol. 56, pp. 887–900, 1989.

[10] Z. Xie and R. W. Hendrix, "Assembly in vitro of bacteriophage HK97 proheads," *J. Mol. Biol.*, vol. 253, pp. 74–85, 1995.

[11] B. K. Ganser, S. Li, V. Y. Klishko, J. T. Finch, and W. I. Sundquist, "Assembly and analysis of conical models for the HIV-1 core," *Science*, vol. 283, pp. 80–3, 1999.

[12] A. Zlotnick, R. Aldrich, J. M. Johnson, P. Ceres, and M. J. Young, "Mechanism of capsid assembly for an icosahedral plant virus," *Virology*, vol. 456, pp. 450–456, 2000.

[13] P. Ceres and A. Zlotnick, "Weak protein-protein interactions are sufficient to drive assembly of hepatitis B virus capsids," *Biochemistry*, vol. 41, pp. 11525–11531, 2002.

[14] S. Mukherjee, M. V. Thorsteinsson, L. B. Johnston, P. A. DePhillips, and A. Zlotnick, "A quantitative description of in vitro assembly of human papillomavirus 16 virus-like particles," *J. Mol. Biol.*, vol. 381, pp. 229–37, 2008.

[15] G. L. Casini, D. Graham, D. Heine, R. L. Garcea, and D. T. Wu, "In vitro papillomavirus capsid assembly analyzed by light scattering," *Virology*, vol. 325, pp. 320–7, 2004.

[16] C. Chen, C. C. Kao, and B. Dragnea, "Self-assembly of brome mosaic virus capsids: insights from shorter time-scale experiments," *J. Phys. Chem. A*, vol. 112, pp. 9405–12, 2008.

[17] P. E. Prevelige, D. Thomas, and J. King, "Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells," *Biophys. J.*, vol. 64, pp. 824–35, 1993.

[18] P. A. Thuman-Commike, B. Greene, J. A. Malinski, J. King, and W. Chiu, "Role of the scaffolding protein in p22 procapsid size determination suggested by T=4 and T=7 procapsid structures," *Biophys. J.*, vol. 74, pp. 559–568, 1998.

[19] K. N. Parent, A. Zlotnick, and C. M. Teschke, "Quantitative analysis of multicomponent spherical virus assembly: scaffolding protein contributes to the global stability of phage P22 procapsids," *J. Mol. Biol.*, vol. 359, pp. 1097–106, 2006.

[20] D. E. Kainov, S. J. Butcher, D. H. Bamford, and R. Tuma, "Conserved intermediates on the assembly pathway of double-stranded RNA bacteriophages," *J. Mol. Biol.*, vol. 328, pp. 791–804, 2003.

[21] H. Fraenkel-Conrat and R. C. Williams, "Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components," *Proc. Natl. Acad. Sci. USA.*, vol. 41, pp. 690–698, 1955.

[22] J. Bancroft and E. Hiebert, "Formation of an infectious nucleoprotein from protein and nucleic acid isolated from a small spherical virus," *Virology*, vol. 32, pp. 354–356, 1967.

[23] Y. Hu, R. Zandi, A. Anavitarte, C. M. Knobler, and W. M. Gelbart, "Packaging of a polymer by a viral capsid: the interplay between polymer length and capsid size," *Biophys. J.*, vol. 94, pp. 1428–36, 2008.

[24] R. F. Bruinsma, W. M. Gelbart, D. Reguera, J. Rudnick, and R. Zandi, "Viral self-assembly as a thermodynamic process," *Phys. Rev. Lett.*, vol. 90, p. 248101, 2003.

[25] S. Katen and A. Zlotnick, *The thermodynamics of virus capsid assembly*, vol. 455, ch. 14, pp. 395–417. Elsevier Inc., 1 ed., 2009.

[26] D. Endres and A. Zlotnick, "Model-based analysis of assembly kinetics for virus capsids or other spherical polymers," *Biophys. J.*, vol. 83, pp. 1217–30, 2002.

[27] R. Zandi, P. van der Schoot, D. Reguera, W. Kegel, and H. Reiss, "Classical nucleation theory of virus capsids," *Biophys. J.*, vol. 90, pp. 1939–1948, 2006.

[28] M. F. Hagan and O. M. Elrad, "Understanding the concentration dependence of viral capsid assembly kinetics–the origin of the lag time and identifying the critical nucleus size," *Biophys. J.*, vol. 98, pp. 1065–74, 2010.

[29] 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*, 2009.

[30] J. Lepault, I. Petitpas, I. Erk, J. Navaza, D. Bigot, M. Dona, P. Vachette, J. Cohen, and F. A. Rey, "Structural polymorphism of the major capsid protein of rotavirus," *EMBO J.*, vol. 20, pp. 1498–507, 2001.

[31] S. Kanesashi, K. Ishizu, M. Kawano, S. Han, S. Tomita, H. Watanabe, K. Kataoka, and H. Handa, "Simian virus 40 VP1 capsid protein forms polymorphic assemblies in vitro," *J. Gen. Virol.*, vol. 84, pp. 1899–1905, 2003.

[32] T. Douglas and M. Young, "Viruses: making friends with old foes," *Science*, vol. 312, pp. 873–875, 2006.

[33] P. Ceres, S. J. Stray, and A. Zlotnick, "Hepatitis B virus capsid assembly is enhanced by naturally occurring mutation F97L," *J. Virol.*, vol. 78, pp. 9538–9543, 2004.

[34] P. van der Schoot and R. Zandi, "Kinetic theory of virus capsid assembly.," *Phys. Biol.*, vol. 4, pp. 296–304, 2007.

[35] A. Y. Morozov, R. F. Bruinsma, and J. Rudnick, "Assembly of viruses and the pseudo-law of mass action," *J. Chem. Phys.*, vol. 131, p. 155101, 2009.

[36] A. Y. Morozov and R. F. Bruinsma, "Assembly of viral capsids, buckling, and the Asaro-Grinfeld-Tiller instability," *Phys. Rev. E*, vol. 81, p. 041925, 2010.

[37] T. Keef, A. Taormina, and R. Twarock, "Assembly models for Papovaviridae based on tiling theory," *Phys. Biol.*, vol. 2, pp. 175–88, 2005.

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

[39] D. C. Rapaport, J. E. Johnson, and J. Skolnick, "Supramolecular self-assembly: molecular dynamics modeling of polyhedral shell formation," *Comput. Phys. Commun.*, vol. 121, p. 231, 1999.

[40] D. C. Rapaport, "Self-assembly of polyhedral shells: a molecular dynamics," *Phys. Rev. E*, vol. 70, p. 051905, 2004.

[41] D. Rapaport, "Role of reversibility in viral capsid growth: A paradigm for self-assembly," *Phys. Rev. Lett.*, vol. 101, pp. 1–4, 2008.

[42] D. C. Rapaport, "Modeling capsid self-assembly: design and analysis," *Phys. Biol.*, vol. 7, p. 045001, 2010.

[43] R. Schwartz, P. W. Shor, P. E. Prevelige, and B. Berger, "Local rules simulation of the kinetics of virus capsid self-assembly," *Biophys. J.*, vol. 75, pp. 2626–36, 1998.

[44] T. Zhang and R. Schwartz, "Simulation study of the contribution of oligomer/oligomer binding to capsid assembly kinetics," *Biophys. J.*, vol. 90, pp. 57–64, 2006.

[45] M. S. Kumar and R. Schwartz, "A parameter estimation technique for stochastic self-assembly systems and its application to human papillomavirus self-assembly," *Phys. Biol.*, vol. 7, p. 045005, 2010.

[46] M. F. Hagan and D. Chandler, "Dynamic pathways for viral capsid assembly," *Biophys. J.*, vol. 91, pp. 42–54, 2006.

[47] M. F. Hagan, "Controlling viral capsid assembly with templating," *Phys. Rev. E*, vol. 77, p. 051904, 2008.

[48] O. M. Elrad and M. F. Hagan, "Encapsulation of a polymer by an icosahedral virus," *Phys. Biol.*, vol. 7, p. 045003, 2010.

[49] H. D. Nguyen, V. S. Reddy, and C. L. Brooks, "Deciphering the kinetic mechanism of spontaneous self-assembly of icosahedral capsids," *Nano Lett.*, vol. 7, pp. 338–344, 2007.

[50] H. D. Nguyen and C. L. Brooks, "Generalized structural polymorphism in self-assembled viral particles," *Nano Lett.*, vol. 8, pp. 4574–4581, 2008.

[51] H. D. Nguyen, V. S. Reddy, and C. L. Brooks, "Invariant polymorphism in virus capsid assembly," *J. Am. Chem. Soc.*, vol. 131, pp. 2606–14, 2009.

[52] M. Hemberg, S. N. Yaliraki, and M. Barahona, "Stochastic kinetics of viral capsid assembly based on detailed protein structures," *Biophys. J.*, vol. 90, pp. 3029–42, 2006.

[53] A. W. Wilber, J. P. K. Doye, A. A. Louis, E. G. Noya, M. A. Miller, and P. Wong, "Reversible self-assembly of patchy particles into monodisperse icosahedral clusters," *J. Chem. Phys.*, vol. 127, p. 085106, 2007.

[54] I. G. Johnston, A. A. Louis, and J. P. K. Doye, "Modelling the self-assembly of virus capsids," *J. Phys.: Cond. Matt.*, vol. 22, p. 104101, 2010.

[55] S. Singh and A. Zlotnick, "Observed hysteresis of virus capsid disassembly is implicit in kinetic models of assembly," *J. Biol. Chem.*, vol. 278, pp. 18249 –18255, 2003.

[56] A. Zlotnick, J. M. Johnson, P. W. Wingfield, S. J. Stahl, and D. Endres, "Article a theoretical model successfully identifies features of hepatitis b virus capsid assembly," *Biochemistry*, vol. 38, pp. 14644–14652, 1999.

[57] A. Zlotnick, "Theoretical aspects of virus capsid assembly," *Journal of Molecular Recognition*, vol. 18, pp. 479–490, 2005.

