Facultad de Ciencias
Universidad de La Laguna

FINAL DEGREE PROJECT

Physical models of properties and structure of viral capsids

Author:
José Javier HERNÁNDEZ HERNÁNDEZ

Supervisor:
Dr. José María GÓMEZ LLORENTE

Department of Applied Physics

September 2020

Contents

Abstract

Los virus son los sistemas biólogicos más simples de la naturaleza, y por ello fueron los primeros en ser tratados matemáticamente. Es fundamental obtener la mayor cantidad de información posible a través de todas las ramas de la ciencia para obtener una imagen completa de sus características, dadas las propiedades emergentes del conocimiento. Es por ello que las propiedades físicas son tan importantes como las biológicas o las químicas. En este trabajo se introducen los principales modelos físicos (de autoensamblaje, cinética, elasticidad, etc.), con especial énfasis en las cápsides icosaédricas debido a sus propiedades de simetría. A continuación, se desarrolla la base de un modelo coarse-grained de 60 unidades asimétricas que junto a las propiedades del grupo de simetría del icosaedro nos permite calcular el número de modos normal de un virus icosaédrico sin hacer cálculos explícitos. También se obtiene información cualitativa del comportamiento de estos modos. Estos resultados son después comparados con cálculos reales de los modos normales del virus del Zika llevados a cabo por un grupo surcoreano [1] con buenos resultados.

Viruses are the simplest biological systems in nature, and because of that they were the first to be treated mathematically. It is fundamental to obtain as much information about them through all branches of science as possible to be able to get a full picture of their characteristics, due to the emergent properties of knowledge. Therefore, their physical properties are as important as their biological or chemical ones. We introduce some of the main physical models (self-assembly, kinetics, elasticity, etc.), with special emphasis on icosahedral capsids because of their symmetry properties. We then develop the basis of a 60 asymmetric units coarse-grained model that in conjunction with the symmetries of the icosahedral point group, allow us to calculate the number of normal modes of an icosahedral virus without making explicit calculations. We also gain some qualitative information about the behaviour of the normal modes. These results are then compared with the actual calculations of the normal modes of the Zika virus made by a South Corean reasearch group [1], with good agreement.

1 Introduction

En esta sección se da una pequeña introducción a la biología de los virus, con especial énfasis en su ciclo vital y en cómo se clasifican de acuerdo a la forma de su cápside. También se motiva el por qué de la investigación en ciencia básica de estos sistemas biológicos.

Viruses are very fascinating biological objects. They were discovered in 1899 by Martinus Beijenrinck and have been studied ever since because of their importance and impact to mankind. As the primary source of many diseases, they are able to cause grave alterations to our social structures, either by directly infecting people or through damaging our crops and livestock. Hence, it is fundamental to understand them as thoroughly as possible in order to protect our society. All these years of research have granted us many technological advances, ranging from the well-known vaccines, to the latest treatments such as virotherapy. This technique uses genetically modified viruses to infect and destroy cancer cells but not healthy cells [2].

Schematically, they are comprised of: 1) the genetic material, usually RNA (although it can be DNA), 2) a protein coat, called the capsid, which acts as a container for the genetic material, and sometimes 3) a lipid bilayer that covers the capsid (envelope). Since they are constituted by only 2 or 3 parts, they may seem simple, but actually a full description of their replication cycle and all their characteristics can be quite complex. Also, they do not fulfil some of the basic requirements to label them as living beings, such as lacking a cell structure, not having a metabolism or needing a living cell to make more copies of themselves. Thus, they are usually described as dwelling on the edge of life and death.

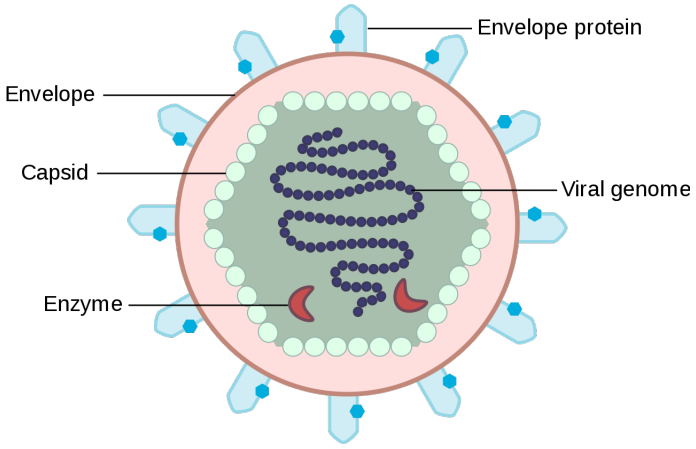

Figure 1: Schematic representation of the parts of a virus

These small structures can replicate themselves by using the metabolism and the molecular machinery of a host cell. Basically, they use the diverse structures

of their capsid or lipid bilayer as a key to enter. Once inside, they release their genetic material and use the molecular machinery to read through it, make copies of the genetic material, and construct the different proteins encoded in it. Then, in some viruses, the proteins are able to self-assemble into the capsid, around the genetic material, forming the complete virus. Some other viruses, usually with more complex forms, assemble a molecular motor that introduces the genetic material into the empty capsid. Once the host cell is full of viruses, they escape by rupturing the cell membrane, ready to infect new cells, repeating the cycle. Is in this release process that some viruses acquire the lipid bilayer from the very membrane of the host cell.

One way to classify them is through the shape of their capsid. According to this classification, there are 3 main groups: icosahedral, helical or irregular. Most of the known species of viruses fall along the first two categories, such as the human rhinovirus (icosahedral) or the tobacco mosaic virus (helical). Irregular viruses are usually more complex and interesting and can vary wildly in structure, as for example the bacteriophage T4 or the variola virus.

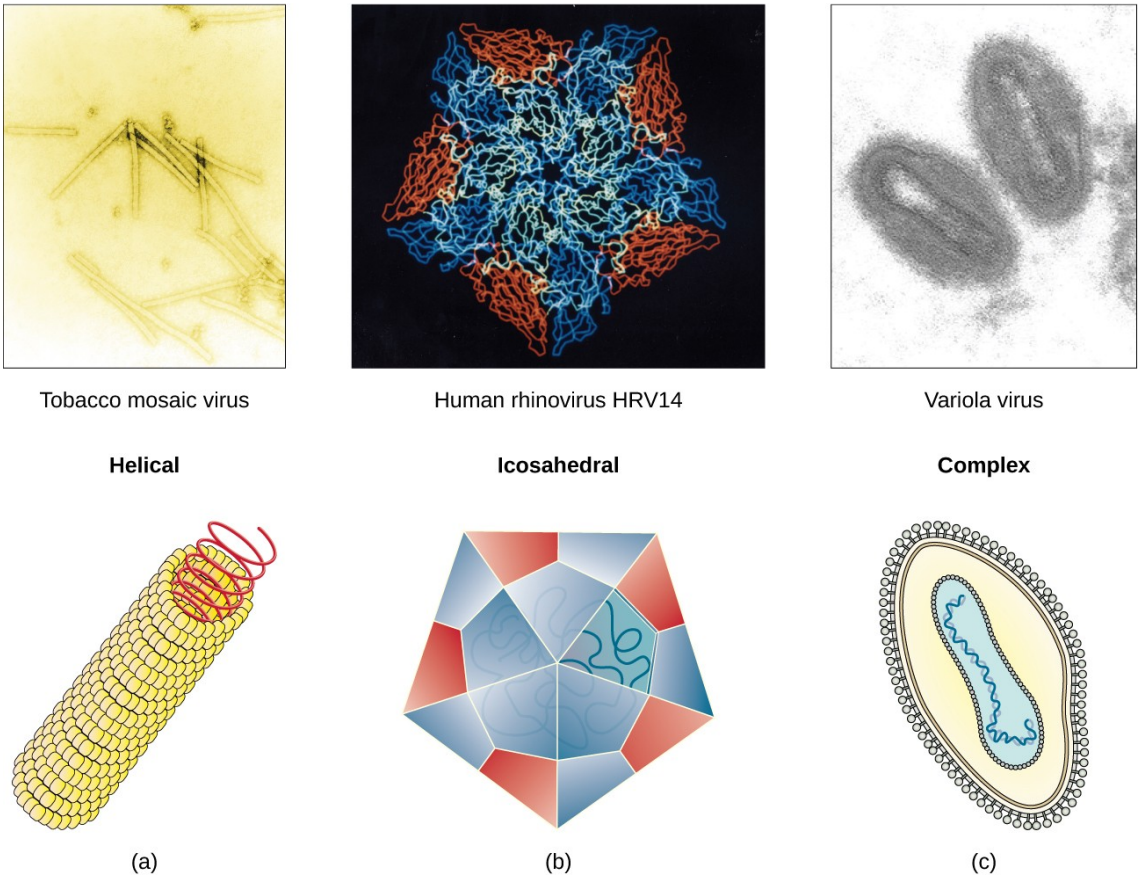

Figure 2: Examples of helical, icosahedral and complex viruses [3]

2 Structure of icosahedral capsids

En esta sección se describen las propiedades geométricas y de simetría que poseen las cápsides virales icosaédricas. En el ámbito geométrico, se explica cómo el modelo de Kaspar y Klug es capaz de generar una cápside partiendo de una red hexagonal bidimensional. En concreto, se demuestra que solo ciertas cápsides con distancias concretas entre sus vértices son posibles. En el ámbito de la símetria se describe cómo partiendo de una unidad asimétrica (como por ejemplo una proteína) y tras aplicar sobre ella los elementos del grupo puntual del icosaedro se obtiene la cápside completa.

