Facultad de Ciencias
Universidad de La Laguna

PHYSICAL MODELS OF THE STRUCTURE AND PROPERTIES OF VIRAL CAPSIDS:

Calculation of minimum potential in the free state

and after undergoing deformation in ligations.

END-OF-DEGREE PROJECT

Written by: Erick Delgado Araujo

Supervised by: José María Gómez Llorente

Acknowledgements

To all the teachers who endured my presence in their classes. Especially to those who went the extra mile for their students.

To my project tutor, to whom I have probably given more headaches than I would like to admit. His guidance was fundamental for developing the simulations.

To my mother, whose patience was put to test while the development of this project.

To all my friends and their unwavering support. They helped me see the light at the end of the tunnel.

To my classmates. Their constant efforts and passion for physics helped me find strength during these last years.

To my family and the faith they put in me to finally defend the results of my research.

To my late lovebird companion. You will never be forgotten.

All of you made me who I am today. I thank you from the bottom of my heart.

Index

Abstract..... 1

Introduction..... 3

Mice Minute Virus: Structure and Characteristics..... 8

Starting assumptions and objectives.....14

Indentation simulations..... 19

Results and conclusion..... 33

References..... 35

Abstract

Comprehending the structure and properties of viral capsids is essential in the field of virology, as it provides valuable insights into the behaviour and mechanics of viruses. This research begins with an overview on these infectious agents, emphasising the significance of capsids as protein-based envelopes that protect the viral genome. Capsids play a crucial role in viral replication, assembly, and interactions with host cells. Understanding their structure is key to deciphering the mechanisms underlying viral infections.

This end-of-degree project focuses on investigating the response of capsids to external forces and developing a Python code to calculate the potential energy based on capsid coordinates. We are also going to focus our attention on whether or not different configurations of a simple virus can affect the outcome of the simulations, and if those results are fully capable of defining the virus itself, and its properties, or if more research for other properties is required to make such claims.

Nonetheless, these data provide interesting information about the stability and energetics of capsids, enabling us to investigate the effects of external forces on viral structures. By utilising simulations and analyses tailored to the icosahedral structure commonly observed in viral capsids, this project aims to enhance our understanding of viral mechanics and which methods are preferable and more accurate to establish their virtual behaviour to correlate their laboratory-generated counterparts.

The Caspar and Klug models are employed to describe the geometric arrangement of capsomers, the protein subunits forming the capsids. These models provide a foundation for studying their assembly and stability in which this project will be based upon. Therefore, the starting point of this project will consist in a brief overview of the basics of virology, specifying which tenets we are upholding for this program to be relevant, and then a review of the previous investigations and results upon which we are basing this work, whose foundations will be investigated to create an alternative method of simulation. Finally, according to the results, we will be able to tell if such changes resulted in a better, worse, or simply different outcome.

Comprender la estructura y propiedades de las cápsides virales es esencial en el campo de la virología, ya que proporciona una valiosa perspectiva sobre el comportamiento y la mecánica de los virus. Este trabajo comenzará con una visión general de estos agentes infecciosos, haciendo hincapié en la importancia de las cápsides como envolturas basadas en proteínas que protegen el genoma viral. Éstas desempeñan un papel crucial en la replicación viral, el ensamblaje y las interacciones con la célula huésped. Entender su estructura es clave para descifrar los mecanismos subyacentes a las infecciones virales.

Este proyecto de fin de carrera se centra en investigar la respuesta de las cápsides a fuerzas externas y en desarrollar un código en Python para calcular la energía potencial basada en sus coordenadas. También nos centraremos en determinar si diferentes configuraciones de un virus simple pueden afectar el resultado de las simulaciones y si esos resultados son completamente capaces de definir el virus en sí y sus propiedades, o si se requiere más investigación para afirmar tal declaración.

No obstante, estos datos proporcionan información muy interesante sobre la estabilidad y la energética de las cápsides, lo que nos permite investigar los efectos de las fuerzas externas en las estructuras virales. Utilizando simulaciones y análisis adaptados a la estructura icosaédrica comúnmente observada en las cápsides virales, este proyecto tiene como objetivo mejorar nuestra comprensión de la mecánica viral y determinar métodos preferibles y más precisos para establecer su comportamiento virtual y correlacionarlo con sus contrapartes generadas en el laboratorio.

Se emplean los modelos de Caspar y Klug para describir la disposición geométrica de los capsómeros, las subunidades proteicas que forman las cápsides. Estos modelos proporcionan una base para estudiar su ensamblaje y estabilidad, en la cual se basará este proyecto. Por lo tanto, nuestro punto de partida consistirá en una breve visión general de los fundamentos de la virología, especificando qué principios estamos sosteniendo para que este programa sea relevante, y luego una revisión de las investigaciones y resultados anteriores en los que basamos este trabajo, cuyos fundamentos se investigarán para crear un método alternativo de simulación. Finalmente, de acuerdo con los resultados, podremos determinar si tales cambios resultaron en un resultado mejor, peor, o simplemente diferente.

Introduction

Abstract. *Los virus exhiben cautivadores mecanismos de autopreservación en numerosos ecosistemas del planeta. Aunque han causado innumerables fatalidades, solo aproximadamente 11,000 de sus millones de especies han sido estudiadas detenidamente en la ciencia moderna. Todos comparten una estructura básica con material genético (ARN o ADN) y una cápside formada por proteínas. Este proyecto indaga en las respuesta de la cápside a fuerzas externas con simulaciones en Python, explorando configuraciones de virus simples para entender mejor su mecánica y métodos precisos para definir su comportamiento virtual.*

Infectious agents that need living cells to replicate, viruses are found in almost all of the ecosystems of our planet [1], being able to spread through a lot of different organisms and exhibiting interesting mechanisms for their self-preservation. An incredibly large number of lives have perished at their hands throughout history, yet to this day modern science has only been able to analyse in detail approximately 14.000 of their millions of species [2].

It was the year 1887 when Dr. John Buist observed for the first time in human history a virus: a cumulus of vaccinia viruses [3]. It was at that moment that an important question was answered to the scientists of his time, as the remains of some formerly infected organisms (such as the tobacco plant) were discovered to be still infectious, even after being liquefied and filtered through a Chamberland filter (whose pores are smaller than bacteria) [4].

Viruses can be very different in nature depending on their species, but all of them share a common structure: at least an ARN or ADN string (simple or double) to store their genetic information (which can be also circular or segmented), and a capsid shell, formed by one or more different kinds of proteins, being standart that the bigger the capsid is, the more types of proteins are found forming it. All capsids are

formed by a number of subunits, groups of proteins that configure a standard building unit. A virus can be formed by a single type of subunit, or by combinations of different ones. They can form a capsid included in one of the three different categories of shape: icosahedral (the most common ones), helical and complex (for those which do not fall into the previous categories, e.g.: Bacteriophages). In addition, an important number of viruses also possess a lipid bilayer gained from their host cell, an example being the recently pandemic originator: the SARS-CoV-2 (which causes COVID-19) [5].