[58] D. Endres, M. Miyahara, P. Moisant, and A. Zlotnick, "A reaction landscape identifies the intermediates critical for self-assembly of virus capsids and other polyhedral structures," *Protein science*, vol. 14, pp. 1518–1525, 2005.

[59] P. Moisant, H. Neeman, and A. Zlotnick, "Exploring the paths of (virus) assembly," *Biophys. J.*, vol. 99, pp. 1350–7, 2010.

[60] C. Uetrecht, C. Versluis, N. R. Watts, W. H. Roos, G. J. L. Wuite, P. T. Wingfield, A. C. Steven, and A. J. R. Heck, "High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, pp. 9216–20, 2008.

[61] J. R. Castón, S. A. Ghabrial, D. Jiang, G. Rivas, C. Alfonso, R. Roca, D. Luque, and J. L. Carrascosa, "Three-dimensional structure of penicillium chrysogenum virus: A double-stranded RNA virus with a genuine T=1 capsid," *J. Mol. Biol.*, vol. 331, pp. 417–431, 2003.

[62] J. M. Johnson, J. Tang, Y. Nyame, D. Willits, M. J. Young, and A. Zlotnick, "Regulating self-assembly of spherical oligomers," *Nano Lett.*, vol. 5, pp. 765–70, 2005.

[63] C. Uetrecht, I. M. Barbu, G. K. Shoemaker, E. van Duijn, and A. J. R. Heck, "Interrogating viral capsid assembly with ion mobility-mass spectrometry," *Nature chemistry*, vol. 3, pp. 126–32, 2011.

[64] A. Rusanov, F. Kuni, and A. Shchekin, "Phase and aggregative characterization of micellar systems," *Colloids and surfaces A*, vol. 128, pp. 13–16, 1997.

[65] F. M. Kuni, A. I. Rusanov, A. K. Shchekin, and A. P. G. A.I., "Kinetics of aggregation in micellar solutions," *Russian J. Phys. Chem.*, vol. 79, pp. 833–853, 2005.

[66] L. D. Landau and I. M. Lifshitz., *Theory of elasticity.* Pergamon, London, 1986.

[67] J. N. Israelachvili, *Intermolecular and surface forces.* Academic Press, 2010.

[68] W. K. Kegel and P. van der Schoot, "Competing hydrophobic and screened-coulomb interactions in hepatitis B virus capsid assembly," *Biophys. J.*, vol. 86, pp. 3905–13, 2004.

[69] H.-X. Zhou and M. K. Gilson, "Theory of free energy and entropy in noncovalent binding," *Chem. Rev.*, vol. 109, pp. 4092–107, 2009.

[70] S. Chandrasekhar, "Stochastic problems in physics and astronomy," *Rev. Mod. Phys.*, vol. 15, pp. 1–89, 1943.

[71] K. F. Kelton and A. L. Greer, *Nucleation in condensed matter: applications in materials and biology*. Pergamon, 2005.

[72] A. C. Steven, J. B. Heymann, N. Cheng, B. L. Trus, and J. F. Conway, "Virus maturation: dynamics and mechanism of a stabilizing structural transition that leads to infectivity," *Curr. Opin. Struc. Biol.*, vol. 15, pp. 227–36, 2005.

[73] T. L. Hill, "Thermodynamics of small systems," *J. Chem. Phys.*, vol. 36, pp. 3182–3197, 1962.

[74] T. L. Hill, *Thermodynamics of small systems: Parts I and II*. W. A. Benjamin Inc., New York, 1963 and 1964.

[75] D. Kashchiev, "On the relation between nucleation work, nucleus size, and nucleation rate," *J. Chem. Phys.*, vol. 76, pp. 5098–5102, 1982.

Chapter 7

Simulation of the Constrained Assembly of Spherical Shells

7.1 Introduction

As we have seen in the previous chapter, the assembly (and disassembly) of viral shells is an activated process, where subunits have to surmount an energy barrier to form a capsid. Although the height of the nucleation barrier is strongly controlled by the concentration of free subunits, its origin comes from the missing contacts at the rim of partial capsids, which is an intrinsic property independent on the concentration of subunits. In the CNT framework, to model this energy penalty we assumed a simple scenario where intermediates have circular boundaries with an associated line tension. This continuum hypothesis allowed us to make simple predictions and to illustrate the main consequences of CNT. Nevertheless, real capsids are discrete systems made of subunits, so it is natural to question whether these simplifying assumptions are accurate enough to describe, in particular, small and medium sized viral structures. But how do we test the accuracy of the continuum assumptions?

One obvious answer to the question would be to estimate the energy of formation of partial capsids directly from experiments. However, the presence of the barrier destabilizes the intermediates, lowering their effective concentrations. This precludes most of the partial structures to be detected experimentally [1, 2].

Alternative, computer simulations could provide the answer. Although *in silico* experiments based on atomistic or quasi-atomistic models are still too expensive computationally to study the assembly of viral capsids [3], more simplified models have already been able to reproduce some of the main features of viral-

like capsid assembly [4–23]. In particular, *in silico* experiments have shown that partial capsids usually grow by adding just a single subunit at a time, and among all possible intermediates, only a small set of them, which is associated to low free energies, contributes significantly to the pathways of assembly [21,23]. Therefore, in principle simulations could provide the required information to test the free energy model of partial capsids assumed in the CNT. However, this analysis has not been performed yet in any previous work.

Thus, our main goal in this chapter will be to study the assembly of viral capsids in order the analyze the mechanisms and the energy of formation of these shells. To that end, we will implement a constrained simulation (Section 7.2) using the physical model introduced in Part I. This will show that in fact the continuum model provides a good description of the assembly and accounts accurately for the intrinsic free energy of formation of partial capsids (Section 7.3). In addition, the simulations will unveil an unexpected phenomenon that promotes the premature closure of shells, and that can be understood in terms of the line tension of the rim (Section 7.4).

7.2 Simulations of spherical capsid formation

The generic capsomer-capsomer interaction model introduced in the first part of the thesis was able to explain successfully the final structures adopted by spherical viruses. So it is natural to use it also in this study. In particular, we will investigate the constrained assembly of this model for capsomers interacting on a spherical template of a fixed radius[1]. This surface will constrain the radial position of the capsomers that integrate the (partial) capsid, and might be interpreted as the spontaneous curvature imposed by protein-protein interactions for small capsids, the protein or membrane scaffold that supports the assembly of larger capsids, or the artificial template provided by nanodrops in recent self-assembly experiments [25].

We emphasize that this model is not adequate to simulate the actual assembly of capsids from *scratch*, i.e., starting from a solution of free subunits (mimicking a common experimental scenario[2]). Instead, we simply aim to characterize the structure and the intrinsic free energies of partial capsids. More precisely, we will

[1] In unconstrained conditions, the Lennard-Jones like potential used will form solid clusters, rather than shells [24].

[2] To properly simulate that situation it will be convenient to add an explicit term related to the directional interactions between subunits, associated to concepts like the bending energy and the spontaneous curvature.

focus on the constrained growth of a single shell, to validate the assumptions of CNT, and for which the actual model will be enough. For simplicity, we only explore the scenario with shells made of only one-type of capsomer, which was discussed in Section 2.2.4 for the equilibrium situation.

In particular, the simulation procedure used to model shell growth is as follows. One starts by placing one capsomer on a spherical template of radius R, which will constrain the radial position of the aggregated subunits. Next, a new capsomer is added at a randomly picked position on the surface, and the new particle is allowed to diffuse until it hits one of the previously added subunits. This cluster of subunits is then equilibrated by applying up to 10^4 Monte Carlo (MC) steps, which ensures the partial capsids to reach a low energy configuration in consonance with observations in more detailed simulations[3]. The first $0.5 \cdot 10^4$ MC are devoted to the equilibration of the cluster, where the second $0.5 \cdot 10^4$ MC steps were used to evaluate the statistical properties of the partial capsid, in particular its energy. Next, a new particle is inserted at a random position on the unoccupied part of the template and the process is continued until the capsid refuses the introduction of further particles after 10^5 random trials searching for remaining available space on the capsid (see Fig. 7.1). The maximum number of particles that can be accommodated on the sphere depends on the radius R of the template. We explored radii going from $R_{min} = 1.82\,\sigma_0$ ($N \sim 10$ subunits) to $R_{max} = 4.82\,\sigma_0$ ($N \sim 80$ subunits) at intervals of $\Delta R_{step} = 0.01\,\sigma_0$ in radius. Moreover, to allow different assembly pathways to compete in the formation of the shell, we repeat the assembly protocol 100 times for each value of the template radius R. Finally, when different structures compete, the results reported correspond to the statistically dominant structure. Using this procedure, the energies per capsomer in the final shells as well as the average energies of the different intermediates were characterized.

Let's first focus on the final structures. The average energy per capsomer obtained in the final structure of our constrained-assembly simulations is plotted in Fig. 7.2 as a function of R (solid line). For comparison, the optimal average energy per capsomer obtained in the equilibrium simulations of Section 2.2 (where the number of capsomers N was fixed and the radius optimized, given by $R_e(N)$) is shown as a discrete set of "+" symbols, connected by a dotted line. The energy per capsomer obtained in the assembly for fixed R is always higher (or

[3]We also performed *fast* assembly simulations allowing only the equilibration of the new attached capsomer. Except for very small capsids, this led to disordered structures. This situation will correspond at high rates of attachment (due to extreme supersaturation conditions) or strong binding interactions, which mislead the rearrangement of bonding interactions and lead to malformed capsids [11, 15].