In this section we will focus on the structure of icosahedral viruses, since they are simple enough to be described mathematically due to the symmetry properties of the icosahedron. The fact that so many viruses have this shape can be argued with "economy" arguments. The size of the capsid puts a constraint in how much genetic material it can contain. Thus, it is seen in nature that most viruses only have the genetic information to build one or a few proteins. For this, the proteins should have some kind of asymmetry in order to be able to form the entire structure [4]. Since the number of ways to construct a solid with cubic point symmetry and with asymmetric units (aus)are limited, it is not a surprise the icosahedral symmetry appears in nature: the icosahedron is the platonic solid that maximises the volume (genetic material) per unit surface (number of proteins it is made of) [5].

2.1 Caspar and Klug model for icosahedral capsids

The Caspar and Klug (CK) model is a way to construct an icosahedral capsid from an hexagonal (or equivalently triangular) lattice [6]. Basically, we will construct a bigger structure, the capsomer, which can be of 2 types: hexamers and pentamers, which are comprised of 6 and 5, respectively, assymetric units. In the hexagonal lattice, we need to replace 12 of the hexagons (hexamers) with 12 pentagons (pentamers). Thus, we will form a closed 3-D surface from the 2-dimensional lattice. Due to the Euler theorem for polyhedra, there are a limited number of ways to construct a closed surface [7]. The way to choose this hexagons are given by the *triangulation numbers* T, which also serve as a classification tool in the CK model. This T numbers are actually the distance squared between adjacent vertices (the pentamers). They are determined by 2 numbers (h, k) which describe the 2 degrees of freedom of our hexagonal lattice.

In vector form, the vector that joins 2 of these pentamers is:

$$\vec{C_T} = h\vec{a_1} + k\vec{a_2}.$$

Due to the geometrical properties of the lattice ($\vec{a_1} \cdot \vec{a_2} = -1/2$), we arrive at the result that T is given by:

$$T = h^2 + hk + k^2.$$

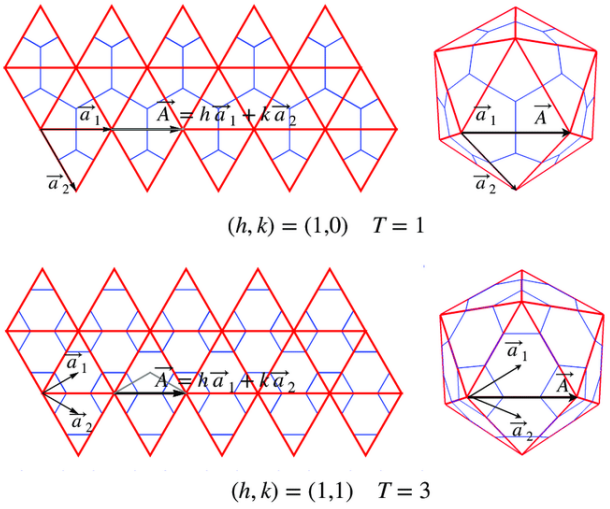

Figure 3: Visualisation of the mapping between the hexagonal lattice and the closed surface. [8]

Thus, it is clear that we have some 'magical' allowed numbers $T = 1, 3, 4, 7, 9...$ that constrain the number of capsomers we need to use to construct our capsid for a given T. Since every icosahedron has 20 faces (each of them will contain T elementary faces of 3 asymmetric units each) the total number of asymmetric units will be

$$N_{au} = 60T.$$

We can also calculate how many capsomers we will have for a given T number. There will always be $N_{pen} = 12$ pentamers, and a variable number of hexamers given by $N_{hex} = 10(T-1)$. Thus, the total number of capsomers will be [5]:

$$N_{cap} = 10T + 2.$$

Also, it is worth noting that T can be degenerate with respect to (h, k). This means that to describe a capsid unequivocally we need to use the (h, k) numbers, since for a given T we can arrange the same number of capsomers in different ways. When our capsid is different from the types $(0, k)$, $(h, 0)$ or (h, h) this degeneration gives rise to different optic isomers, either levorotary or dextrorotary.

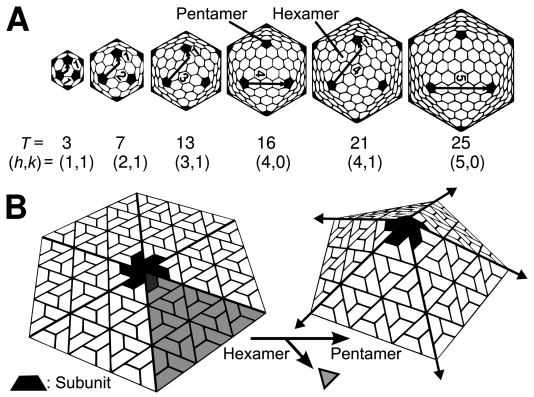

Figure 4: **A**: icosahedral capsids for different values of T. **B**: Representation of how to obtain a pentamer from a hexamer by removing an asymmetric unit. [9]

2.2 Symmetry properties of the icosahedron

A regular icosahedron is the platonic solid with more faces. It has 20 triangular faces, 30 edges and 12 vertices. It also has 3 different types of symmetry axis, related to these 3 elements. We have 2-fold axis in each of the edges, 3-fold axis in each of the faces, and 5-fold axis in each of the vertices. Thus, the icosahedral rotation group I, which contains all possible rotations that leave invariant the icosahedron, has order 60.

The 60 elements that form I can be grouped in conjugacy classes, which are disjoint subsets containing elements of the icosahedron group that are related by a conjugacy relation. What this means is that each of the elements in a certain class produce a *different* effect (since they are not the same element) but the effect is of the same *type*. The conjugacy classes (described by the effects of the elements they contain) are the following:

- the identity element (E).
- 12 rotations of 72^o, order 5 ($12C_5$).
- 12 rotations of 144^o, order 5 ($12C_5^2$).
- 20 rotations of 120^o, order 3 ($20C_3$).
- 15 rotations of 180^o, order 5 ($15C_2$).

These symmetry properties indicate why the number of asymmetric units of an icosahedral capsid derived in the previous subsection has to be a multiple of 60 ($60T$). If it was not a multiple of 60, the action of all the elements on a single asymmetric unit (or on T aus, when we have more than 60) would not form the capsid.

Thus, the icosahedron can be thought of as comprised of (at least) 60 asymmetric units. If we applied the 60 elements of the group to a single au, we would obtain the entire platonic solid. This asymmetric unit can be just a single protein, or a combination of them. Also, this means that each of the 20 faces of the icoshaedron contains 3 aus. The fact that the icosahedron is made of 60 asymmetric units does not mean that all viruses with icosahedral symmetry are made of 60 aus (or a multiple of 60): the icosahedral point group symmetry does not impose it. Several studies, such as the ones made by Watson and Crick [10] show that there are viruses with this symmetry that have a different number of subunits.

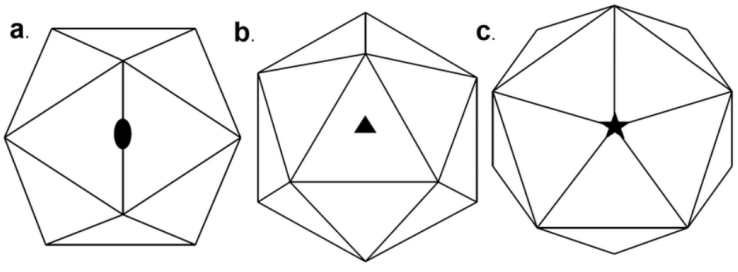

Figure 5: a) 2-fold b) 3-fold and c) 5-fold rotation axis of the icosahedron [11].

3 Formation of the capsid: self-assembly

En esta sección se demuestra como el autoensamblaje de una cápside viral vacía es posible y de hecho favorecido bajo determinadas circunstancias de equilibrio termodinámico. Se obtiene una ley de acción de masas que predice que cuando la concentración de unidades asimétricas supera un cierto valor crítico, el autoensamblaje es favorecido. Además, se describe un posible modelo de cinética del ensamblaje en el que las unidades asimétricas solo pueden unirse o separarse de los agregados intermedios de una en una. Esto produce un comportamiento sigmoideo en el tiempo de la concentración de cápsides completas, como se observa experimentalmente.