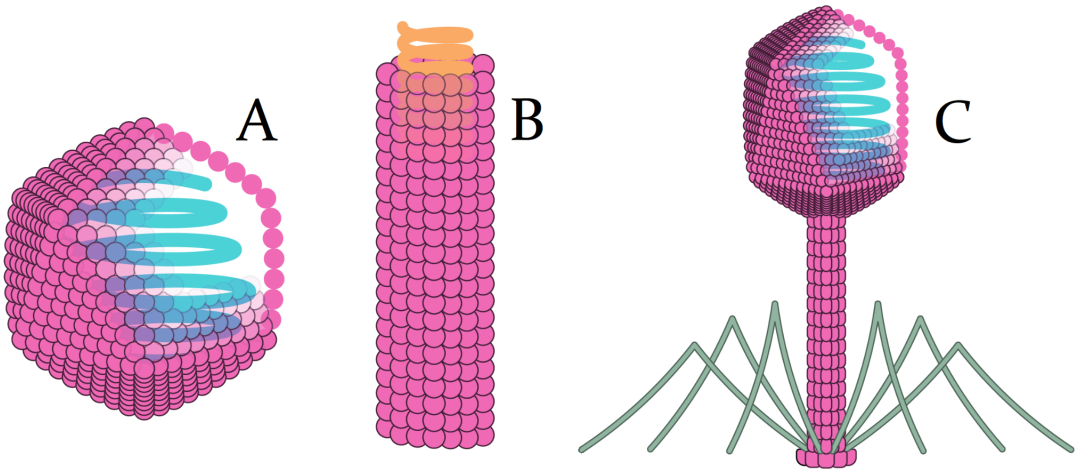


Figure 1: A visual example of the three categories of shape a virus can be classified: icosahedral (A), helical (B) and complex, represented by a bacteriophage (C). Modified from [6].

Simple virus' capsids tend to adopt shapes that minimise their bonding energy, as they are easier to both replicate and assemble. But in nature, an immense variety of shapes and sizes can be found, forming complex and intricate structures that target specific types of host cells. Being able to replicate effectively, having a stable composition in the environment they inhabit and creating mechanisms to avoid their destruction by their host's immune system (in those cases where they have one) will translate in a successful replication, hence the efficiency of the virus affects its survivability.

Poorly constructed viruses that are prone to crumble, or those who do not replicate fast enough to infect other host cells before ending its life cycle, are bound

to disappear. For those reasons, a very interesting targeting method to fight against viral pathogens called Capsid-Targeted Retroviral Inhibition (CTVI) is to inhibit the formation of its capsid by incorporating an inhibitor protein, which will then mingle with the capsid proteins, form strong bonds with them, and impede their use as a capsid protein [7].

In the discrete space, the icosahedral shape is the most symmetrical, being part of the *I* group of three-dimensional finite structures along with the dodecahedron (following Schoenflies notation [8]) and not qualifying to be part of the I_h symmetry group due to capsid proteins being asymmetric, and therefore icosahedron-shaped capsids lacking an inversion centre. Those capsids have three possible rotational symmetries: Of order 2 (180º rotations), around an axis through midpoints of edges of the icosahedron (2-fold symmetry, or also *C2*); of order 3 (120º rotations), around an axis through face centres of the icosahedron (3-fold symmetry, or also *C3*); and of order 5 (72º rotations), around an axis through vertices of the icosahedron (5-fold symmetry, or also *C5*). These three symmetry configurations will be studied in this project, as we predict differences between their data output.

Another common method to categorise different kinds of viruses is the Baltimore classification [9] (named after its creator, Nobel Prize-winning biologist David Baltimore), which consists in the separation of the pathogens into seven categories depending on their mechanism of mRNA production. It is usually combined with the unified taxonomy method of the International Committee on Taxonomy of Viruses (ICTV) [10], which puts great weight on the viruses' properties to maintain uniformity in their categories.

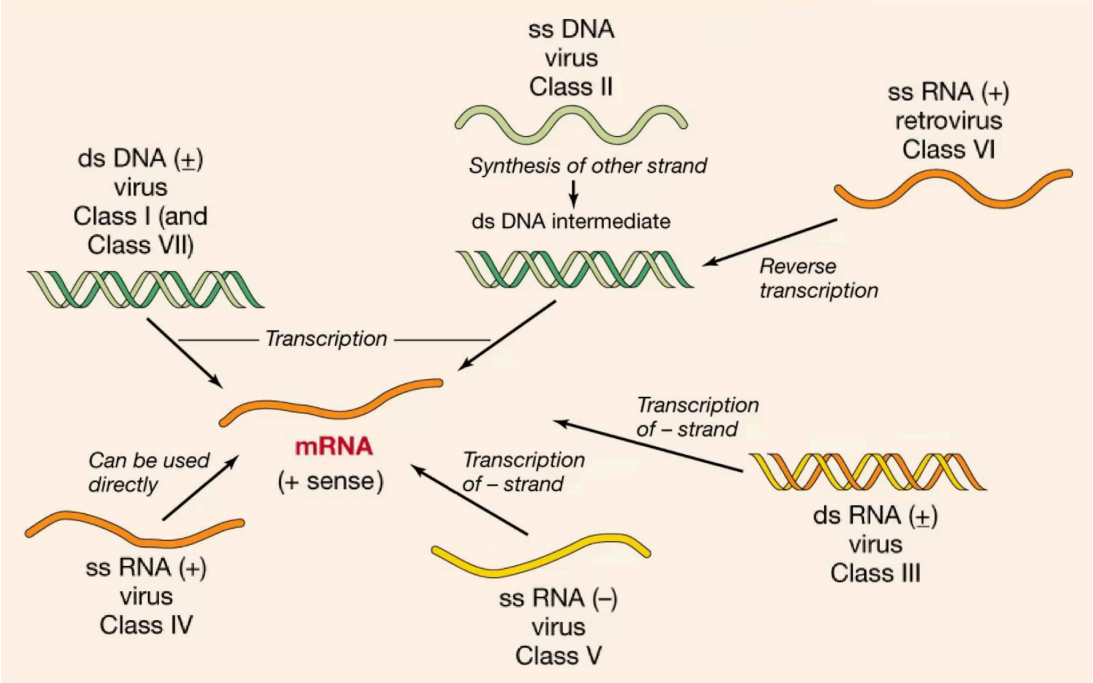


Figure 2: A diagram depicting how each one of the genomes found in a different category of the Baltimore classification undergoes the process of transcription. It is interesting to note the difference between Class I and Class VII viruses: the latter goes through an intermediate state of RNA in their life cycle. Taken from [11].