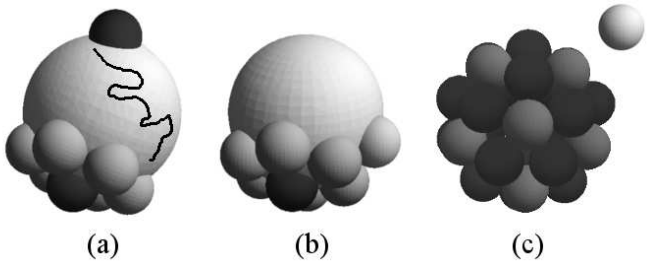


Figure 7.1: Sketch of the constrained assembly simulation. (a) Starting from one subunit, for every assembly growth-step, a new subunit is added randomly on the accessible surface of the spherical template, and rambles on the sphere until it hits the partial capsid (the black line illustrates a path). Subsequently, the new partial capsid is equilibrated using the MC algorithm (b). The process is repeated until a new subunit cannot fit on the template (c).

equal) than the minimum energy per subunit for $R = R_e(N)$, as it should. The shaded sectors in Fig.7.2 indicate the intervals $[R_a, R_c]_N$ of radii over which a shell structure is produced with the same number N and architecture. Remarkably, the final arrangement of subunits in these structures is the same as that of the optimal equilibrium all-pentamer capsids, which corresponded to Tammes Magic Number structures [26]. In particular, the $N = 12$ and $N = 32$ structures show icosahedral symmetry ($T = 1$ and $T^\star = 3$, respectively), while, for instance, the $N = 24$ and $N = 48$ structures have a more complex cubic symmetry. As the radius increases from R_a to R_c for given N, the symmetry of the structure does not change, but it carries an increasing level of elastic strain, as indicated by the steep rise in the energy per capsomer. With only one exception, the equilibrium radius $R_e(N)$ of a Magic Number structure is located outside the stability interval $[R_a, R_c]_N$ for constrained assembly. The exception is the $N = 12$ dodecahedral shell where the structure assembles near the optimal radii $R_a \simeq R_e(N = 12)$.

Therefore, in general, it becomes impossible to accommodate all the subunits N in a capsid having a fixed radii equal to the equilibrium one $R_e(N)$. The reason is that, at the last stages of the growth, a capillarity pressure associated to the open hole strains the capsid and promotes its premature closure. Hence, bigger radii are required to prevent this implosion and to assemble the structure with the right number of subunits (see Fig. 7.2). We will analyze this interesting phenomenon in more detail later. First, let's focus on the intrinsic energies of partial capsids before this implosion takes place.

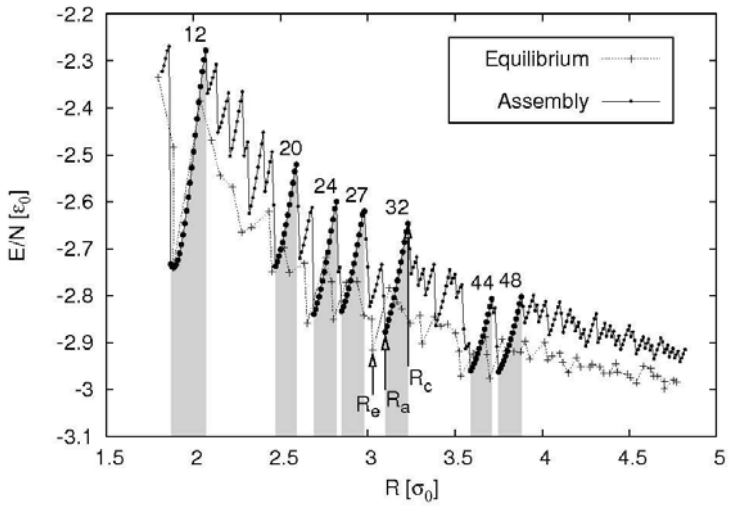


Figure 7.2: Energy per particle of the equilibrium shell structures (+ symbols joined by a dashed line), and of the constrained-assembly shells (dots joined by a solid line), both as a function of the template radius R expressed in units of σ_0. The bigger dots and shaded areas indicate the range R_a to R_c of sphere radii where the constrained assembly shells adopt a particular equilibrium "Magic Number" structure. As an example, the equilibrium radius $R_e(32)$ as well as $R_a(32)$ and $R_c(32)$ are indicated for $N = 32$.

7.3 The energy of formation of partial capsids

As shown in Section 6.3, the assembly energy of a partial shell can be expressed as

$$E(n) = \mu_e\, n + E_b(n) \tag{7.1}$$

where μ_e is the equilibrium energy per capsomer of the completed and unstrained shell, while $E_b(n)$ is defined to be the "boundary energy" of the partial shell.

As discussed in the CNT of viral capsids this boundary energy would be equal to the line tension λ times the perimeter length l, $E_b(n) = \lambda\, l(n)$. Within a continuum model, a partial capsid can be represented as a spherical cap with a circular rim, which is the shape that would minimize this line energy. An expression for the length of such boundary was already derived in Section 6.3, yielding $l(n) = (4\pi R/q_R)\sqrt{n(q_R - n)}$, where R is the radius of the capsid and q_R is the number of subunits that could be in principle accommodated in such a

capsid. Assuming a constant density of coat subunits in the growing phase, the area per subunit in any partial capsid will be the same as the equilibrium one given by

$$a_1 = \frac{4\pi R_e^2}{N} \tag{7.2}$$

Thus, the value of q_R for any radius within the stability region $[R_a, R_c]_N$ of the structure N, and before the closure, will be given by

$$q_R = \frac{4\pi R^2}{a_1} = N\left(\frac{R}{R_e}\right)^2 \tag{7.3}$$

For the equilibrium radius we recover $q_{R_e} = N$, but bigger radii can accommodate a larger total number of capsomers, i.e., $q_{(R>R_e)} > N$, as it should.

Combining these results, one gets

$$E_b(n, R) = \lambda \frac{4\pi R}{q_R}\sqrt{n(q_R - n)} \tag{7.4}$$

It is useful then to reexpress the previous result using a dimensionless line tension. From dimensional arguments, one expects the line tension λ to be proportional to μ_e/σ_0, the characteristic energy per assembly subunit divided by the characteristic length scale. This leads to a boundary energy

$$E_b(n, R) = \Lambda \frac{|\mu_e|}{\sigma_0}\frac{R}{q_R}\sqrt{n\,(q_R - n)} \tag{7.5}$$

where

$$\lambda = \Lambda \frac{|\mu_e|}{4\pi\sigma_0} \tag{7.6}$$

is the actual line tension, and Λ is a dimensionless line tension, which absorbs also the geometric constant prefactor. Let's try to estimate the value of this line tension for this particular model.

7.3.1 Line energy of a flat hexagonal sheet

To have a clearer insight of Λ, let's first discuss the line tension for a flat hexagonal lattice. One can compute the line tension by just splitting the hexagonal lattice and counting the energy cost of the lost contacts per unit length (see Section 6.3.1). However, this will depend on the direction of the cut. In this case,

one could think about two plausible cut directions. Cutting along the crystallographic axis, λ_{cry}, each capsomer, occupying a length $2\sigma_0$, will lose two associated contacts, $2\varepsilon_c$ (see Fig. 7.3a), so

$$\lambda_{cry} = \frac{1}{2}\frac{|\varepsilon_c|}{\sigma_0} \tag{7.7}$$

Equivalently, we can perform the same computation along the *zigzag* direction, λ_{zig}. In this situation four contacts, $4\varepsilon_c$, are lost every length step $3b$, where $b = 2\sigma_0/\sqrt{3}$ is the edge of the hexagon. Thus, one gets

$$\lambda_{zig} = \frac{1}{\sqrt{3}}\frac{|\varepsilon_c|}{\sigma_0} \tag{7.8}$$

which is slightly larger than λ_{cry}. The contact energy in a hexagonal lattice is directly related to the energy per capsomer μ_e. Each contact is shared by two capsomers, and each capsomer has six neighbors, so

$$\varepsilon_c = \frac{1}{3}\,\mu_e \tag{7.9}$$

Using this into λ_{cry} and λ_{zig}, and comparing with the expression for the boundary energy, Eq. (7.6), we get two reference values for the dimensionless line tension, Λ. One for the crystallographic direction

$$\Lambda_{cry} = \frac{2\pi}{3} \tag{7.10}$$

and the other for the zigzag cut

$$\Lambda_{zig} = \frac{4\pi}{3\sqrt{3}} \tag{7.11}$$

In the next subsection, the accuracy of this simple prediction will be tested. Our strategy is to first see whether Eq. (7.5) can be used to fit our numerical results for ordered capsids and, if so, to extract a value for Λ.

7.3.2 Comparison with the simulation results

In order to validate the proposed expressions for the boundary energy, average energies of all intermediates having n capsomers for a given radius, $E(n)$, were evaluated in the constrained assembly simulations. Thus, these values combined

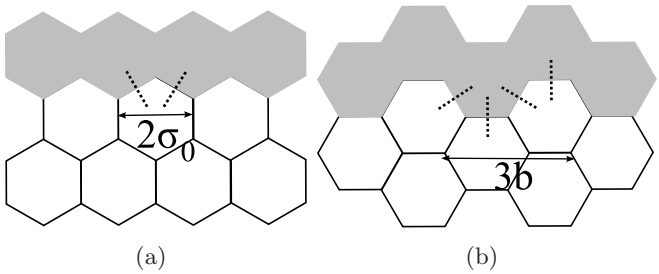


Figure 7.3: Line tension of a hexagonal lattice. (a) Cut along the crystallographic direction, the missing hexagons are colored in gray. For every step of length $2\sigma_0$ two contacts are broken (dashed lines). (b) Cut along the zig-zag direction. For every step of length $3b$ four contacts are lost (dashed lines), where $b = 2\sigma_0/\sqrt{3}$ is the edge of the hexagon.

with Eq. (7.1) determine the boundary energy $E_b(n)$, at a fixed radius R for any partial shell n of the final structure N as

$$E_b(n) = E(n) - \mu_e n. \tag{7.12}$$

It is important to emphasize that, for a given structure with N total capsomers, this $\mu_e(N)$ corresponds to the optimal value of the energy per capsomer obtained in the simulations of Chapter 2, and plotted in Fig. 7.2.