One of the most remarkable characteristics of the viral capsids is the fact that they are able to self-assemble. This means that the proteins that are encoded in the genetic material are able to join together spontaneously. In fact, this process can be carried out *in vitro*, as several experiments show [12][13]. This happens because within the physiochemical conditions of the host cell (or the test tube), this process is favoured thermodynamically.

We will study the thermodynamics and the kynetics of the assembly for an empty capsid, since for most viruses this is a usual first step in the building of the capsid. As we said in the introduction, some viruses construct the capsid around the cargo, whereas others use a molecular motor to introduce the cargo inside the capsid.

3.1 Thermodynamics of the self-assembly for an empty capsid

We will derive a typical law of mass action (LMA) for the concentrations of intermediate capsids, comprised of n subunits [14]. Our system is a collection of identical subunits that will assemble into a $T = 1$ capsid. We are going to assume that the conditions for thermodynamic equilibrium are possible within the physiological conditions of the host cell, as well as that there is only one possible structure -intermediate- that has n subunits (in order to make the computation tractable). Thus, a possible way of writing the free energy of a system of subunits, intermediates and capsids in solution is:

$$F = \sum_{n=1}^{N} \left(k_B T \rho_n [\ln(\rho_n v_0) - 1] + \rho_n G_n^{cap} \right),$$

where ρ_n is the density of intermediates with n subunits, v_0 a standard volume and G_n^{cap} is the energy of interaction for the respective n-intermediate. This interaction energy depends on the geometry of the particular intermediate, the number of ways the n-intermediate can be formed, as well as the temperature, the ionic strength and the pH. A suitable model for this energy can be found in [15].

In order to obtain the equilibrium concentration we need to minimise the free energy F with the constraint that the total concentration of subunits, ρ_T, remains

unchanged:

$$\sum_{n=1}^{N} n\rho_n = \rho_T.$$

If we minimise the free energy subject to the above-mentioned constraint we obtain the desired LMA:

$$\rho_n v_0 = e^{-\beta(G_n^{cap} - n\mu)}, \tag{1}$$

with

$$\mu = k_B T \ln(\rho_1 v_0), \quad \beta = \frac{1}{k_B T}.$$

To be able to solve equation (1) we need to use computational methods, due to the constraint over the total concentration ρ_T. But it is observed that, at equilibrium, the intermediate state concentrations are negligible. Thus, the LMA can be simplified by neglecting all concentrations but those of the free subunits and those of the full capsids:

$$\rho_T = \rho_1 + N\rho_N. \tag{2}$$

If we now call the fraction of subunits that form part of a capsid f_c, by arranging equations (1) and (2) we arrive at

$$\frac{f_c}{1 - f_c} = N(\rho_1 v_0)^{N-1} e^{-\beta G_N^{cap}}.$$

It is interesting to observe what happens in the limit $N \gg 1$. It can be shown that:

$$\frac{f_c^{1/N}}{1 - f_c} = \frac{\rho_T}{\rho^*},$$

with

$$\rho^* v_0 = \left(\frac{e^{\beta G_N^{cap}}}{N}\right)^{\frac{1}{N-1}} \approx e^{\beta \frac{G_N^{cap}}{N}},$$

where ρ^* is the critical concentration of free subunits. And if we observe the asymptotic behaviour for the ratio ρ_T/ρ^*, we obtain the following equations:

$$f_c = \left(\frac{\rho_T}{\rho^*}\right)^N \qquad \text{when } \frac{\rho_T}{\rho^*} \ll 1, \tag{3}$$

$$f_c = 1 - \frac{\rho^*}{\rho_T} \qquad \text{when } \frac{\rho_T}{\rho^*} \gg 1. \tag{4}$$

Thus, we have shown that under conditions of thermodynamical equilibrium, the spontaneous formation of capsids is possible and in fact favoured when the total concentration surpasses the critical one [14].

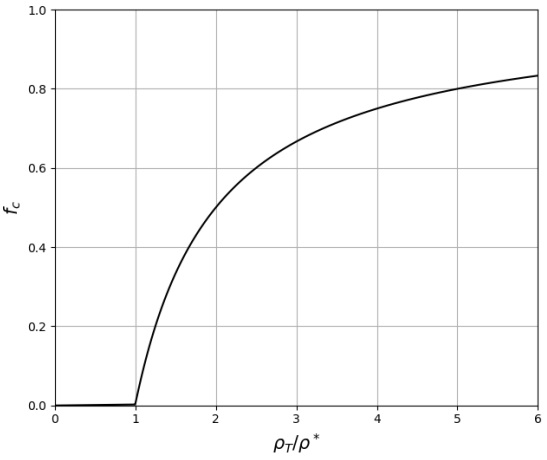

Figure 6: We can clearly see that when the total concentration is less than the critical concentration, no capsid formation occurs. But as soon as we surpass the critical concentration, the fraction of subunits that form part of a capsid quickly grows up to its maximum value $N\rho_N/\rho_T$. This change of behaviour is more drastic as the number of subunits that are needed to form the full capsid increases, i.e. for large T. [This particular plot was made for $N = 1000$.]

3.2 Kinetics of the self-assembly for an empty capsid

There have been several experiments as well as computational calculations that show that the kinetics of the self-assembly are sigmoidal with respect to time evolution [14]. This means that the self-assembly takes place very quickly after a certain period of time has elapsed, which depends on certain parameters such as the concentration of subunits, the strength of the inter-subunit interactions, among other physiochemical parameters [14].

In order to describe the assembly reactions that take place to form the entire capsid starting from single subunits, we need to specify the intermediate steps that conform it. As a first thought, one could think of the several paths this reactions might take, such as adding subunits one at a time or forming first certain pieces of the capsid that are later combined into the full structure. We know thanks to simulations that the first idea is the predominant one, which could be a consequence of the different probabilities all possible events have [5]. Here we will explain the model developed in [5] which only allows the subunits to bind (or unbind) with just a single subunit at every step of the reaction. These can be expressed symbolically as:

$$c_n(t) \underset{\alpha_{n+1}(t)}{\overset{\beta_n(t)}{\rightleftharpoons}} c_{n+1}(t),$$

where the $c_n(t)$ show the number of aggregates formed of n subunits, while the $\beta_n(t)$ and $\alpha_n(t)$ are binding and unbinding rates, respectively. The dependence of these two last sets of functions with n and time can be motivated by the results shown in the previous subsection.

Thus, in general the time evolution of $c_n(t)$ will be given by the following master

equation (the time dependence is assumed implicitly):

$$\frac{\partial c_n(t)}{\partial t} = J_{n-1} - J_n, \tag{5}$$

where

$$J_n = \beta_n c_n - \alpha_{n+1} c_{n+1} \tag{6}$$

is the current of aggregates growing from n to $n+1$.

In equation (5) we have all the possibilities in which the number of aggregates made of n subunits may change. Obviously, we have N equations, one for every possible aggregate (with the aggregate with N subunits being the full capsid). These means we have a system of coupled equations which can be solved if one knows the form of the binding and unbinding rates (see the Smoluchowski's theory of aggregation [16]). We also need to include the conditions $\beta_N = 0$ and $\alpha_1 = 0$, since a full capsid cannot grow any further, and a single subunit cannot lose another subunit. If we found a solution for this set of equations, we would obtain the time-dependant concentrations of all aggregates: free subunits, intermediates and full capsids. This can be a very difficult task, specially for large N.

By the use of the adiabatic approximation, the set of N equations can be solved in a more manageable way. In this approximation, it is assumed that the kinetic reaction is sufficiently slow as to be characterised by a quasi-steady current $J(c_1(t))$ (whose explicit form can be found in [5]). As explained in the previous subsection, the total concentration of subunits (which is fixed) can be approximated by

$$c \approx c_1 + N c_N.$$

If we assume that locally the reaction is in a steady-state for a given concentration $c_1(t)$, then the rate of formation of full capsids will depend on c_N only:

$$J(c_N(t)) = \frac{\partial c_N}{\partial t}. \tag{7}$$

And by integrating that implicit equation we would obtain the time evolution of the comple

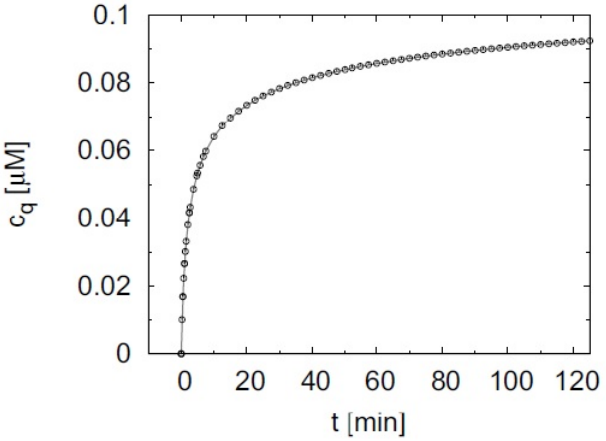

Figure 7: Plot obtained by [5] after solving the set of master equations with the help of the adiabatic approximation for the case $N = 72$. We can clearly see the sigmoidal behaviour explained at the beginning of the subsection. Also, it is worth noticing that it takes a considerable amount of time to approach the maximum concentration of complete capsids (≈ 1 hour).