The reproductive cycle of a virus initiates when it manages to enter a host cell. There, the capsid disassembles, and the genetic information inside it unpacks. Each virus has its own process of replication, being the bigger and more complex ones able to make it mostly by themselves, only using some resources of the host cell; unlike simple-structured viruses, where the genetic code has to infiltrate the nucleus to be replicated there, forcing the cell to create itself new copies of the virus. As for the capsid formation, it entails a complex process in which the proteins group themselves into capsomeres, the capsid's basic unit of construction. From there, depending on the virus's nature and size, the capsomeres connect into either penton, hexon or trimer subunits (the latter mostly common in tiny viruses) and then start to self assemble into the capsid, given that the conditions around them are adequate (subunit concentration, free energy of the capsid in each stage of its growth, pH of the surroundings) [12]. This process has also been studied and adapted to computer

simulations, existing numerous models for a grand variety of viruses [13]. The possibility of a capsid formed by dimer subunits (subunits formed by two proteins) exists, but for this project that prospect will not be taken into account, as it happens in very few cases.

As it has been stated before, we currently hold very little information involving the composition of most of the existing viruses. That is why numerous studies are being researched right at this moment, most of them related to the search of a universal understanding of their microscopic behaviour (which has proven to be quite difficult and expensive to investigate) related to macroscopic properties easily measurable. A breakthrough of that nature would surely become one of the most important discoveries of the century, making the production of vaccines of future viruses much more simple, and thus, saving a gargantuan amount of lives and resources. It is an honour to be a microscopic cog on the wheel of change for a brighter future.

Mice Minute Virus: Structure and Characteristics

Abstract. *El estudio se centra en el Virus Minuto del Ratón (MMV), un patógeno altamente contagioso en ratones de laboratorio que se propaga a través de secreciones como la orina y las heces. Está caracterizado por una cadena simple de ADN y una cápsida icosaédrica con 60 proteínas idénticas. Explorando variaciones en las configuraciones de las subunidades, el proyecto aplica el modelo de Caspar y Klug, revelando los diversos diseños posibles para virus icosaédricos en una cuadrícula hexagonal según su tamaño y número de proteínas.*

The Mice Minute Virus (MMV) is a highly contagious pathogen found commonly in laboratory mice. It can be transmitted through an infected mouse's secretions, especially their urine and faeces, and has two main variant forms in MMVp, a strain that infects fibroblastic cells, and MMVi, which reproduces through T lymphocytes [14]. Although their prevalence has declined over the years [15], it can still be found as a contaminant while other studies of laboratory mice are taking place. Examples range from tumours or cell mutations to the growth of other pathogens inside the poor rodent subject. In those cases, the results normally show deviations of the expected outcomes, and while a simple correction can be made, it is still not the ideal situation. There are no obvious signs of infection on adult mice, but it can produce multiple organ damage during their foetal development [16].

The MMV is usually showcased as an example of tiny viruses' behaviour due to the fact that it is the smallest virus that can be titrated with ease, which is very helpful to counter the contamination effect it can have on mice-related experiments [17]. This virus is also the main focus of this study, being that the calculations have been made specifically for a capsid of its characteristics.

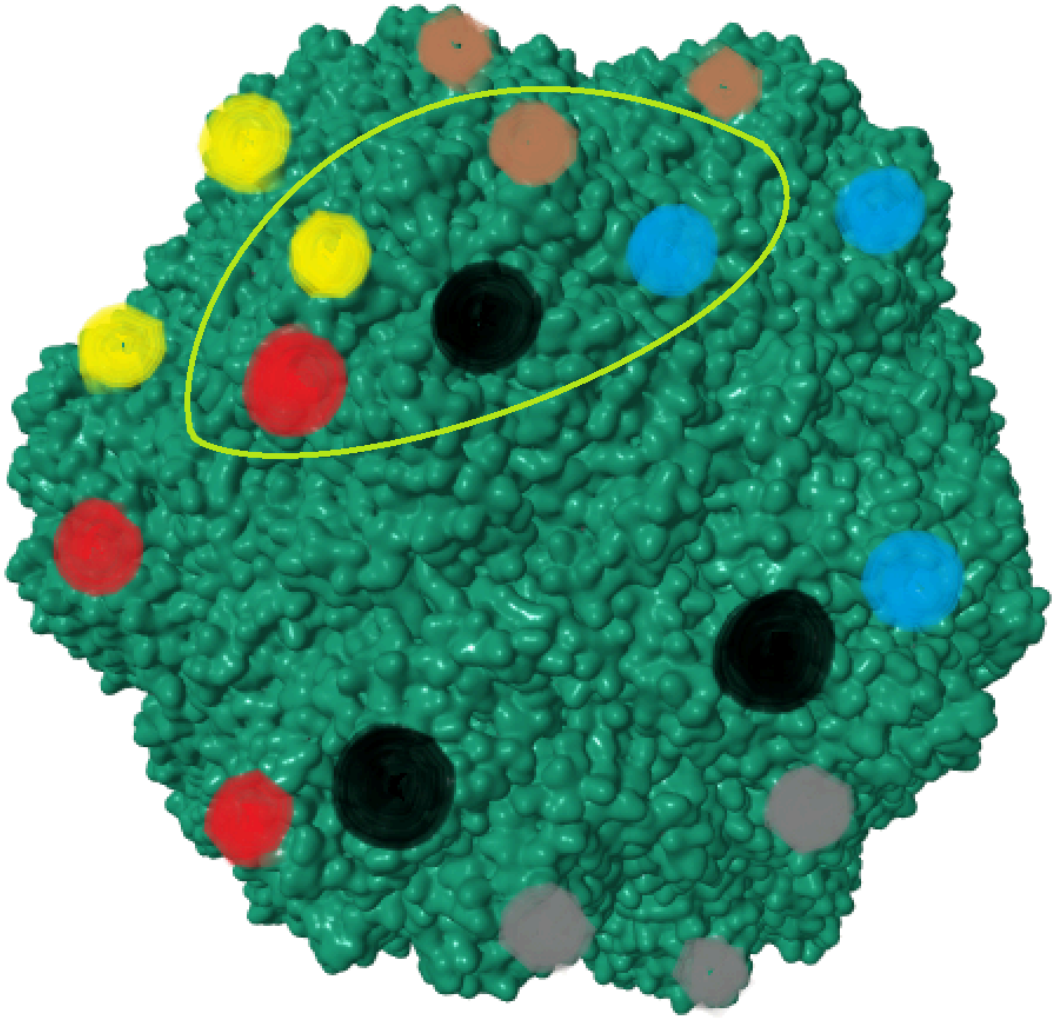


Figure 3: A representation of the MMV capsid found in the VIPERdb website. To help distinguish between different proteins, each of them has been given a coloured node, whereas every trio of same-coloured proteins represent the subunit of the capsid: the trimers. An area including five different-coloured nodes is also marked, as it represents our assumed-to-be subunit for this project: the pentons. Modified from [18]