Now, one can fit Eq. (7.5), with Λ as the sole fitting parameter, to the values of the boundary energy E_b derived from the simulations for $n = 2$ up to $n = N-1$. In particular, Fig. 7.4 shows the boundary energies of a $T^\star = 3$ structure with $N = 32$ for the two stability region limits, i.e., $R = R_a(32)$ and $R = R_c(32)$. The theoretical curves are in good agreement, specially for the larger radius $R = R_c(32)$. The quality of this agreement is comparable for other N values, and is particularly accurate for the Magic Numbers (see Appendix A). We stress that the only relevant deviations occur at the last stage of the assembly, which is a consequence of the premature closure phenomenon that will be discussed later. More importantly, the values of Λ obtained from the fit for $N = 32$ are almost constant, ranging from $\Lambda_a(32) = 2.45 \pm 0.02$ to $\Lambda_c(32) = 2.47 \pm 0.01$, and are very close to the zigzag hexagonal reference $\Lambda_{zig} \approx 2.42$, as we can see in Fig. 7.5.

The discontinuities observed in Λ when comparing the values derived for the different assembly radii (see Fig. 7.5) track the jumps in Fig. 7.2, i.e., they are related to changes in the final number of subunits (and the symmetry) of the corresponding equilibrium shell structures. Thus, Fig. 7.5 illustrates that discreteness effects are important for the capillary action exerted on small partial

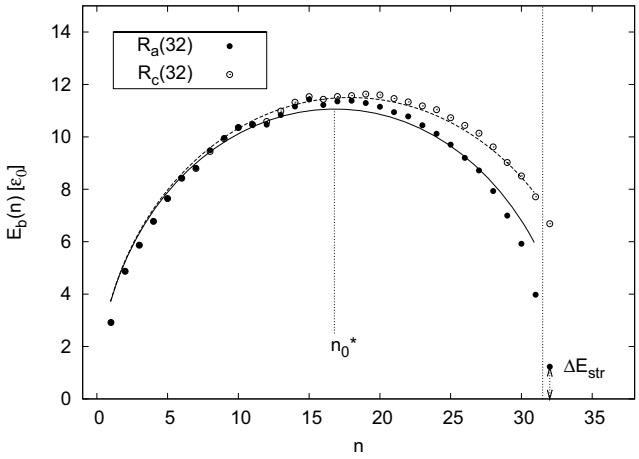


Figure 7.4: Boundary energy $E_b(n)$ of a partial shell, in units of ε_0, as a function of the number of assembly units n. Solid dots and open dots are the line energy results obtained in the simulation for the stability radius of $N = 32$, and $R = R_c(32)$ and $R = R_c(32)$, respectively. Solid and dashed lines show the results of a fit to Eq. (7.12). The elastic energy stored in the closed shell is (for $R = R_a(32)$) indicated as ΔE_{str}. The critical nucleus size corresponding to the top of the barrier is indicated by $n_0^\star$.

shells. Interestingly, the Magic Number shells have dimensionless line energies significantly higher than those of the less stable structures. The larger values of Λ for the Magic Number structures seem to indicate that Magic Number shells are exceptionally prone to implosion, as will be discussed later in detail. Interestingly, the values of Λ lay generally in the interval between the reference values Λ_{cry} and Λ_{zig}. In particular the values of the Magic Numbers structures and large shells are very close to Λ_{zig}. This indicates that the partial capsids of these structures have rims that on average may tend to adopt a zigzag configuration. For instance, a $T = 1$ will tend to grow around a 3-fold axis, while a $T^\star = 3$ will promote the capsid formation around a 5-fold axis[4].

Therefore, the line tension concept gives a good qualitative description of the boundary energy of partial capsids, despite the discreteness effects. More importantly, for the Magic Number structures and large shells the growth is

[4]This does not mean that a *penton* nucleates the assembly. We emphasize that, specially at the initial stages of the assembly, the partial capsid structure fluctuates and the disclinations, like pentons, are not necessarily stacked.

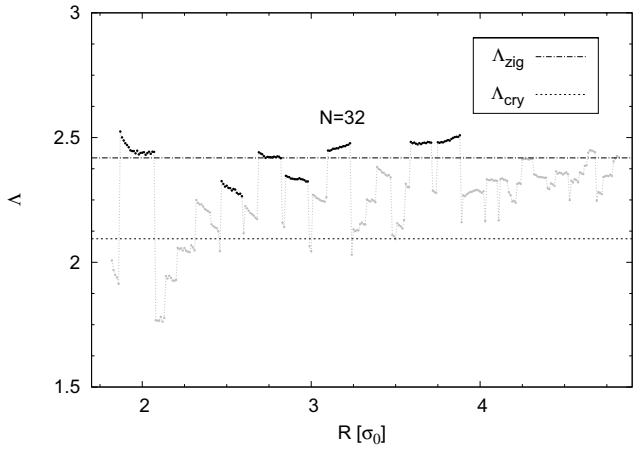


Figure 7.5: Dimensionless line tension. Grey dots correspond to the fitted values of Λ using Eq. (7.5) for all radii of assembly. Dots are connected as a guide for the eye. The black dots highlight the results for the Magic Numbers structures. The predictions for the dimensionless line tensions Λ_{cry} and Λ_{zig} for the two principal directions of an hexagonal lattice, given by Eqs. (7.10) and (7.11), respectively, are also plotted.

particularly well described by a circular rim approximation and a line tension associated to Λ_{zig}, and given by

$$\lambda_0 = \frac{1}{3\sqrt{3}} \frac{|\mu_e|}{\sigma_0} \tag{7.13}$$

7.4 Closure by implosion

As we have seen in Fig. 7.2, in general it is not possible to assemble a structure N at its radius of equilibrium $R_e(N)$. In particular, Fig. 7.6a shows a typical situation observed in the assembly of a shell at the equilibrium radius $R = R_e(32)$. At the last stages of the growth, a partial capsid of $n = 29$ subunits presents a hole that still would be able to accommodate a number of extra capsomers. However, when one more subunit is added, the hole undergoes an inward collapse (implosion) and the shell adopts a closed $N = 30$ structure. After this, it is not possible to increase the number of particles any further.

When the assembly radius is increased, $R > R_e(N)$, at some point the implosion can be avoided and we get an expanded version of the equilibrium structure

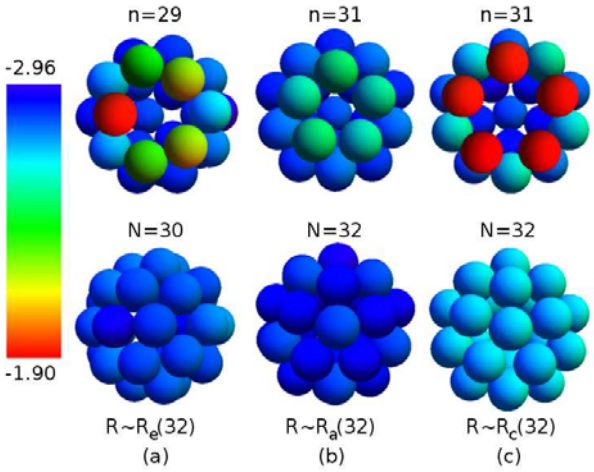


Figure 7.6: Implosion of incomplete shells. The assembly units are shown as spheres colored according to the energy scale shown on the left (in units of ε_0). (a) For a template with radius equal to the $N = 32$ equilibrium radius $R_e(32)$, the capsid closes prematurely at $N = 30$ due to inward collapse of the hole in the $N = 29$ structure. Note that the $n = 29$ hole still could accommodate more subunits. (b) For larger template radius (exceeding $R_a(32)$) this premature collapse no longer takes place and a uniformly stretched version of the equilibrium $N = 32$ icosahedral structure is produced. (c) As the template radius is further increased, shell closure proceeds by the collapse of the five-fold symmetric hole in the $n = 31$ structure when a subunit is inserted. If the radius is increased beyond $R_c(32)$, then a hole at the $n = 32$ shell remains stable, which allows to add more subunits leading to a $N > 32$ shell.

N. This defines the assembly radius $R = R_a(N)$. Figure 7.6b illustrates the successful assembly of the icosahedral shell $T^\star = 3$ at $R = R_a(32)$. We remark that this structure is strained, so it has a higher energy than the equilibrium one (see Fig. 7.2 and Chapters 3 and 4). Remarkably, the completion of the $N = 32$ structure proceeds around a 5-fold axis, which is in agreement with the structural insight provided by the dimensionless line tension analysis, $\Lambda(32) \simeq \Lambda_{zig}$, in the previous section.

If the template radius is increased, beyond $R = R_a(N)$, the structure N is recovered up to a limit radius $R = R_c(N)$. As an example, the assembly of $N = 32$ at $R = R_c(32)$ is shown in Fig. 7.6c. Interestingly, we observe that at $n = 31$ the remaining hole would be able to accommodate more than one

extra capsomer, but it prematurely closes at $n = 32$ leading to the icosahedral structure $N = 32$. Thus, in this case the implosion effect is responsible for the proper completion of the equilibrium structure N.

The range determined by $R_a(N)$ and $R_c(N)$ defines the stability region of assembly of the structure N, i.e., $[R_a, R_c]_N$. We remark that the lowest production of well-defined N capsids takes place at the assembly radius $R_a(N)$. Contrarily, at $R_c(N)$ the yield is extremely high (see Appendix B). The implosion effect shows a cyclic behavior along the assembly of structures. At the beginning of any stability region this phenomenon does not occur, although it is responsible of the alternative smaller capsids that reduces the yield; then as the radius increases it is responsible for properly close the N shell until $R_c(N)$ is reached; and, after that, it generally starts a new stability region for another N, where the implosion is not present initially.

The origin of this phenomenon is related again to the line tension of the rim in the partial capsids. Nevertheless, to properly describe the implosion effect present at the last stage of the assembly, it is necessary to extend the original continuum model given by Eqs. (7.1) and (7.5).