4 Physical properties of the capsid

En esta sección se proponen sendos modelos que describan las propiedades mecánicas y electrostáticas de una cápside viral. En relación con las propiedades mecánicas, se utiliza la teoría de láminas (TST) para obtener de forma teórica los perfiles de indentación que pueden ser obtenidos mediante las técnicas de microscopía de fuerza atómica. Con respecto a las propiedades electrostáticas, se calcula la energía propia de atracción y se compara con los valores experimentales de la energía de repulsión, mostrando el delicado equilibrio electrostático en la cápside de estabilidad/inestabilidad.

Since the capsid is the main tool (usually the only one) that a virus uses to interact with their surroundings, it is natural to think that the physical properties of the capsid play a key role on its functionality, such as its stability against exterior forces. We need to take into account that viral capsids need to be stable enough to protect and carry the genetic material around but unstable enough to be able to release it at the appropriate time and place. Therefore it is critical to understand their physical properties to gain more insight into the aforementioned characteristics. In particular, we will describe some of the discoveries regarding the indentations produced by force probes on viral capsids as well as the electrostatic self-energy related to the self-assembly.

4.1 Mechanical properties of the capsid

A useful way of measuring the stability and resistance of a capsid against external forces is to simply "poke it". With this simple method of applying forces onto a capsid and measuring how much it deforms, if it recovers its original form or if it breaks, we can gain some understanding on the mechanical properties of these structures and how they resist external "threats". Numerous experiments have been conducted on this track, and specially significant are the ones performed with an atomic force microscope (AFM), whose techniques are summarised in [17].

We can relate the applied force with the corresponding change in the capsid, obtaining a so-called force-deformation curve (FDC). As said before, the different reactions of the structure after the force is removed reveal the regimes of reversible and irreversible deformation. This transition between reversible and irreversible deformation occurs when the force probe moves the viral structure away from the state of minimum energy (maximum stability). Therefore, if we modelled the deformation free energy of the capsid and knew the applied force, we could compute the deformation profile of the structure, which can be compared with experimental measures (FDC) [17].

In various fields of engineering it is of great importance to know the effects of external forces or pressures on thin-walled or shell-like materials, such as airplanes, gas tanks, etc. For that matter they use the "thin-shell theory" (TST), which can

be applied to viral capsids as well. We can model the capsid as a thin spherical shell of radius R and uniform thickness. Even though we have assumed in previous sections that the viruses are empty for several calculations, in this model we need to include the osmotic pressure exerted by the genetic material, since it can be as high as 50 atm [18]. Since we want to compare this with FDC, we will try to express the free energy of deformation in terms of the radial indentation profile $\zeta(\mathbf{r})$, which is the radial displacement of the sphere in terms of a 2D coordinate system on the surface of the sphere. When taking the limit for small $\zeta(\mathbf{r})$, the TST gives us the following expression for the free energy [17]:

$$\Delta F = \int dS \left[\frac{1}{2}\kappa(\Delta\zeta)^2 + \frac{1}{2}\tau(\nabla\zeta)^2 + \frac{1}{2}Y\left(\frac{2\zeta}{R}\right)^2 \right]. \tag{8}$$

The first term of this expression is the bending energy of the shell (notice how $\Delta\zeta$ represents the curvature and that κ has units of energy), the second term is the work done by the probe against the osmotic pressure (τ is a surface tension), and the third term measures the stretching of the thin layer (Y is the 2D Young modulus) [17]. By computing the functional derivative of the free energy with respect to the radial displacement and setting it equal to the probe force per unit area, we can obtain the indentation profile. In fact, for the ideal case of a punctual force, the equation

$$\frac{\delta\Delta F}{\delta\zeta(\mathbf{r})} = f\delta(\mathbf{r})$$

can be solved analytically. One obtains that the force creates a crater with a radius of order $\sqrt{Rl_b}$ (where $l_b = \sqrt{k/Y}$ is a characteristic length scale) and when the force applied is small, it has a linear response, which means it acts like a spring [17].

For larger forces, equation (8) cannot be used, Instead, one has to solve a group of daunting equations known as the Föppl von Kármán equations. Even computationally they are rather difficult to solve, so its better to minimise the elastic energy by methods such as the finite-element modelling (FME) [17].

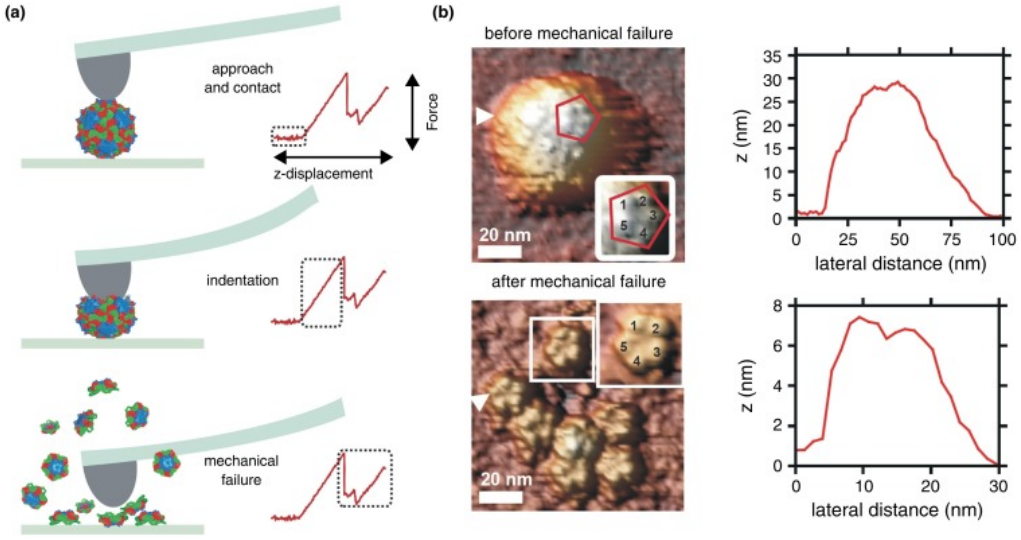

Figure 8: a) scheme of how the AFM creates the indentation (or breaks) the capsid. Notice the abrupt change in response after the linear regime is no longer valid. b) On the left, Triatoma virus before and after irreversible deformation. Even individual pentamers can be seen after the mechanical failure; on the right, indentation profiles for the top of the capsid (top) and the pentamer (bottom) [19].

4.2 Electrostatic properties of the capsid

We know that capsids interact electrostatically, since that is the way some viruses attach to different surfaces, such as enteric viruses on lettuce [20]. Viral capsids are formed by proteins, which through their charge distributions acquire the secondary (and upper) structures. In fact, the interactions between the charges in the proteins that conform the capsid and the charges in the genetic material (which are arranged in a complementary fashion) increase the electrostatic interaction, decreasing the energy and facilitating the assembly [21]. This creates, in general, a layered distribution of charge.

The easiest way to model the interaction self-energy of the capsid is to calculate that of a uniformly charged, permeable and infinitely thin spherical shell of radius R (surface charge density σ). One approach to computing this is by treating the ions as an ideal gas, which adjust to the external potential and contribute to it via their charge density. This approach is called the Poisson-Boltzmann (PB) approach, and gives us a non-linear differential equation for the potential ϕ. We can obtain the PB equation by minimising the appropriate free energy, which is [22]:

$$F_{PB}[\phi(r), \nabla\phi(r), c^i(r)] = \int f_{PB}(\phi(r), \nabla\phi(r), c^i(r)) d^3r,$$

with

$$f_{PB}(\phi(r), \nabla\phi(r), c^i(r)) = -\frac{1}{2}\epsilon\epsilon_0\nabla\phi(r)^2 + \sum_{i=\pm} e_i c^i(r)\phi(r) + e_0\rho_p(r)\phi(r) + \\ + \sum_{i=\pm} k_B T[c^i(r) ln c^i(r) - c^i(r) - (c_0^i ln c_0^i - c_0^i)].$$

For the above equation, $e_0$is the charge of the electron, $e_0\rho_p(r)$ represents the charge density of the capsid, c^i are the concentrations of the salt ions (with c_0^i being their bulk concentrations) and $\epsilon\epsilon_0$ is the permittivity of water. And after a minimisation with respect to the three functions of the F_{PB} functional, we arrive at the PB equation [21]:

$$-\epsilon\epsilon_0\nabla^2\phi(r) = e_0\rho_p(r) + \sum_{i=\pm} e_i c_0^i e^{-\beta e_i\phi(r)},$$

where $\beta^{-1} = k_B T$.