It is formed by a single strand of DNA (making it a Class II virus by the standards of the Baltimore classification), and an icosahedral capsid with an average radius of 138 Å and constituted by 60 identical proteins, called **Murine Minute Virus Coat Protein**. This information has been gathered from the Virus Particle Explorer Database (most known as VIPERdb [18]), where it can also be noted that the method used to examine the MMV's properties was the crystallisation and posterior

diffraction with X-rays [19]. Previous experiments have proven that this certain virus' subunits take the form of trimers (three capsid proteins joint in a triangular formation), creating a 20-piece structure [20]. Even so, for the duration of this project, we will presume that its subunits are pentons (five capsid proteins joint in a pentagonal formation), as another of our objectives is to establish if a different initial assumption of its configuration does change the results of the calculations significantly, or in the other hand, a similar result is found, having to make extra experiments to define a certain virus' subunits.

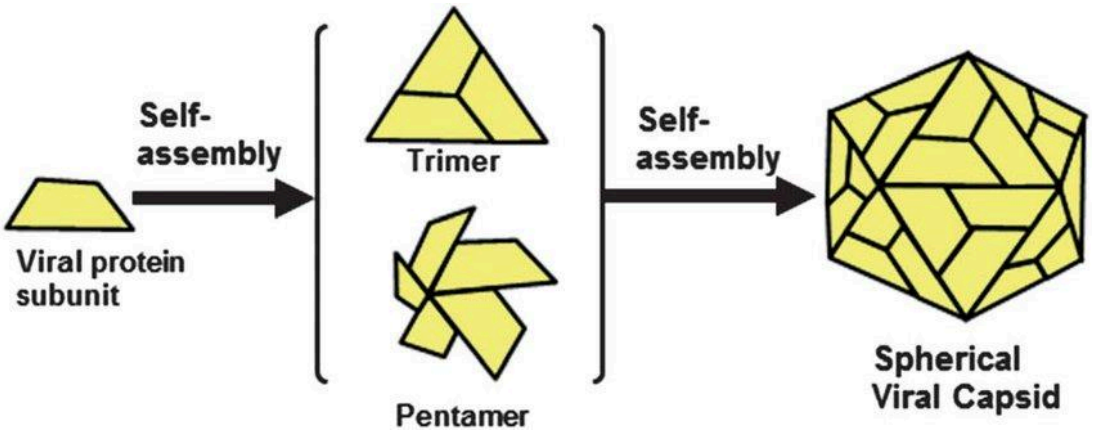


Figure 4: A visual representation of what a trimer and a penton subunit represents when discussing capsid self-assembly in a $T = 1$ virus: three or five capsomeres, respectively, joint as a subunit which in turn serve as a middle step for the final capsid configuration. In this particular scenario, as the viral capsid has only space for 60 proteins, hexons do not have the possibility of appearing. Taken from [21].

We are also going to utilise Caspar and Klug's model of icosahedral viruses and their structure [22], which is a generalised representation of a great variety of virus sizes and their properties. A two-dimensional hexagonal grid with the 20-faced icosahedron unfolded in it [Figure 6], aligning each vertex with the centre of a hexagon, can display how viruses with Caspar and Klug's capsomere arrangement are structured. The size of the figure is represented by the "Triangulation number", or T, which is calculated by the formula:

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

where *h* and *k* are whole numbers which represent the minimum distance path in between two adjacent vertices counting the two possible movements from hexagon to hexagon [Figure 5].