7.4.1 Hole implosion and the capillary pressure

The inward collapse observed in the assembly of spherical shells is driven by the formation of new bonds that can be produced by closing the shell, at the cost of stretching the existing bonds. We have seen that even for small and medium capsids the line tension concept is able to describe accurately the effect of the missing bonds on the rim of a partial capsid. But to properly study the closure effect, we must add a stretching term, E_{str} to the original line energy model,

$$E(n) = \mu_e n + \lambda l(n) + E_{str}(n) \tag{7.14}$$

Interestingly, hole collapse was reported in experimental studies of the closure of holes in lipid bilayer fluid shells [27]. In that case, the energetics of hole closure was described also by assigning to the perimeter of the hole a line tension. So it is useful to examine, in a similar way, the premature closure of a shell with a hole.

For the sake of simplicity, we focus on the simpler situation of hole collapse in a flat hexagonal sheet, thus ignoring the curvature effects in the capsid. Let's assume that a circular hole of area ΔA_0 has been removed from a flat and unstrained sheet of area A, and that the rim of the hole is subjected to a line tension

λ. Then, the stretching energy is given by

$$E_{str}(\Delta A) = \frac{1}{2} K \frac{(\Delta A_0 - \Delta A)^2}{A} \tag{7.15}$$

where, K is the 2D area modulus of the sheet, which is an intrinsic property. For a fixed number of units n, the line energy and the new stretching term are the only energies required to study the closure process. Their combination leads to the elastic energy

$$E_{el}(\Delta A) = \frac{1}{2} K \frac{(\Delta A - \Delta A_0)^2}{A} + 2\sqrt{\pi \Delta A}\, \lambda \tag{7.16}$$

where the length of the rim has been expressed in terms of the area ΔA of the hole, $\Delta A = \pi r^2$, where r is the hole radius.

Setting the derivative of $E_{el}(\Delta A)$ with respect to ΔA to zero gives

$$K \frac{\Delta A_0 - \Delta A}{A} = \lambda \sqrt{\frac{\pi}{\Delta A}} \tag{7.17}$$

This equates the radial elastic stress, $K(\Delta A_0 - \Delta A)/A$, exerted by the partial shell on the rim of the hole with the contractile capillary pressure, λ/r. For a given partial capsid n the values of K, λ, and A will be essentially constant. However, the hole of the unstretched partial shell, ΔA_0, will depend on the assembly radius R. Then, for radii where the hole is big enough, there are two solutions to Eq. (7.17): an energy minimum near $\Delta A \simeq \Delta A_0$ that corresponds to a mechanically stable state, and an unstable energy maximum for $\Delta A \ll \Delta A_0$ (see Fig. 7.7). The energy maximum separates the stable state from a boundary minimum at $\Delta A = 0$, that corresponds to the collapsed state. The energy maximum thus acts as an activation barrier against hole collapse. As the radius of assembly reduces, the unstretched hole ΔA_0 also reduces and the two solutions approach and merge at a critical hole area

$$\Delta A_0^\star = \left(\frac{3\sqrt{3\pi}}{2} \frac{\lambda}{K} A \right)^{2/3} \tag{7.18}$$

For shell radius corresponding to $\Delta A_0 \leq \Delta A_0^\star$, the hole is mechanically unstable and collapses (see Fig. 7.7).

This continuum description appears to account, at least qualitatively, for the observations of the simulations: for a certain radius once we reach the partial shell n with $\Delta A_0 \leq \Delta A_0^\star$, the structure implodes and precludes the formation of the expected structure with $N > n$; but when the radius is increased the hole remains

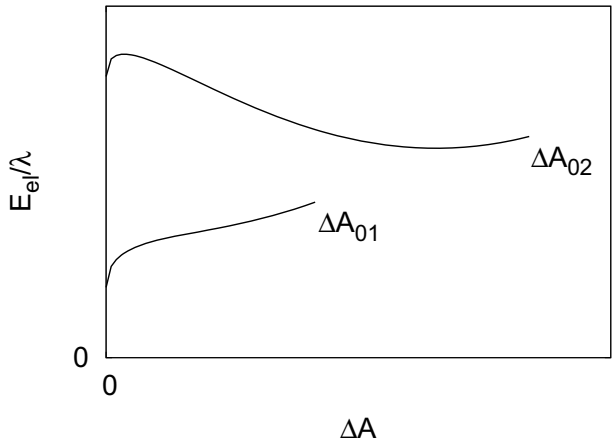


Figure 7.7: Capillary balance and implosion. Two characteristic scenarios related the elastic energy given by Eq. (7.16) are illustrated by plotting E_{el}/λ versus the variable hole area ΔA. The value $\Delta A = 0$ means that the capsid is closed. For a given partial shell n the term $K/\lambda A$ is constant. Then for an assembly radius R_1 an initial hole of area ΔA_{01} is unstable and leads to implosion in a downhill process. Instead, for a larger radius of assembly R_2, the initial hole ΔA_{02} will slightly compress to balance the capillary pressure and the strain of the shell, and will find a stable size. In this case, an energy barrier prevents the closure of the capsid.

stable, which allows the completion of N. If we keep increasing the radius we will repeat cyclically the same process for bigger shells. This justifies the existence of a stability region for bigger radii than the expected equilibrium radius of the structure. According to Eq. (7.18), the critical area of the hole required to prevent implosion grows as $R^{4/3}$ assuming $A \sim R^2$. So the collapse of partial shells is predicted to become increasingly pronounced as the shell radius grows, which explains why the $N = 12$ capsid was able to assemble at the equilibrium radius.

We notice that at the beginning of the assembly, the capillary pressure acts in the opposite direction and compresses the partial shell, rather than stretching it. In that situation one could also perform a similar analysis.

Finally, the fact that in Fig. 7.4, the fit for an assembly radius $R_a(32)$ deviates more than for $R_c(32)$ with respect to the theoretical curve at the end of the assembly could be easily explained in this context. It is clear now that this

originates because the hole at $n = 31$ is stable in both cases, but for the smallest radius the stretching will be higher, which is a term that was not considered in the original model. On the other hand, at the beginning of the assembly the compression of the partial shell produced by the capillary pressure of the rim affects both radii in a similar way, and is the origin of the small deviations respect the line energy curve in Fig. 7.4.

7.5 Conclusions

We have seen that the growth of highly-symmetrical shells can be properly described by continuum concepts. In particular, the line energy model introduced in CNT is surprisingly accurate to characterize the intrinsic energy of partial capsids, specially for the Magic Numbers structures. Moreover, the line tension associated to those architectures can be accurately estimated theoretically. The results of this chapter thus reinforce the applicability of CNT to describe the thermodynamics and kinetics of viral assembly.

Remarkably, a premature closure phenomenon related to the negative capillary pressure of the rim at the last stages of the assembly has been unveiled. This implosion effect precludes the formation of optimal capsids at their optimal radius, and shifts the assembly towards bigger radii. It is worth to notice that this same closure mechanism is responsible for the stability region of the structures, which is particularly wide for the Magical Numbers structures, due to the larger associated line tension.

The underlying physics of the implosion effect is related to the balance of the capillary pressure of the rim with the stretching of the partial shell. This can be analyzed in a continuum framework by a simple model of the line and stretching energies, which confirms and explains the observations of the simulations.

Interestingly, it follows from our findings that, except for the $T = 1$ shell, to properly form a complete capsid in a constrained assembly scenario (which seems common for large shells) it is convenient to impose a spontaneous radius bigger than the equilibrium one. This leads to a final capsid that is strained. So it suggests that post-assembly (or maturation) mechanisms would be required to shrink the radius of the shell to obtain the true equilibrium structure. In order to *stabilize* the structure and minimize the strain, viral capsids could rely on scaffold expulsion, cleavage of the coat proteins, or cross-linking of the subunits. Alternatively, as we saw in Chapter 5, the additional stress obtained by the strained shell could be useful to reinforce the mechanical response of the capsid. Therefore, from a physical standpoint, the maturation of viral capsids do not seem

to have a univocal motivation, and probably tunes the mechanical properties of the shells to optimize the infection strategy of each virus.

We finally stress that the closure phenomenon described above could also be present in unconstrained assembly scenarios. However, in those situations the proper closure of the shell could be facilitated by fluctuations of the radius.

Appendices

A Line energy fit

Fig. A.1 shows the line energies obtained from the simulation, using Eq. (7.12), for the shaded local minima in Fig. 7.2 (except for $N = 32$ that was already plotted in Fig. 7.4). The lines show the fitted curves (before the closure) using Eq. (7.5) and Λ as the only fitting parameter. The line tension model is in very good agreement even for small shells.

Fig. A.2 provides the correlation coefficient[5] R^2_{fit} of the line energy model Eq. (7.5) when fitted to the simulation results, using the NonlinearModelFit function implemented in Mathematica [28]. Except for some small shells and some transition radii, the quality of the fits is excellent, specially for the deep local minima structures.

[5] R^2_{fit} is given by $1 - SS_{mod}/SS_{tot}$, where SS_{mod} is the sum squares of the fitted model respect the data, and SS_{tot} is the total sum of squares of the data respect the average. Thus, for $R^2_{fit} = 1$ one obtains a perfect fit.