If the expression on the exponential is sufficiently small (small potentials in the solution) and we have a symmetric system ($c_0^+ = c_0^-$), the PB equation can be linearised, which gives the Debye-Hückel equation for the potential [21]:

$$-\nabla^2\phi(r) = \kappa^2\phi(r) + \frac{e_0\rho_p(r)}{\epsilon\epsilon_0} + ..., \tag{9}$$

where $\kappa^2 = \beta\left(\sum_{i=\pm} e_i^2 c_0^i\right)/\epsilon\epsilon_0$ its the square of the screening length. This equation can be solved with several techniques, such as the Green's method. Basically we obtain a Coulomb-screened potential, which we can integrate over the two surfaces of the thin shell to obtain the self-energy of the capsid [21]:

$$F_{DH} = \frac{1}{2}\frac{\sigma^2}{4\pi\epsilon\epsilon_0}\int dS_1 \int dS_2 \frac{e^{-\kappa|\mathbf{r}_1-\mathbf{r}_2|}}{|\mathbf{r}_1-\mathbf{r}_2|},$$

where the factor of 1/2 eliminates double counting of pair interactions. Given the conditions of the physiological solution where capsids lie inside living beings and the typical radius for a virus, we see that $\kappa R >> 1$, which helps us to simplify the integral. This condition implies that the range of integration is effectively cut on the scale of κ^{-1}, which separates the two integrals and gives the following result for the self-energy [21]

$$F_{DH} = \frac{\pi\sigma^2 R^2}{\epsilon\epsilon_0\kappa}. \tag{10}$$

This energy is the energy required to bring infinitesimal charges from the infinite to the solution to form the capsid [21]. We can estimate this value given the typical size of a virus, the usual physiological conditions and the the aminoacid content, which gives us the charge density [23], and we get a value of $F_{DH} \approx 10^4 k_B T$. Now we can compare this attractive energy of interaction with the corresponding repulsion energy, which has been measured for the hepatitis B virus with a value of $\approx 1.5 \cdot 10^3 k_B T$ [24]. Thus, we see the delicate equilibrium between stability and instability, which is necessary for the capsid to release the genetic material at the appropriate place.

5 The computational approach: 2 different models of interaction

En esta sección se describen las dos técnicas principales usadas en simulaciones moleculares por ordenador, los modelos all-atom y los modelos coarse-grained. Dentro de cada modelo se contraponen sus puntos fuertes a sus puntos débiles y se pone en valor la utilidad de cada uno. Además, se comenta la posibilidad de poder combinar ambos métodos en uno solo para intentar compensar mutuamente sus debilidades.

Up to this point we have used continuous models to describe some of the physical properties of the capsid. But since matter is discrete, to model reality in a continuous manner is actually an approximation. Therefore, if we want to increase the accuracy of the description of our system we inevitably need to take the discrete nature of reality we into account.

Thus, we need to introduce how the different parts (atoms, molecules, or other subunits) interact between them, and the way to do that is through the use of intermolecular potentials. These are very varied and they choice of using ones instead of others depends on many variables such as the scale of the interactions, their strength, etc. For example, we could model the intermolecular interactions in a capsid with the famous Lennard-Jones potential, or with a simple spring like interaction.

In theory, if we wanted to study the interactions between a finite number of subuits we would construct the corresponding set of equations taking into account the relevant interactions for our model and solve it, either with pen and paper or, for a large number of atoms, with the help of a computer. Here, we will describe two different approaches to tackle this mathematical problem computationally: all-atom and coarse-grained simulations.

5.1 All-atom models

The approach all-atom models take is to attempt to simulate all the atoms (sometimes reducing the radicals of a molecule) that conform the system and all their mutual interactions. Therefore, the results obtained through this method are in principle exact up to the approximations one may introduce in the interaction potentials of the system. The problem we encounter is that for a sufficiently large number of atoms, we can run into problems even with the help of a computer, since it can need an incredible amount of time to perform the calculation.

Let's check the magnitude of the problem if we tried to model an empty capsid. Proteins are constituted, roughly, by 6000 atoms [25]. This means that a virus is constituted by approximately $60 \cdot 6000 = 3.6 \cdot 10^5$ atoms (T1 virus with 1 protein per asymmetric unit). And this rapidly increases the amount of time the computer takes to compute every step of the simulation. To put into perspective the scales

of computing time we are dealing with, the current record for the longest all-atom simulation of an empty capsid (an HIV-1 capsid) simulated 64,423,983 atoms for only (or as long as) 1.2 μs [26]. Even with a supercomputer, it could only calculate 8 ns of the simulation per day, which means it took about 150 days of computations to obtain the entire sequence [26]. Also, it requires huge amounts of memory power: every 8 ns weighed 315 gigabytes! [26].

Thus, it seems like all-atom models can only be applied making a balance between the size of the system and the computational resources one has access to. But this could change soon once quantum computers increase in computational power and efficiency, making all-atom models more broadly usable.

5.2 Coarse-grained models

Coarse-grained (CG) models attempt to simplify the all-atom approaches in order to make them more manageable computationally. Instead of taking into account all possible interactions between all the atoms that conform the system, coarse-grained models split the system into groups of atoms or molecules and treats each of them as single particles, thus reducing the complexity of the problem. Instead of particles one could also use rigid solids. This gives us freedom of choice to choose the scale of the simulations, as well as how many degrees of freedom we want to take into account. With a less complex problem, we can achieve longer simulations with shorter times of calculation. Usually, the model is considered coarse-grained when one considers as subunits of the system at least the individual proteins.

Obviously, how we group the atoms or the molecules can change drastically the behaviour of the system in comparison to the all-atom prediction. This can happen either by making a poor choice of the elements that conform each group, or by putting too many atoms (or molecules) in each group. A poor choice of any (or both) could lead to unrealistic movements in the system and/or to eliminate the finer details of the simulation. Therefore, one has to make a balance between the computational power available and the details and realism one wants to obtain.

This model is broadly used across all kinds of molecular dynamics simulations: from single proteins to viral capsids as well as all types of biomolecules [27]. It is a consequence of its scalability and lesser requirements of computational power, which compensate the lack of precision when compared with all-atom simulations. It is also very interesting to compare how both techniques can be combined, using the all-atom model for certain parts of a molecule and the coarse-grained for others, increasing the details of the simulation without a lot of computational cost [27].

6 The 60 asymmetric units model (AUM-60)

En esta sección se construye un pequeño modelo coarse-grained de una cápside viral formada por 60 unidades asimétricas, las cuales son modeladas como sólidos rígidos. Por un lado, se construye una matriz de contactos que servirá para calcular la matriz del potencial necesaria para calcular ciertos parámetros físicos como las frecuencias de oscilación. Por otro, se utilizan las técnicas de teoría de grupos para calcular cuantos modos normales posee la cápside sin calcularlos explícitamente y se describen de forma cualitativa sus posibles simetrías.

After developing a model of interaction, we basically have solved all of our problems, since from that we can calculate almost all of the physical properties of interest, such as the equilibrium configurations [28], the kinetics [29]... and in particular the normal modes and its associated frequencies. Thus, it is of great importance to introduce as much detail as possible into the model while keeping it simple enough so it remains tractable.

We have seen that even though the coarse-grained model is not the most precise method of simulating a molecule computationally, it is the best when we take into account the computational cost versus the detail we obtain. And we can enrich and simplify our model further with the use of symmetry arguments, which will allow us to obtain information about the normal modes without making explicit calculations. Because of their usefulness, CG models that rely on symmetry properties have already been developed, using as a basic unit trimers and pentamers [30].

In this chapter we will take a step further from those models and lay the basis to simulate an empty viral T1 capsid comprised of 60 asymmetric units using a coarse-grained with symmetry arguments model. Instead of considering each au as point particles as it is usually done, we will model them as rigid bodies, thus increasing the number of degrees of freedom of the system and the fineness of the model (3 rotations and 3 translations).

6.1 The construction of the contact matrix

To be able to simulate any molecule we need to know how the different atoms (or the asymmetric units, in our capsid) interact. We will make the approximation that each asymmetric unit only interacts with their near-neighbours, which leaves us with 5 interactions per au. Moreover, the fact that each au is, by nature, asymmetric, means that these 5 interactions will all be different, but a certain level of symmetry remains thanks to the icosahedral shape. If we use as a reference an asymmetric unit with its short base up, interactions that involve right-top sides have been labelled as “Type 1”, interactions left-bottom as “Type 2” and interactions bottom-bottom as “Type 3”. Types 1 and 2 are asymmetric, since the aus involved have different effective shapes.