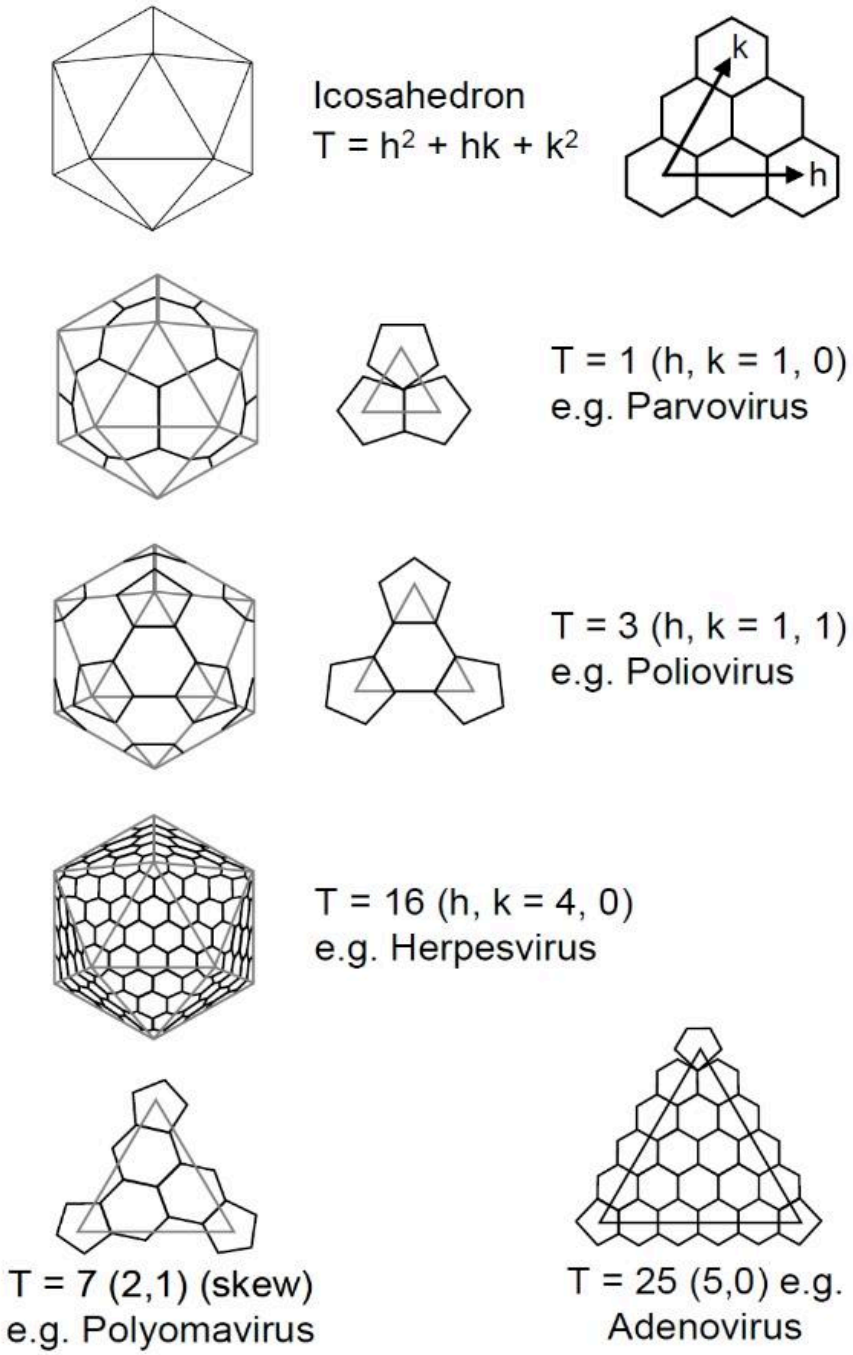


Figure 5: A simple diagram to understand the basic foundations of the Caspar and Klug theorem of icosahedral viruses' configuration. It is shown how viruses with a *T* number different to 1, the basic subunit to form the capsid are pentons and hexons, instead of trimers. Taken from [23].

Each sixth of a hexagon represents a capsid protein. By how the T number operates, an icosahedral virus will have a number of capsid proteins equal to 60*T. According to its formula, the triangulation number can only take certain values, such:

$$T(h,k) = 1(1,0),\ 3(1,1),\ 4(2,0),\ 7(2,1),\ 9(3,0),\ 12(2,2),\ 13(3,1),$$
$$16(4,0),\ 19(3,2),\ 21(4,1),\ 25\ (5,0),\ 27(3,3),\ 28(4,2),\ 31(5,1)\ldots$$

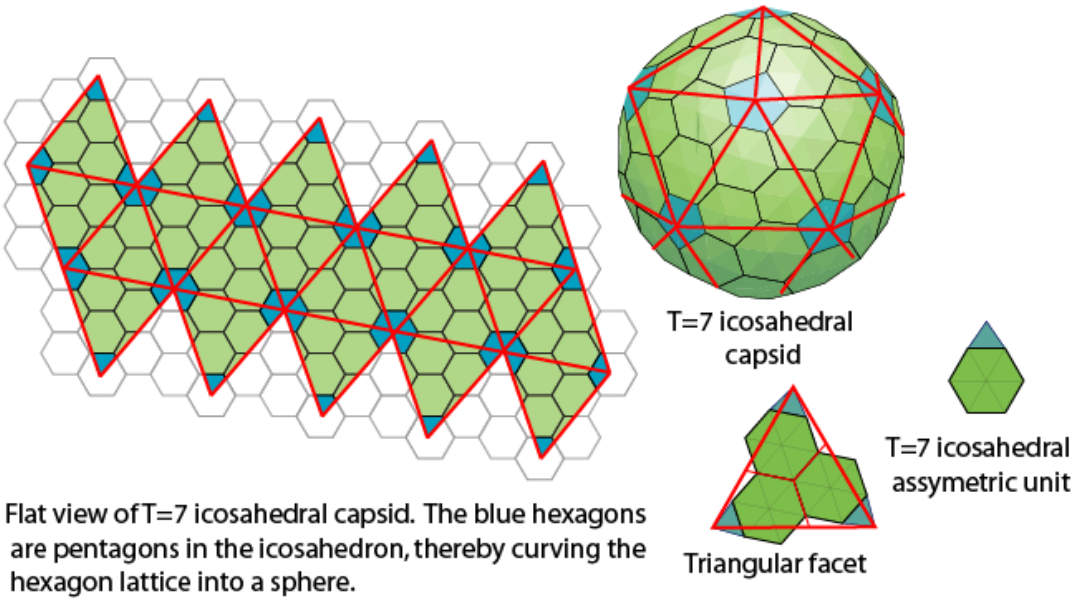


Figure 6: An example of the subunits and the unfolded outline of a $T = 7$ icosahedral capsid. The illustration accurately depicts the hexagon displacement of $h = 2$ and $k = 1$ discussed before, and a capsid made by hexons and pentons as its subunits. Taken from [24].

Additionally, if we consider the space between two centres of hexagons a unit of distance, the value of T is equivalent to the squared length of a side of the triangles. There are numerous capsids that only fulfil certain properties, and are categorised with a pseudo-T number. Most of those have a stable structure but have evolved with asymmetric capsomeres, and thus Caspar and Klug's model cannot correctly represent their appearance and characteristics.

By all these terms and definitions, the MMV is a $T = 1$ icosahedral virus, one of the simplest to analyse, and our choice for this simulated experiment.

Starting assumptions and objectives

Abstract. *Para facilitar la realización de la simulación en un ordenador doméstico, simplificamos la estructura de la cápsida del virus. Agrupamos los capsómeros en pentones, asumiendo simetría axial para facilitar la simulación. Utilizando una fórmula de potencial de energía, modelamos interacciones entre pentones, variando parámetros según sea conveniente. El objetivo es estudiar las variaciones de la energía de enlace en una estructura icosaédrica bajo fuerzas de hendimiento adiabáticas. El código creado para la simulación calcula la energía de enlace y la matriz jacobiana para un conjunto de 60 coordenadas (x, y, z, θ, φ) que representan 12 pentones. Este trabajo prepara el terreno para la siguiente etapa: simular los pentones en las tres simetrías rotacionales de los icosaedros.*

In order to create a simulation simple enough to be able to get processed in a common household PC, some concessions and assumptions about the structure of the capsid of the virus have to be made. Other methods, like the all-atom method (which consists in the representation of every atom of every amino acid of every protein forming the viral capsid) have been studied in multiple examples, providing very useful and specific data about the properties of those concrete structures. It will not be the one we utilise through this project, not only because the simulation is meant to be made in a common household PC, but also to achieve our goal of understanding the general behaviour of every capsid with similar geometrical characteristics, which is hard to do when so much information is gathered, as it is sometimes hard to discern the values corresponding to the general structure to those specific of the proteins of that particular capsid [Figure 7].

It is common to treat groups of objects as one in this kind of experiment, and as we have stated before, we intend to create a simulation where the capsid subunits are pentons. For that reason, we will simplify the simulation by applying a

coarse-grained model, where each penton is treated as a regular cone, assuming then that all of them have axial symmetry (and therefore a rotational coordinate becomes irrelevant to the description of the system). That leaves us with 12 rigid units with five degrees of freedom each: a system described by 60 different coordinates.