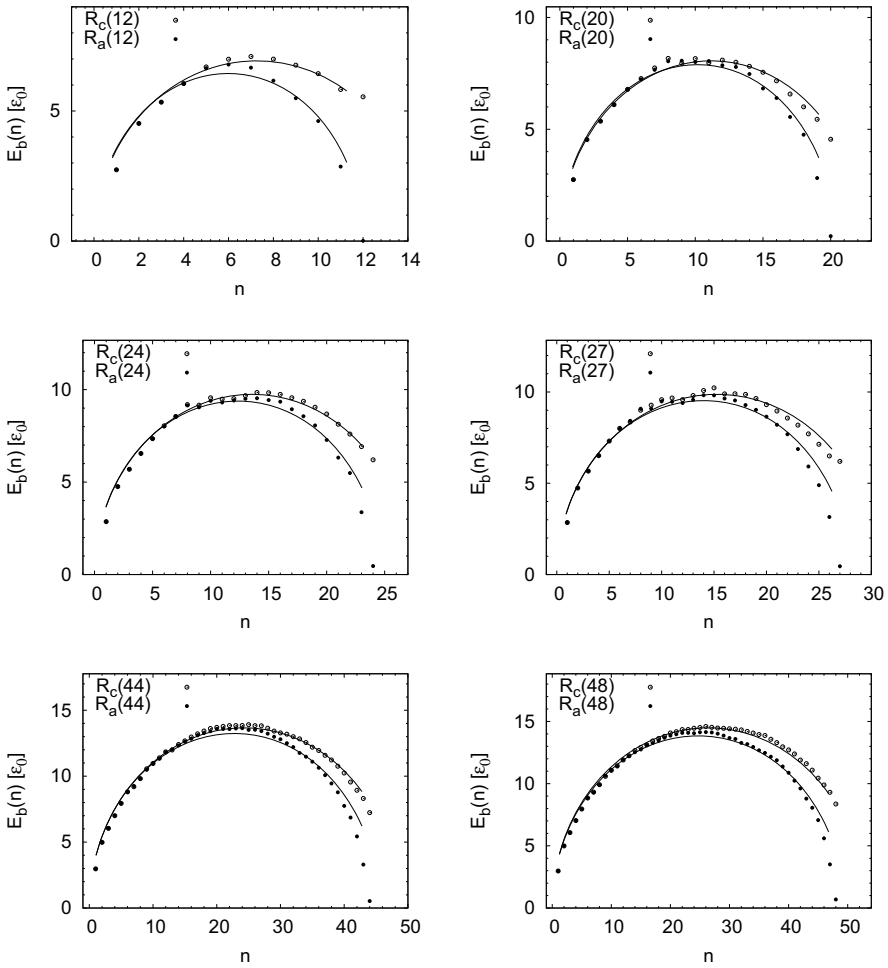


Figure A.1: Boundary energy $E_b(n)$ of a partial shell, in units of ε_0, as a function of the number of assembly units n. Solid dots: $R = R_a$; open dots: $R = R_c$. The lines show the results of the fits to Eq. (7.12). The panels correspond to the structures N that are deep local minima in Fig. 7.2 (except $N = 32$).

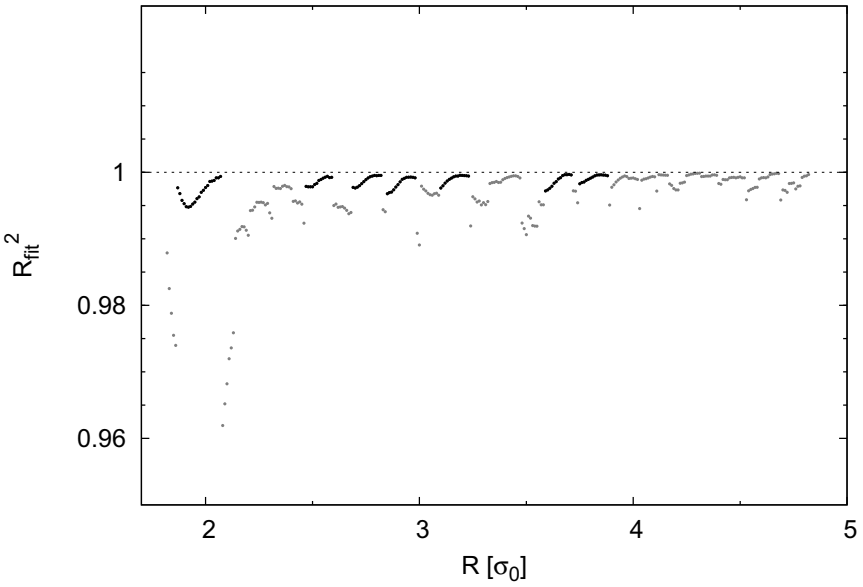


Figure A.2: Correlation coefficient R^2_{fit} of the line energy fit for all the assembly radii explored. The black dots highlight the deep local minima shade in Fig. 7.2.

B Yield of assembly

Fig. B.1 plots the production yield f_{yield} of assembly structures over the 100 simulations performed at each radius. The scale goes from 0 (no production) to 1 (maximum yield). When several structures compete at the same radius, we select the most efficient one. So the yield is always $f_{yield} \geq 1/2$. The structures related to optimal minima at equilibrium typically reach high yields over the whole stability regions $[R_a(N), R_c(N)]_N$. Notice that for these highly symmetrical structures the production is lower the closer one gets to the initial assembly radius $R_a(N)$.

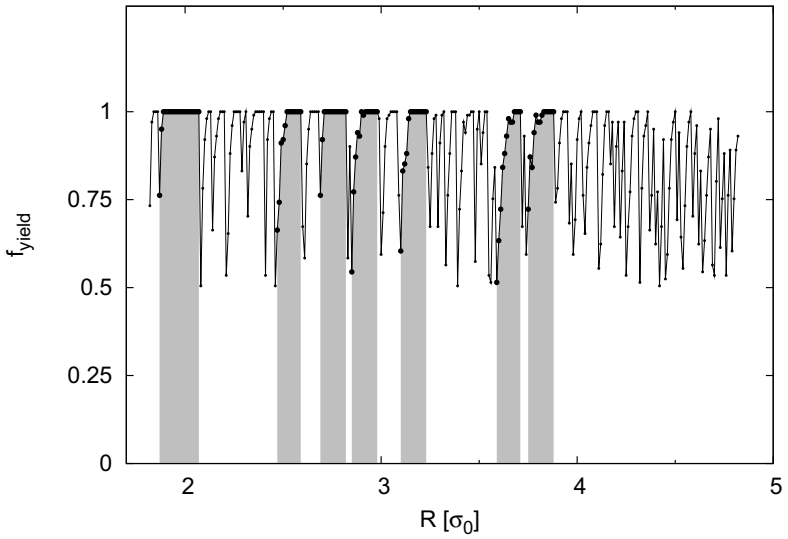


Figure B.1: Yields of assembly. The efficiency of capsid production f_{yield} is plotted for all the assembly radii. The yield is normalized to 1, which means that all attempts led to the same N structure. The shaded areas highlight the stability regions of the capsids related to optimal equilibrium structures, which are plotted with bigger dots.

References

[1] D. E. Kainov, S. J. Butcher, D. H. Bamford, and R. Tuma, "Conserved intermediates on the assembly pathway of double-stranded RNA bacteriophages," *J. Mol. Biol.*, vol. 328, pp. 791–804, 2003.

[2] C. Uetrecht, I. M. Barbu, G. K. Shoemaker, E. van Duijn, and A. J. R. Heck, "Interrogating viral capsid assembly with ion mobility-mass spectrometry," *Nature chemistry*, vol. 3, pp. 126–32, 2011.

[3] P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson, and K. Schulten, "Molecular dynamics simulations of the complete satellite tobacco mosaic virus," *Structure*, vol. 14, pp. 437–49, 2006.

[4] D. C. Rapaport, J. E. Johnson, and J. Skolnick, "Supramolecular self-assembly: molecular dynamics modeling of polyhedral shell formation," *Comput. Phys. Commun.*, vol. 121, p. 231, 1999.

[5] D. C. Rapaport, "Self-assembly of polyhedral shells: a molecular dynamics," *Phys. Rev. E*, vol. 70, p. 051905, 2004.

[6] D. Rapaport, "Role of reversibility in viral capsid growth: A paradigm for self-assembly," *Phys. Rev. Lett.*, vol. 101, pp. 1–4, 2008.

[7] D. C. Rapaport, "Modeling capsid self-assembly: design and analysis," *Phys. Biol.*, vol. 7, p. 045001, 2010.

[8] R. Schwartz, P. W. Shor, P. E. Prevelige, and B. Berger, "Local rules simulation of the kinetics of virus capsid self-assembly," *Biophys. J.*, vol. 75, pp. 2626–36, 1998.

[9] T. Zhang and R. Schwartz, "Simulation study of the contribution of oligomer/oligomer binding to capsid assembly kinetics," *Biophys. J.*, vol. 90, pp. 57–64, 2006.

[10] M. S. Kumar and R. Schwartz, "A parameter estimation technique for stochastic self-assembly systems and its application to human papillomavirus self-assembly," *Phys. Biol.*, vol. 7, p. 045005, 2010.

[11] M. F. Hagan and D. Chandler, "Dynamic pathways for viral capsid assembly," *Biophys. J.*, vol. 91, pp. 42–54, 2006.

[12] M. F. Hagan, "Controlling viral capsid assembly with templating," *Phys. Rev. E*, vol. 77, p. 051904, 2008.

[13] M. F. Hagan and O. M. Elrad, "Understanding the concentration dependence of viral capsid assembly kinetics–the origin of the lag time and identifying the critical nucleus size," *Biophys. J.*, vol. 98, pp. 1065–74, 2010.

[14] O. M. Elrad and M. F. Hagan, "Encapsulation of a polymer by an icosahedral virus," *Phys. Biol.*, vol. 7, p. 045003, 2010.

[15] H. D. Nguyen, V. S. Reddy, and C. L. Brooks, "Deciphering the kinetic mechanism of spontaneous self-assembly of icosahedral capsids," *Nano Lett.*, vol. 7, pp. 338–344, 2007.

[16] H. D. Nguyen and C. L. Brooks, "Generalized structural polymorphism in self-assembled viral particles," *Nano Lett.*, vol. 8, pp. 4574–4581, 2008.

[17] H. D. Nguyen, V. S. Reddy, and C. L. Brooks, "Invariant polymorphism in virus capsid assembly," *J. Am. Chem. Soc.*, vol. 131, pp. 2606–14, 2009.

[18] M. Hemberg, S. N. Yaliraki, and M. Barahona, "Stochastic kinetics of viral capsid assembly based on detailed protein structures," *Biophys. J.*, vol. 90, pp. 3029–42, 2006.

[19] A. W. Wilber, J. P. K. Doye, A. A. Louis, E. G. Noya, M. A. Miller, and P. Wong, "Reversible self-assembly of patchy particles into monodisperse icosahedral clusters," *J. Chem. Phys.*, vol. 127, p. 085106, 2007.

[20] I. G. Johnston, A. A. Louis, and J. P. K. Doye, "Modelling the self-assembly of virus capsids," *J. Phys.: Cond. Matt.*, vol. 22, p. 104101, 2010.