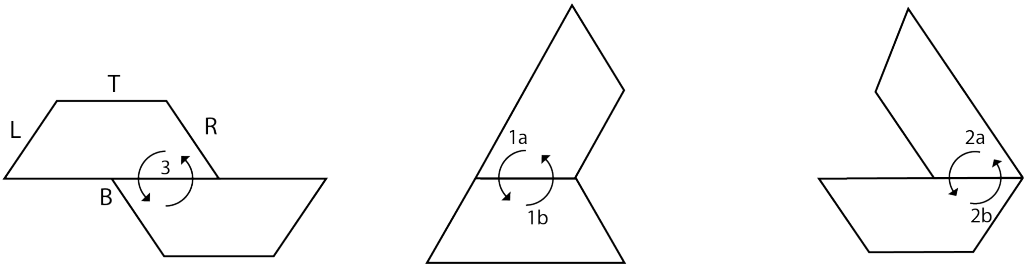

Figure 9: The three different types of interaction in our model (types 3, 1 and 2 from left to right). Notice the asymmetry of types 1 and 2.

In order to construct a matrix that contains all the information regarding which units interacts with which, we need to number each asymmetric unit in the capsid. This is in principle a *free-to-choose* task, since there is no immediate way to do it (this is worth investigating further!). I chose the numeration provided by my tutor to be able to compare results easily. Not having an ubiquitous way to number the capsid means the ways of coding the contacts between aus are limited. I found that using the property that the contact matrix is symmetric and that the numbers of the aus follow certain simple recurring relations for the different interactions works best for simplifying the code. Of course, the recurrence relations are dependent on the choice of the numeration. The resulting matrix as well as the numeration used are available in the appendix.

The contact matrix has a size of 60×60. But this is not the size of the matrix that contains the interactions. The matrix we developed only connects the appropriate asymmetric units. Since we are modelling the aus as rigid bodies, in the actual potential matrix we need to include the 6 coordinates of each au. Thus, every element of the contact matrix is in fact a 6×6 matrix, making the real size of the matrix that contains the interactions 360×360. Once we describe the potential, we can calculate physical properties of the capsid such as its normal modes and their associated frequencies. There is, in principle, a maximum number of normal modes given by the amount of degrees of freedom of the system. Dealing with 60 aus, we have $6(N_{au} - 1) = 354$ possible normal modes (eliminating translations and rotations of the capsid as a whole). But the fact that the system is invariant under the action of the elements of I means the actual number of normal modes is less (and even more the number of different frequencies!).

6.2 Group theory: reducing the complexity

The problem of calculating the normal modes (in the linear regime) of a physical system is often reduced to finding the eigenvalues (the frequencies) and the eigenvectors (normal modes) of a certain matrix H that depends on the kinetic and potential energy of the system. Basically, we construct the Hamiltonian of the system and then Taylor-expand the potential around the equilibrium configuration so that the corresponding equations of motion have harmonic-type solutions. This gives us the small oscillations of the system around such equilibrium configuration. This seemingly simple problem can be easily solved with pen and paper for systems

with a small number of elements, or with computational methods when this number increases. But if the system is too big, this last method start to cost a lot of computational time. Therefore, one is compelled to look for other restrictions that can be imposed on the system, trying to reduce the complexity of the matrix.

The methods developed in group theory can help us decrease the complexity of the problem by transforming the matrix H into another that is block-diagonal, where one can obtain the eigenvalues and eigenvectors in a much simple manner. One only has to perform a similarity transformation upon H with a certain matrix that is determined by the symmetry of the system and the choice of coordinates for the normal modes [31]. This is called to *reduce the representation*. Basically, we express our matrix as the direct sum of a certain combination of the irreducible representations (irrep) of the symmetry group. And even without making explicit calculations, we can still gain some insight into the behaviour of the system.

As proof of this, we will show how many normal modes there are, if their associated frequencies are degenerate or not, and gain certain insight on how the capsid vibrates. Group theory tells us that a certain irreducible representation appears in our reducible representation a specific number of times (we ignore the complex conjugation since the icosahedron point group is real) [32]:

$$n^{(\alpha)} = \frac{1}{g} \sum_{G} \chi^{red}(G)\chi^{(\alpha)}(G). \tag{11}$$

Here α denotes the irreducible representation, g is the order (number of elements) of the symmetry group, and $\chi(G)$ represents the character of correspondent representation (the sum is carried out over all the elements of the group). Thanks to a character table, readily available in the literature, we can obtain easily the values of the characters. We obtain that our reducible representation can be expressed as

$$S = 6A \oplus 16T_1 \oplus 18T_2 \oplus 24G \oplus 30H, \tag{12}$$

where the irreducible representations associated with translations and rotations of the capsid as a whole have already been removed.

Since the matrix S is block diagonal (whose blocks are given by equation (12)) we can diagonalise each block individually. For every irreducible representation of dimension f, we have an f-degenerate frequency [33]. The normal modes that belong to the same irreducible representation are the equal, albeit rotated [31]. This means that an empty viral capsid comprised of 60 aus has 94 different frequencies and normal modes, which is a great reduction from the 354 that we had at the beginning (we have only taken into account *distinct* normal modes, because if they belong to the same irrep they are related through linear transformations: thus we have one frequency per irrep).

Also, by noticing which type of functions are invariant under each irreducible representation we can gain some understanding on the behaviour of the normal modes [34]. In fact, since only symmetric functions of the form $f(x^2+y^2+z^2)$ are invariant under the A irrep, we are able to assure that the 6 normal modes associated with A

are completely symmetric. When these symmetric modes involve translation coordinates, they are called respiration or breathing modes. Also, symmetric oscillations mean that almost all of the aus of the capsid are moving. For the rest of irreducible representations, we only have degenerate asymmetric motions [31]. Visualisations of some of these normal modes both symmetrical and asymmetrical (although for a different model) can be found in [30].

We also need to take into account that the greater the granularity in the CG model (the bigger the basic unit), the lower the frequencies, as they are inversely proportional to the mass. This gives us the intuition that the more symmetric a movement is, a bigger number of basic units are involved in the oscillation, and thus the frequencies are lower, since there is more mass involved in the movement. This is worth investigating further.

I	E	$12C_5$	$12C_5^2$	$20C_3$	$15C_2$
A	+1	+1	+1	+1	+1
T_1	+3	$-2\cos(4\pi/5)$	$-2\cos(2\pi/5)$	0	-1
T_2	+3	$-2\cos(2\pi/5)$	$-2\cos(4\pi/5)$	0	-1
G	+4	-1	-1	+1	0
H	+5	0	0	-1	1

Table 1: Character table of the icosahedron point group

7 Application of the AUM-60 to a real virus

En esta sección comparamos las predicciones de nuestro modelo de 60 unidades asimétricas con los hallazgos de un grupo surcoreano en el virus del Zika, y comprobamos como únicamente con el uso de las propiedades de simetría se pueden predecir de forma cualitativa el número de modos normales completamente simétricos.

Even though the Zika virus has been relatively unknown for the public since its discovery in 1947, it has gained recent popularity due to the epidemic that affected Central America, the Caribbean and South America between 2015-2016. In particular, it was particularly well-know for its association with a huge increase in the cases of microcephaly in new-born children whose mothers were infected by the virus. This has motivated the latest research on the virus, specially directed towards a vaccine or a treatment. The knowledge acquired regarding the biological process of infection, among other processes, has been limited so far [1]. There is no vaccine or treatment available at the moment.

Thanks to cryo-electron microscopy, it has been revealed that the Zika virus capsid possesses an icosahedral symmetry [1]. The capsid under study has as unit structure a dimer which is formed by 2 heterodimers and each dimer is connected to the lipid membrane of the virus [1]. Since there are 90 dimers, in principle we would not be able to apply our model (made for 60 au), but every trio of heterodimers acts *de facto* as an assymetric unit. Therefore, we have 60 aus and our model is applicable.

A research group in South Korea has made explicit calculations of 6 lowest normal modes of the capsid. They treated each heterodimer as a single particle, but simulated the shell with a very complete method that even includes the thermodynamical properties of the simulation box (like pressure, temperature...), as well as other molecules, such as water [1]. As a potential, they used a simple spring-like interaction with a cut-off distance to suppress long-range (and thus weak).

Despite the fact that the South Korean group did not model the heterodimers as rigid bodies, their findings about the symmetries of the lowest frequencies are in good agreement with the predictions of the AUM-60. They found that the 6 lowest frequencies have fully symmetric movements around the symmetry vertices, although only 5 of them present breathing movements [1]. This is in agreement with the predictions of the AUM-60, which predict 6 fully symmetric modes. Modes A and C have breathing movements about the 3-fold vertices, with the C mode movement originated by torsion of certain parts of the heterodimers [1]. Modes B and E also have breathing movements, but around the 5 fold vertices: they only differ in the narrowness of the oscillation [1]. Mode D its similar to A, B and E, but around the 2-fold vertices [1]. Finally, mode F is purely torsional, around the 5-fold vertices [1].

About the interpretation of the simulation results of the South Korean group, it

is merely qualitative since they only show a certain number of arrows in each mode which is substantially less than the 60 arrows our model would produce. Therefore, the movements shown are an average over the real movements that our 60 aus model could obtain. It would be interesting, through the adequate choice of force constants, to reproduce their results with our model.