Coarse-grained modelling is a widely known and recognized method to simplify most biomolecular systems [25], as a reduction of independent coordinates always helps to create swifter models and faster calculations, albeit the simplification has to be appropriate to truly help the experiment overall. The all-atom method, the coarse-grained method and the continuum method (which consists on the assumption that the whole structure is a single sphere with its own singular elastic constant) are the three main processes of experimental simulations that are used for this kind of experience.

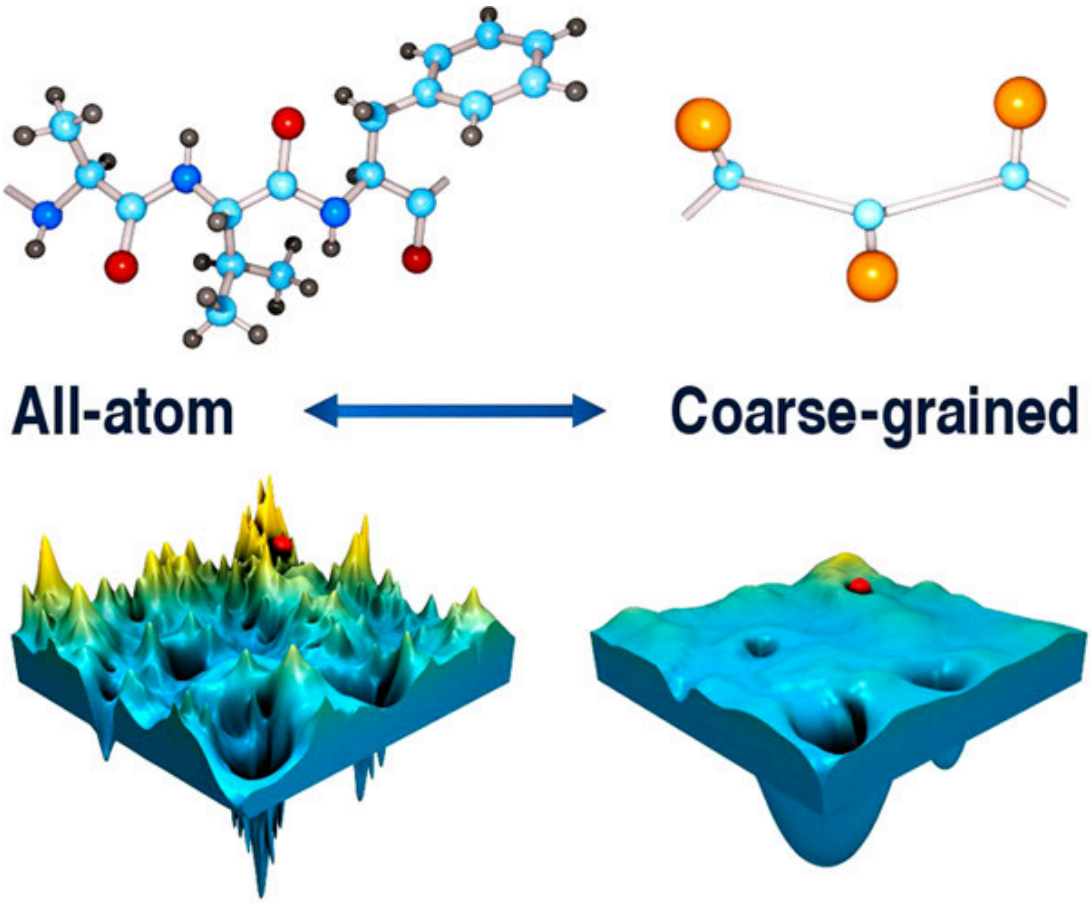


Figure 7: A diagram comparing a biomolecular representation and its energy landscape with their corresponding simplification via coarse-grained modelling. It also serves as an example of how making a coarse-grained

model of a system, while removing an important amount of data and smoothing the results, correctly represents the general features of the original outcome. As such, it proves ideal to simulations like ours, where the main objective lies not on local values or energy, but on the overall of the whole system. Taken from [25].

For this configuration, we will borrow and use an energy potential formula found in my supervisor's group paper [26], where a binary interaction between every two elements (*i* and *j*) is calculated, using the minimum multipole expansion interaction terms needed for the pentons to form an icosahedral structure: monopole-monopole, monopole-dipole, dipole-dipole and monopole-quadrupole:

$$V_{ij} = p_0 F_0(r_{ij}) + p_1 F_1(r_{ij})[1 - \vec{v}_{zi} \cdot \vec{v}_{zj} + 2(\vec{v}_{zi} \cdot \vec{n}_{ij})(\vec{v}_{zj} \cdot \vec{n}_{ij})]$$

$$+ p_2 F_2(r_{ij})[(\vec{v}_{zi} \cdot \vec{n}_{ij} + cos\theta_{en})^2 + (\vec{v}_{zj} \cdot \vec{n}_{ij} - cos\theta_{en})^2]$$

$$F_0(r) = (r_{en}/r)^{2m} - 2(r_{en}/r)^m$$

$$F_i(r) = (r_{en}/r)^{m'm}\ ;\ i \geq 1,\ m' \in (1, 2)$$

In this equation, r_{ij} represents the distance between the centre of the pentons, $\mathbf{n_{ij}}$ is the unitary vector of the distance between *i* and *j*, the parameter *m* escales the magnitude and range of interactions between pentons, and $\boldsymbol{v}_z$ is the normalised vector of the Z axis of either *i* ($\boldsymbol{v}_{zi}$) and *j* ($\boldsymbol{v}_{zj}$). In other words, where the cone-shaped structure we have established for the pentons is directed at. $F_0(r)$ and $F_i(r)$ are functions depending on the distance between the pentons, the latter decaying rapidly while the value of $F_0(r)$ becomes its minimum in the instance when $r = r_{en}$:

$$F_0(r_{en}) = (r_{en}/r_{en})^{2m} - 2(r_{en}/r_{en})^m = 1 - 2 = (-1)$$

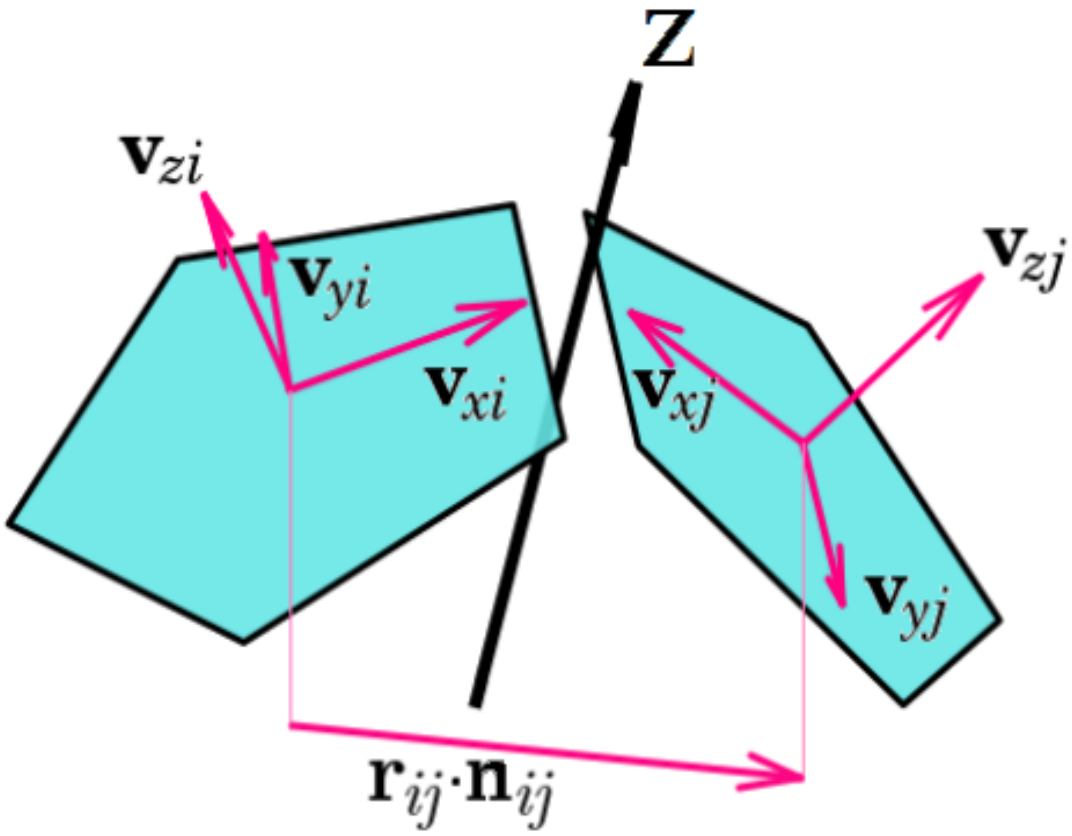


Figure 8: A visual representation of the vectors needed for the representation of V_{ij}, seen from three different reference systems: two from each of the pentons *i* and *j*, and the third being the standard computer three-dimensional axis (the Z-axis being featured in the diagram, marking it as the *C2* symmetry representation). Taken from [27].

On the other hand, p_0, p_1 and p_2 are constant values of energy that can be varied to obtain different results, and r_{en} and θ_{en} are the given distance and angle, respectively, which provide the structure the least potential energy. In this model, we strive to simulate the twelve pentons of the icosahedral capsid, so θ_{en} has a fixed value: the arctangent of the golden ratio (Φ).

$$\theta_{en} = arctan(\Phi) \simeq 58.283°;\ \Phi = (1 + \sqrt{5})/2 \simeq 1.618$$

At the end, the bonding energy stored in the system (V) is calculated by the summation of every individual term of V_{ij} in half, as every interaction is calculated twice:

$$V = \frac{1}{2}\sum_{i}\sum_{j\neq i} V_{ij}$$

This value, alongside the terms of its Jacobian matrix, will be extremely useful in the next stage of the study: the simulation of the pentons in an icosahedral structure. These data can be more thoroughly researched, as adding more terms of the multipolar expansion would prove to replicate with more accuracy the capsid's behaviour and characteristics.

It is worth noting that, in the case of the project of the La Laguna group [26], the model investigated the circumstances around a MMV capsid whose subunits were formed by trimers, and in that assumption, the first six terms of the multipolar expansion were needed to form the icosahedral shape (each of the 20 rigid units representing a face of the icosahedron).

The main objective of this project is to study the variations of the bond energy between the simulated pentons of the capsid (along its three rotational symmetries: *C5*, *C2* and *C3*) when subjected to an adiabatic force of indentation, and to compare the results obtained with the data gathered by the trimer coarse-grained model in the ULL group [26]. The simulation will attempt to represent a force measurement experiment with an atomic force microscope (AFM) [Figure 12], which consists in the usage of the tip of the microscope to generate a piezoelectric force over the desired material, which will return a force curve that can be analysed to understand better its characteristics.

Our goal is to study if different capsid models can achieve similar results, thus proving if both models can be used as a tool to research viruses that are difficult to be experimented in a laboratory. To achieve that, we created a code that determined the bond energy of a given set of 60 coordinates (corresponding the x, y, z, θ and φ of each of the 12 pentons) and the values of the parameters p_1, p_2, m, and m'; calculating the V_{ij} between every penton and the energy V of the system. Another function, the jacobian matrix of the function V (dV), is created as well.

Indentation simulations

Abstract. *Establecemos los parámetros de la función de energía y analizamos los valores iniciales de las coordenadas para introducirlos al proceso de la simulación, donde utilizaremos dos algoritmos de la librería de fuente abierta Scipy. Teniendo en cuenta todos los cálculos previos y tras un análisis de los elementos necesarios para su correcto funcionamiento, creamos una rutina de Python donde se simule unas condiciones de hendimiento para la cápside, donde obtenemos como respuesta sus energías, fuerza de hendimiento y constante elástica. Éstas son representadas gráficamente en función del hendimiento, y posteriormente se valora la efectividad de los parámetros elegidos inicialmente.*