[21] T. Keef, A. Taormina, and R. Twarock, "Assembly models for Papovaviridae based on tiling theory," *Phys. Biol.*, vol. 2, pp. 175–88, 2005.

[22] D. Endres, M. Miyahara, P. Moisant, and A. Zlotnick, "A reaction landscape identifies the intermediates critical for self-assembly of virus capsids and other polyhedral structures," *Protein science*, vol. 14, pp. 1518–1525, 2005.

[23] P. Moisant, H. Neeman, and A. Zlotnick, "Exploring the paths of (virus) assembly," *Biophys. J.*, vol. 99, pp. 1350–7, 2010.

[24] D. Wales and J. Doye, "Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms," *J. Phys. Chem. A*, vol. 101, pp. 5111–5116, 1997.

[25] C. B. Chang, C. M. Knobler, W. M. Gelbart, and T. G. Mason, "Curvature dependence of viral protein structures on encapsidated nanoemulsion droplets," *ACS Nano*, vol. 2, pp. 281–286, 2008.

[26] D. A. Kottwitz, "The densest packing of equal circles on a sphere," *Acta Cryst. A*, vol. 47, pp. 158–165, 1991.

[27] E. Karatekin, O. Sandre, H. Guitouni, N. Borghi, P.-H. Puech, and F. Brochard-Wyart, "Cascades of transient pores in giant vesicles: Line tension and transport," *Biophys. J.*, vol. 84, pp. 1734–1749, 2003.

[28] I. Wolfram Research, *Mathematica edition: version 7.0.* Wolfram Research, Inc., 2008.

Conclusions and Perspectives

In this thesis we have tried to elucidate the general physical principles that play a major role in the morphology, stability, and assembly of viral capsids. The main conclusions of this work are summarized in the following:

In Part I of the thesis, we showed that the structure of viral capsids can be rationalized using basic geometrical rules and physical principles.

Thus, in Chapter 1, we developed a general theory that characterizes both spherical and bacilliform (or prolate) capsids under the same geometrical framework. This theory is the natural extension of the classical Caspar and Klug construction of icosahedral shells (CK) [1] and the prolate construction of Moody [2, 3], and enumerates all possible structures based on hemi-spherical icosahedral caps with axial symmetry. Our study indicates that the specific symmetry and T-number of the caps strongly constrain the possible architecture of the central bodies. Indeed, we have derived the precise geometrical rules that determine, for any structure, the discrete set of radii, lengths, helicity, and number and position of coat proteins. The theory also predicts the existence of isomers for some architectures, which could interfere with (or enrich) the assembly of certain viruses. More importantly, this general theory allows to infer the structure of viral capsids based on simple experimental data, thus facilitating a rapid cataloging of their structures. Accordingly, it might be useful to interpret different types of experiments, and could guide high-resolution studies of viral capsids.

Subsequently, in Chapter 2, we demonstrated that the structures derived in the previous geometrical study arise naturally from a simple physical model, which captures the basic ingredients involved in the interactions among viral capsomers. The results corroborated the theoretical selection rules, and pointed out that not all possible designs are equally favorable from an energetic standpoint. This confirms that icosahedral symmetry in elongated viral capsids is obtained spontaneously from the free energy minimization of a very generic interaction. Moreover, our model shows the possibility of building some non-icosahedral capped prolates that could also play a relevant role in the assembly of viruses.

The combination of both our geometrical and physical studies pave the way to design *artificial* elongated nanostructures with well defined properties.

In the second part of the thesis, we analyzed the role of the discrete nature of viral capsids and the organization of capsomers in the actual mechanical properties of shells.

We first introduced a simple theoretical framework, in Chapter 3, connecting the intermolecular interactions among coat subunits and the global elastic properties of the capsid. We particularized the study to spherical and polyhedral structures, observing that a buckling transition naturally occurs by the expansion of spherical capsids, in consonance with the maturation of several viruses. Furthermore, the polyhedral structures turned out to be more resistant than their spherical counterparts, as far as the intermolecular interactions remain unchanged. This could be a biological advantage for some viruses, especially those that rely on pressurization as an infection strategy. We stress that the principles involved in this theory could be easily adapted to other types of structures, and establishes a molecular basis to interpret the mechanical properties of viral capsids.

Then, in Chapter 4, the general mechanical properties of spherical and polyhedral capsids predicted by the theory in Chapter 3 were confirmed, using our simple interaction model. In addition, it was shown that the T-number (or more precisely the P-class) of the quasi-spherical structure plays a relevant role in the stability and local mechanical properties of the shell. We remark that this aspect of the study was elusive for both the traditional continuum elasticity theory and the mean field molecular theory introduced previously. In particular, $P = 1$ capsids resulted to be more prone to adopt a polyhedral shape, whereas $P = 3$ shells showed the opposite behavior. In the limit of large capsids the spherical morphology became less stable for any T than the polyhedral counterpart. The different tendencies to buckle are related to the local distribution of stress associated with the class P and could play a relevant role in the decapsidation, bursting, or local mechanical response of viral capsids. Therefore, we have shown that the characterization of capsids in terms of the class P, rather than just the T-number, could be particularly useful in order to elucidate common mechanical properties for different viruses.

Additionally, Chapter 5 unveiled some unexpected properties of bacilliform capsids. The study indicated that the non-uniform stress distribution of capsids, related to the discrete nature of the structure, plays a crucial role determining the mechanical properties of bacteriophage $\phi 29$ procapsids. More precisely, the existence of built-in stress present in those empty bacilliform shells is able to

reinforce their mechanical properties, inverting the classical anisotropic response expected from continuum elasticity theory. Interestingly, this behavior is similar to the pressurization of a prolate. Thus, the built-in stress could play an important biological role facilitating the accommodation of the pressure produced by the highly compacted dsDNA in the mature ϕ29 virion.

The self-assembly of viral capsids was explored in the third block. In particular, we showed that viral assembly and disassembly are activated processes controlled by a nucleation barrier.

The first chapter of Part III demonstrated how classical nucleation theory (CNT) can be adapted to explain not only the assembly [4] but also the disassembly of viral capsids. We focused on the case of spherical shells, being able to justify and characterize the hysteresis reported in several experiments. This phenomenon originates because the association and disassociation of capsids are activated processes, controlled by an energy barrier, and become kinetically stopped at typical accessible time-scales. Thus, hysteresis is a fundamental property of viral capsids, and will be present even in the absence of maturation processes. In addition, we indicated how experiments of assembly and disassembly can be combined in order to extract thermodynamic properties of capsids, like the average binding energy. We emphasized that the height of the energy barrier and the size of the critical nucleus depend on the concentration of free subunits, which clarifies some controversies reported in the literature. A special effort was made to express CNT in accessible experimental magnitudes to facilitate its application. Remarkably, the predictions of CNT could be a good starting point to guide the design of broad spectrum antivirals, aimed at interfering with the assembly or disassembly processes.

Finally, in Chapter 7 we confirmed that the underlying assumptions of CNT, are surprisingly good in characterizing the assembly of discrete shells, which were simulated using the physical model introduced in the previous parts. In particular, assumptions like the line tension or the use of an average binding energy were shown to be accurate to describe the thermodynamics of capsids. Furthermore, an interesting closure mechanism for spherical capsids was unveiled. This phenomenon originates due to the negative capillary pressure present in the rim of a partial shell at the last stages of the assembly. Surprisingly, this implosion effect precludes the constrained assembly of optimal structures at their respective optimum radius, but promotes a wide stability region for bigger radii, where perfect viral capsids are easily assembled. Interestingly, due to the implosion effect, the final shells are strained, which increases their built-in stress. Hence, this closure phenomenon opens up interesting questions regarding the physical

origin and biological motivations of virus maturation. In addition, the implosion effect could be exploited as a potential physical antiviral strategy to promote the malformation of capsids.

The general theoretical frameworks and physical principles developed in this thesis are by no means limited to the particular cases discussed. The flexibility of the ideas presented may become very useful to the systematic description of more complex situations. In this sense our findings may open up new research lines to carry out future investigations. Concerning these perspectives, we can quote some short-term objectives and broader generalizations.

One of our first goals will be to confirm the predictions derived from the general theory and the physical model of bacilliform capsids. To this end, the case of alfalfa mosaic virus (AMV) is particularly interesting, because independent experiments indicate that this plant virus might adopt an elongated structure with 3-fold axial symmetry. This is one of the new architectures predicted in our study, and has never been reconstructed for any virus. Curiously, all bacteriophage prolates characterized up to now adopt a 5-fold structure instead. Thus, we plan to compare the mechanical and assembly properties of both type of architectures to elucidate the potential physical advantages in relation also to the infection strategy of each virus.

A natural extension of the geometrical and physical model is the study of capsids with conical shape. These shells are similar to the elongated particles, but in this morphology each cap has a different number of disclinations (pentamers). Interestingly, one of the most studied viruses today, HIV, adopts precisely a conical nucleocapsid, and the rationalization of its structural principles is still a focus of intense research [5–7]. Therefore, this line of study is particularly appealing, and could shed some light on understanding this popular virus.

A priority in our future research is to progressively develop more realistic physical models for the intermolecular interactions in the viral capsids. The first step will be to introduce an explicit term in the capsomer-capsomer potential that will take into account the preferential angle of interaction. In this way, it will be possible to remove the spherical and spherocylindrical templates that we used to explore the equilibrium structures, the mechanical properties, and the assembly of viral capsids. This will allow us to refine different aspects studied here, and also extend the research to new scenarios.

For instance, we could explore the dynamic process involved in the buckling transition, related to maturation, and also generate intermediate shapes, instead of only comparing the final states for *perfect* spherical and icosahedral morphologies. In addition, this could naturally confirm if faceted $P = 3$ structures adopt

a more dodecahedral shape rather than icosahedral, as some viral structures also suggest [8].