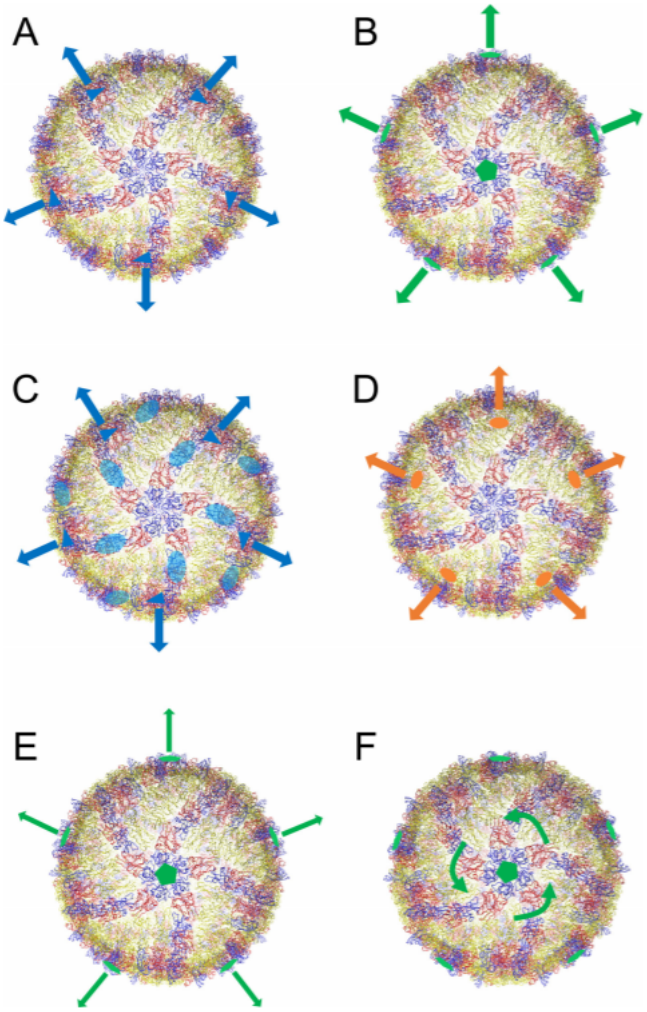

Figure 10: The 6 fully symmetric modes of the Zika virus that belong to the A irreducible representation [1].

8 Conclusions

En este apartado condensamos las conclusiones y resultados a los que hemos llegado utilizando el modelo de 60 unidades asimétricas para capsides icosaédricas y proponemos varias líneas de trabajo futuro.

In this project we have shown how important is the application of physical theories to better understand biological objects such as viruses. We have seen how these two branches of science, biology and physics, can be intertwined to offer a better and deeper description of reality and actually complement each other.

But the main work has been to develop the basis of a coarse-grained model that uses 60 asymmetric units and symmetry arguments for the description of icosahedral capsids of T1 viruses. First we have created a contact matrix for the asymmetric units under the restriction that only the *contact-neighbours* interact with each other. Then we calculated how many normal modes and associated frequencies a 60 aus capsid has to have due to the symmetry properties of the icosahedral point group. In particular, we proved that there are 94 different frequencies and normal modes, and that 6 of those frequencies belong to fully symmetric oscillations. In fact, we compared our results to the explicit calculations of a south corean research group that computed the 6 lowest normal modes of the Zika virus, since it presents an icosahedral symmetric capsid that can be described with 60 asymmetric units [1]. The south corean group, without making symmetry arguments, obtained 6 fully symmetric modes that correspond to the ones our model predicted.

From here, we propose the following paths to continue this work:

- Investigate the possible relation between the symmetry of a normal mode and its frequency. Since in general symmetric modes involve the movement of more asymmetric units than other modes, and the frequency is inversely proportional to the square root of the mass, it is intuitive to think that a symmetric mode would have a lower frequency than other modes. In particular, fully symmetric modes involving translations (breathing modes) should have lower frequencies than fully symmetric rotational modes.
- Develop a better code to construct the contact matrix and try to find an optimal way of numerating the capsid (if it exists).
- Complete the AUM-60 model by proposing an interaction potential and constructing (with the help of the contact matrix developed in this project) a potential matrix for the capsid, and to compute explicitly the normal modes and its frequencies, as well as other parameters of interest such as the critical concentration for optimal self-assembly, etc.
- Make a fit using the force constants (free parameters) of the completed AUM-60 model to reproduce the results of the south corean group [1].

References

[1] Byung Ho Leea, Soojin Joa, Moon-ki Choia, Min Hyeok Kimb, Jae Boong Choia,and Moon Ki Kima: "Normal mode analysis of Zika virus". Computational Biology and Chemistry, vol.72, pp. 53-61, 2018.

[2] Hiroshi Fukuhara, Yasushi Ino and Tomoki Todo: "Oncolytic virus therapy: A new era of cancer treatment at dawn". Cancer Sci, vol. 107, pp. 1373–1379, 2016.

[3] https://courses.lumenlearning.com/microbiology/chapter/viruses/

[4] Martina Pozar and Rudolf Podgornik: "Physical virology". http://www-f1.ijs.si/ rudi/sola/Physicalvirology(M.Pozar).pdf

[5] Antoni Luque Santolaria: "Structure, Mechanical Properties, and Self-Assembly of Viral Capsids", PhD thesis, University of Barcelona, 2011

[6] D.L.D. Caspar and A. Klug: "Physical principles in the construction of regular viruses". Cold Spring Harbor Quant. Bio., vol. 27, pp. 1-24, 1962.

[7] H.S.M. Coxeter: "Introduction to geometry". Wiley, 2nd ed., 1989.

[8] Roya Zandi, Bogdan Dragnea, Alex Travesset, Rudolf Podgornik: "On virus growth and form". 2019.

[9] Ranjan V. Mannige and Charles L. Brooks: "Periodic Table of Virus Capsids: Implications for Natural Selection and Design". PLoS ONE 5(3): e9423, 2010.

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

[11] Lokendra Poudel: "Deciphering of packing signal hypothesis in bacterophage RNA recognition by the MS2 capsid protein in virus assembly". PhD thesis, University of Missouri-Kansas City, 2014.

[12] Ruben D. Cadena-Nava, Mauricio Comas-Garcia, Rees F. Garmann, A. L. N. Rao, Charles M. Knobler, and William M. Gelbart: "Self-Assembly of Viral Capsid Protein and RNA Molecules of Different Sizes: Requirement for a Specific High Protein/RNA Mass Ratio". Journal of Virology, vol. 86, pp. 3318-3326, 2012.

[13] Minna M. Poranen, Anja O. Paatero, Roman Tuma and Dennis H. Bamford: "Self-Assembly of a Viral Molecular Machine from Purified Protein and RNA Constituents". Molecular Cell, vol. 7, pp. 845.854, 2001.

[14] Michael F. Hagan: "Modeling Viral Capsid Assembly". Advanced Chemichal Physics, vol. 155, pp. 1-68, 2014.

[15] Endres D. and Zlotnick A.: "Model-based analysis of assembly kinetics for virus capsids or other spherical polymers". Biophysics Journal, vol. 83, pp. 1217–1230, 2002.

[16] S. Chandrasekhar: "Stochastic problems in physics and astronomy". Reviews of Modern Physics, vol. 15, pp. 1-89.

[17] W. H. Roos, R. Bruinsma and G. J. L.Wuite: "Physical virology". Nature Physics, vol. 6, pp. 733.743.

[18] A. Evilevitch, L. Lavelle, C. M. Knobler, E. Raspaud and W. M. Gelbart: "Osmotic pressure inhibition of DNA ejection from phage". Proceedings of the National Academy of Sciences of the United States of America, vol. 100, pp. 9292-9295.

[19] M. Marchetti, G. J. L. Wuite and W. H. Roos: "Atomic force microscopy observation and characterisation of single virions and virus-like particles by nano-indentation". Current opinion in Virology, vol. 6, pp. 82-88.

[20] Everardo Vega, Jay Garland and Suresh D. Pillai: "Electrostatic Forces Control Nonspecific Virus Attachment to Lettuce". Journal of Food Protection, vol. 71, No. 3, pp. 522–529, 2008.

[21] Antonio Siber, Anze Losdorfer Bozic and Rudolf Podgornik: "Energies and pressures in viruses: contribution of nonspecific electrostatic interactions". Physical Chemistry Chemical Physics, vol. 14, pp. 3746-3765, 2012.

[22] S. Safran: "Statistical Thermodynamics Of Surfaces, Interfaces, and Membranes". Cornell University Press, Westview Press, 1st edn, 2003.

[23] E. C. Holmes: "The Evolution and Emergence of RNA Viruses". Oxford University Press, Oxford, 1st edition, 2009.

[24] P. Ceres and A. Zlotnick: "Biochemistry". Vol. 41, pp. 11525–11531, 2002.