The extended version of the capsomer-capsomer interaction model will permit us to deform the viral particles in multiple ways, leading to a better understanding of the mechanical properties of capsids. For example, we could mimic using simulations the standard atomic force microscopy (AFM) setup to perform local deformations, and compare the results more directly with the experiments. From this comparison, reasonable values for intermolecular interactions could be estimated, giving a more accurate insight than the classic elastic analysis commonly used today [9]. In addition, these studies could systematically establish the conditions where continuum elasticity theory is not a good approach, like in the built-in stress scenario observed for bacteriophage $\phi 29$.

Another immediate application of the extended intermolecular interaction will be the simulation of unconstrained capsid assembly. In this way, we could study in detail the kinetics of viral assembly, and determine the degree of accuracy of quantitative CNT predictions.

We are also interested in analyzing in more detail using the CNT model two particular set of experiments. First, we want to study the hysteresis experiments of hepatitis B virus [10]. Thus, applying the ideas derived in the CNT model we expect to extract valuable thermodynamic information, and estimate the conditions where assembly and disassembly are expected, respectively. Regarding the kinetics of capsid production, it will be worth to study, for instance, the light scattering experiments on the assembly of papilloma virus [11], and compare it with the rate of formation derived in the CNT context. Analyzing experiments at different concentrations of subunits might allow us to precisely characterize its kinetics, and make predictions. Remarkably, the ideal scenario would be to study both type of experiments for the same virus, although in the literature these data is very scarce.

In addition, the CNT could be extended to describe more complex assembly scenarios. One example is the formation of capsids in the presence of scaffolding proteins. In this case the new ingredients will be, essentially, the interaction between coat and scaffolding proteins, their stoichiometry in the capsid, and the concentration of scaffold proteins, which could be easily introduced in the CNT framework. Another scenario of special interest corresponds to the coassembly of coat proteins and the genetic material. Nevertheless, in this case there is not a unique assembly mechanism, which varies depending on the length of the genome and the capsid-genome interactions [12,13]. In addition, the configurations of the genetic material are more challenging to characterize energetically, so this will be

a more complicated case of study.

In a more general perspective, a long term goal is to connect the interactions of the physical phenomenological model with experimental accessible parameters. For instance, it has been shown that hydrophobic and electrostatic forces dominate the assembly of viruses, and can be modeled by effective potentials [14]. The rationalization of those interactions can be applied to analyze the assembly of specific viruses at different pH or salt concentrations [15]. Therefore, expressing the capsomer-capsomer model in terms of these realistic forces will allow us to simulate the assembly or mechanical properties of viruses in consonance with the actual physicochemical conditions. In this way, we could predict, for example, the assembly phase diagram (pH/ionic-strength) of specific viruses. Furthermore, the combination of *in vitro* experiments, these type of simulations, and the CNT theory might build a solid bridge to understand *in vivo* scenarios of virus assembly, which are particularly difficult to monitor and interpret today.

Another interesting line of investigation concerns the study of the reasons why viral coat proteins shapes can be grouped in few structural lineages [16]. Are the shape and interactions of these structures optimal designs for the assembly or the stability of viral capsids? To answer this question, we will adapt the capsomer-capsomer physical models to the case of model proteins, exploring different coarse-grained shapes for the subunits, and simulating the stability and mechanical properties of icosahedral capsids made of these coat proteins. Later on, we could study possible advantages regarding the assembly kinetics of capsids.

Finally, the principles governing the structure and assembly of viral shells are quite universal, so they should also be present in the formation and stability of other kinds of biological structures like gas vesicles [17], vault shells [18], photosynthetic vesicles [19], or ferritin [20]. More importantly, the rationalization of these principles could be the first step in the design of artificial materials in soft matter. Remarkably, novel structures mainly based on hydrophobic and electrostatic interactions have already been designed, mimicking the assembly of viruses [21, 22].

In summary, despite the significant achievements in virology during the last century, and the emerging interest in the interdisciplinary field of physical virology, several fundamental questions regarding the properties of viruses remain open. The understanding of the underlying physical mechanisms involved in the life cycle of viruses might help to control their impact in our society, and also to take advantage of their astonishing properties to develop fascinating nanotechnological applications.

References

[1] D. L. D. Caspar and A. Klug, "Physical principles in the construction of regular viruses," in *Cold Spring Harbor Symp. Quant. Biol.*, vol. 27, pp. 1–24, Cold Spring Harbor Laboratory Press, 1962.

[2] M. F. Moody, "The shape of the T-even bacteriophage head," *Virology*, vol. 26, pp. 567–76, 1965.

[3] M. F. Moody, "Geometry of phage head construction," *J. Mol. Biol.*, vol. 293, pp. 401–33, 1999.

[4] R. Zandi, P. van der Schoot, D. Reguera, W. Kegel, and H. Reiss, "Classical nucleation theory of virus capsids," *Biophys. J.*, vol. 90, pp. 1939–1948, 2006.

[5] J. A. G. Briggs, J. D. Riches, B. Glass, V. Bartonova, G. Zanetti, and H.-G. Kräusslich, "Structure and assembly of immature HIV," *Proc. Natl. Acad. Sci. USA.*, vol. 106, pp. 11090–5, 2009.

[6] A. Levandovsky and R. Zandi, "Nonequilibirum assembly, retroviruses, and conical structures," *Phys. Rev. Lett.*, vol. 102, pp. 13–16, 2009.

[7] O. Pornillos, B. K. Ganser-Pornillos, and M. Yeager, "Atomic-level modelling of the HIV capsid," *Nature*, vol. 469, pp. 424–427, 2011.

[8] X. Yan, N. H. Olson, J. L. van Etten, M. Bergoin, M. G. Rossmann, and T. S. Baker, "Structure and assembly of large lipid-containing dsDNA viruses," *Nat. Struct. Biol.*, vol. 7, pp. 101–103, 2000.

[9] W. H. Roos, R. Bruinsma, and G. J. L. Wuite, "Physical virology," *Nature Phys.*, vol. 6, pp. 733–743, 2010.

[10] S. Singh and A. Zlotnick, "Observed hysteresis of virus capsid disassembly is implicit in kinetic models of assembly," *J. Biol. Chem.*, vol. 278, pp. 18249 –18255, 2003.

[11] G. L. Casini, D. Graham, D. Heine, R. L. Garcea, and D. T. Wu, "In vitro papillomavirus capsid assembly analyzed by light scattering," *Virology*, vol. 325, pp. 320–7, 2004.

[12] O. M. Elrad and M. F. Hagan, "Encapsulation of a polymer by an icosahedral virus," *Phys. Biol.*, vol. 7, p. 045003, 2010.

[13] J. Rudnick and R. Bruinsma, "Icosahedral packing of RNA viral genomes," *Phys. Rev. Lett.*, vol. 94, pp. 1–4, 2005.

[14] J. N. Israelachvili, *Intermolecular and surface forces*. Academic Press, 2010.

[15] W. K. Kegel and P. van der Schoot, "Competing hydrophobic and screened-coulomb interactions in hepatitis B virus capsid assembly," *Biophys. J.*, vol. 86, pp. 3905–13, 2004.

[16] D. H. Bamford, J. M. Grimes, and D. I. Stuart, "What does structure tell us about virus evolution?," *Curr. Opin. Struc. Biol.*, vol. 15, pp. 655–63, 2005.

[17] A. E. Walsby, "Gas vesicles," *Microbiol. Rev.*, vol. 58, pp. 94–144, 1994.

[18] J. Querol-Audi, A. Casanas, I. Uson, D. Luque, J. R. Caston, I. Fita, and N. Verdaguer, "The mechanism of vault opening from the high resolution structure of the N-terminal repeats of MVP," *EMBO J.*, vol. 28, pp. 3450–3457, 2009.

[19] M. Sxener, J. Struumpfer, J. A. Timney, A. Freiberg, C. N. Hunter, and K. Schulten, "Photosynthetic vesicle architecture and constraints on efficient energy harvesting," *Biophys. J.*, vol. 99, pp. 67–75, 2010.

[20] G. C. Ford, P. M. Harrison, D. W. Rice, J. M. a. Smith, a. Treffry, J. L. White, and J. Yariv, "Ferritin: Design and formation of an iron-storage molecule," *Philos. Trans. R. Soc. London. [Biol.]*, vol. 304, pp. 551–565, 1984.

[21] A. Verhoeff, M. Kistler, A. Bhatt, J. Pigga, J. Groenewold, M. Klokkenburg, S. Veen, S. Roy, T. Liu, and W. Kegel, "Charge regulation as a stabilization mechanism for shell-like assemblies of polyoxometalates," *Phys. Rev. Lett.*, vol. 99, p. 066104, 2007.

[22] J. Zhang, D. Li, G. Liu, K. J. Glover, and T. Liu, "Lag periods during the self-assembly of {Mo(72)Fe(30)} macroions: connection to the virus capsid formation process," *J. Am. Chem. Soc.*, vol. 131, pp. 15152–9, 2009.

List of publications

- *Optimal architectures of elongated viruses*
  A. Luque, R. Zandi, and D. Reguera
  Proc. Nat. Acad. Sci. USA 107, 5323–5328 (2010).

- *The structure of elongated viral capsids*
  A. Luque and D. Reguera
  Biophys. J. 98, 2993–3003 (2010).

- *Built-in mechanical stress in viral shells*
  C. Carrasco, A. Luque, M. Hernando-Pérez, R. Miranda, J. L. Carrascosa, P. A. Serena, M. de Ridder, A. Raman, J. Gḿez-Herrero, I. A. T. Schaap, D. Reguera, and P. J. de Pablo
  Biophys. J. 100, 1100–1108 (2011).

- *Relevance of capsid structure in the buckling and maturation of spherical viruses*
  M. Aznar, A. Luque, and D. Reguera
  submitted to Phys. Biol.

- *The assembly of spherical shells: Line energy, implosion, and closure catastrophe*
  A. Luque, D. Reguera, A. Morozov, J. Rudnick, and R. Bruinsma
  preprint to be submitted to Phys. Rev. Lett.

- *Entropic particle splitter*
  D. Reguera, A. Luque, P.S. Burada, G. Schmid, J.M. Rubí, and P. Hänggi
  preprint to be submitted to Phys. Rev. Lett.