[25] Inmuno genetics information system, IMGT. http://www.imgt.org/IMGTeducation/Aide-memoire/_UK/aminoacids/abbreviation.html

[26] Juan R. Perilla and Klaus Schulten: "Physical properties of the HIV-1 capsid from all-atom molecular dynamics simulations". Nature communications, vol.8, pp. 15959, 2017.

[27] Parimal Kamar and Michael Feig: "Hybrid All-Atom/Coarse-Grained Simulations of Proteins by Direct Coupling of CHARMM and PRIMO Force Fields". Journal of Chemical Theory and Computation, vol. 13, pp. 5753-5765, 2017.

[28] J. M. Gómez Llorente, J. Hernández-Rojas and J. Bretón: "A minimal representation of the self-assembly of virus capsids". Soft Matter, vol. 10, pp.3560.3569, 2014.

[29] D. Reguera, J. Hernández-Rojas and J. M. Gómez Llorente: "Kinetics of empty viral capsid assembly in a minimal model". Soft Matter, vol. 15, pp. 7166-7172, 2019.

[30] M. Martín-Bravo, J. M. Gómez Llorente and J. Hernández-Rojas: "A minimal coarse-grained model for the low-frequency normal mode analysis of icosahedral viral capsids". Soft Matter, vol. 16, pp. 3443-3455, 2020.

[31] Herman W. T. van Vlijmen and Martin Karplus: "Normal Mode Calculations of Icosahedral Viruses with Full Dihedral Flexibility by Use of Molecular Symmetry". Journal of Molecular Biology, vol. 350, pp. 528-542, 2005.

[32] L. D. Landau and E. M. Lifshitz: "Non-relativistic quantum mechanics". Pergamon Press, 3rd edition, 1977.

[33] Kasper Peeters and Anne Taormina: "Group theory of icosahedral virus capsid vibrations: A top-down approach". Journal of theoretical biology, vol. 256, pp. 607-624, 2009.

[34] Character table for point group I: http://symmetry.jacobs-university.de/cgi-bin/group.cgi?group=905option=4

Appendix: the contact matrix of the AUM-60

Due to the size of the matrix (60x60) and thanks to its symmetry only the non-zero elements above the diagonal are shown. To obtain the non-zero elements below the diagonal, just swap $i \longleftrightarrow j$ and $a \longleftrightarrow b$.

$a_{i,j}$	Int. Type	$a_{i,j}$	Int. Type	$a_{i,j}$	Int. Type	$a_{i,j}$	Int. Type
$a_{1,2}$	1a	$a_{12,41}$	2a	$a_{25,27}$	1b	$a_{38,49}$	2a
$a_{1,3}$	1b	$a_{13,14}$	1a	$a_{25,28}$	2a	$a_{39,52}$	2b
$a_{1,4}$	2a	$a_{13,15}$	1b	$a_{26,27}$	1a	$a_{39,54}$	3
$a_{1,13}$	2b	$a_{14,15}$	1a	$a_{26,56}$	3	$a_{40,41}$	1a
$a_{1,15}$	3	$a_{14,44}$	3	$a_{26,57}$	2b	$a_{40,42}$	1b
$a_{2,3}$	1a	$a_{14,45}$	2b	$a_{27,28}$	3	$a_{40,46}$	3
$a_{2,15}$	2a	$a_{15,44}$	2a	$a_{27,29}$	2b	$a_{40,47}$	2b
$a_{2,32}$	3	$a_{16,17}$	1a	$a_{27,56}$	2a	$a_{40,51}$	2a
$a_{2,33}$	2b	$a_{16,18}$	1b	$a_{28,29}$	1a	$a_{41,42}$	1a
$a_{3,4}$	3	$a_{16,19}$	2a	$a_{28,30}$	1b	$a_{41,46}$	2a
$a_{3,5}$	2b	$a_{16,28}$	3	$a_{29,30}$	1a	$a_{42,49}$	2b
$a_{3,32}$	2a	$a_{16,39}$	2b	$a_{29,59}$	3	$a_{42,51}$	3
$a_{4,5}$	1a	$a_{17,18}$	1a	$a_{29,60}$	2b	$a_{43,44}$	1a
$a_{4,6}$	1b	$a_{17,30}$	2a	$a_{30,59}$	2a	$a_{43,45}$	1b
$a_{4,7}$	2a	$a_{17,47}$	3	$a_{31,32}$	1a	$a_{43,48}$	2a
$a_{5,6}$	1a	$a_{17,48}$	2b	$a_{31,33}$	1b	$a_{43,58}$	3
$a_{5,35}$	3	$a_{18,19}$	3	$a_{31,55}$	3	$a_{43,59}$	2b
$a_{5,36}$	2b	$a_{18,20}$	2b	$a_{31,56}$	2b	$a_{44,45}$	1a
$a_{6,7}$	3	$a_{18,46}$	2a	$a_{31,60}$	2a	$a_{44,58}$	2a
$a_{6,8}$	2b	$a_{19,20}$	1a	$a_{32,33}$	1a	$a_{45,46}$	2b
$a_{6,35}$	2a	$a_{19,21}$	1b	$a_{32,55}$	2a	$a_{45,49}$	3
$a_{7,8}$	1a	$a_{19,22}$	2a	$a_{33,58}$	2b	$a_{46,47}$	1a
$a_{7,9}$	1b	$a_{20,21}$	1a	$a_{33,60}$	3	$a_{46,48}$	1b
$a_{7,10}$	2a	$a_{20,50}$	3	$a_{34,35}$	1a	$a_{47,48}$	1a
$a_{8,9}$	1a	$a_{20,51}$	2b	$a_{34,36}$	1b	$a_{49,50}$	1a
$a_{8,37}$	3	$a_{21,22}$	3	$a_{34,52}$	3	$a_{49,51}$	1b
$a_{8,38}$	2b	$a_{21,23}$	2b	$a_{34,53}$	2b	$a_{50,51}$	1a
$a_{9,10}$	3	$a_{21,50}$	2a	$a_{34,57}$	2a	$a_{52,53}$	1a
$a_{9,11}$	2b	$a_{22,23}$	1a	$a_{35,36}$	1a	$a_{52,54}$	1b
$a_{9,38}$	2a	$a_{22,24}$	1b	$a_{35,52}$	2a	$a_{53,54}$	1a
$a_{10,11}$	1a	$a_{22,25}$	2a	$a_{36,55}$	2b	$a_{55,56}$	1a
$a_{10,12}$	1b	$a_{23,24}$	1a	$a_{36,57}$	3	$a_{55,57}$	1b
$a_{10,13}$	2a	$a_{23,53}$	3	$a_{37,38}$	1a	$a_{56,57}$	1a
$a_{11,12}$	1a	$a_{23,54}$	2b	$a_{37,39}$	1b	$a_{58,59}$	1a
$a_{11,41}$	3	$a_{24,25}$	3	$a_{37,43}$	3	$a_{58,60}$	1b
$a_{11,42}$	2b	$a_{24,26}$	2b	$a_{37,44}$	2b	$a_{59,60}$	1a
$a_{12,13}$	3	$a_{24,53}$	2a	$a_{37,48}$	2a		
$a_{12,14}$	2b	$a_{25,26}$	1a	$a_{38,39}$	1a		

Note: the positions $a_{i,j}$ of the non-zero elements as well as the type of interaction between aus depend on the numbers allocated to each of them, which is an arbitrary choice. The numeration used for this particular matrix is the shown below.

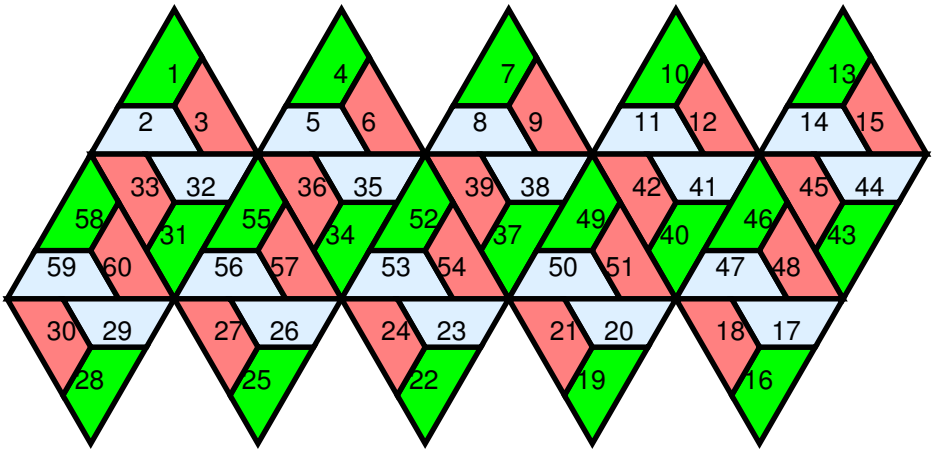

Figure 11: [Original work from Dr. José María Gómez Llorente]