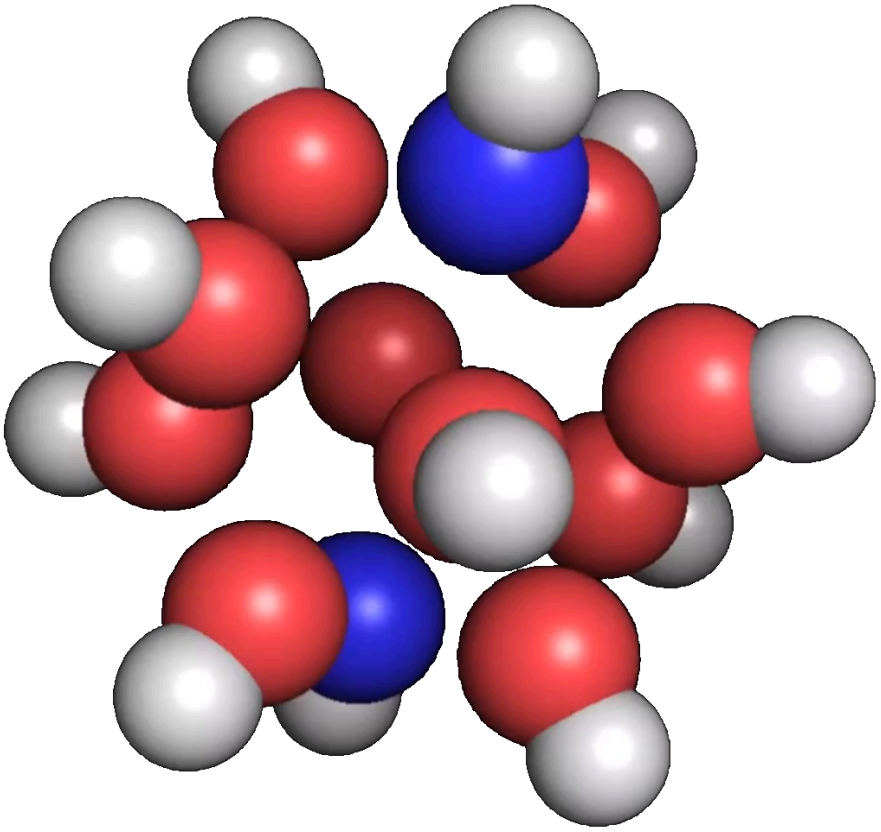
In this project, the distance of equilibrium between pentons $\boldsymbol{r_{en}}$ has been established to be 1, as it can therefore be used as a unit of distance. Moreover, the twelve pentons are situated as the vertices of an icosahedron. For that to be in accordance to the coordinates of the pentons (with the origin of the coordinate system in the centre of the capsid) it has to be taken into account that, for any given regular icosahedron with an edge length of $\boldsymbol{a}$, the distance between the centre of the icosahedron and any of its vertices equals $0.95105652 \cdot a$.

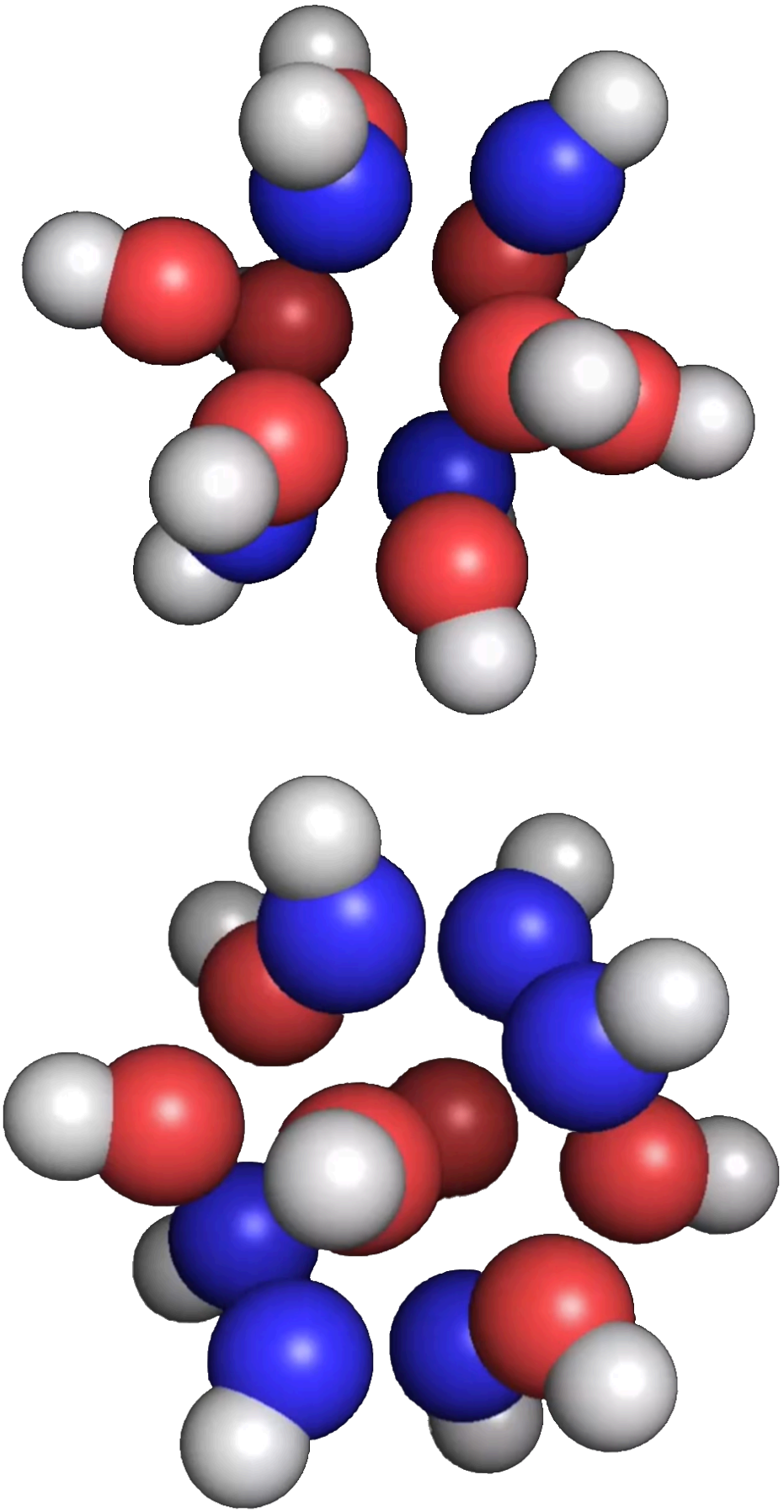
For scaling purposes and to indicate it as the basic energy unit of the system, the constant term p_0 will be set up as 1. The values of p_1, p_2, m, and m' will be fixed for every symmetry as it follows:

$$p_1 = 10;\ p_2 = 10;\ m = 12;\ m' = 1.5 \Rightarrow m \cdot m' = 18$$

These have been selected after numerous trials and errors with the code, to accommodate qualitatively the experimental results. Further into the study, the nature of the code when tweaking these constants will be observed, to further prove their convenience.

For starters, an array of 60 elements is created, where every five elements represent the coordinates of a penton, as stated before. Taking the precautions mentioned beforehand, this array has the coordinates of a regular icosahedron with two of its vertices fixed in the Z-axis. As the simulated force will be represented as a minute downwards displacement in the z-term of the coordinate of the penton or pentons with higher *z* value, this arrangement will represent that of the *C5* symmetry. A simple rotational function is all that is needed to generate the other two configurations needed in the project: the *C2* and *C3* symmetries.





Figures 9, 10 and 11: A representation of the initial states of the three rotational symmetries the simulations will experiment with: *C5*, *C2* and *C3*, respectively. The pentons representing the bottommost (which will remain immobile) and uppermost (which will suffer the indentation) of the capsid are depicted in blue, indicating clearly which figure corresponds to each rotational symmetry. A smaller white orb is attached to all pentons, representing the direction the penton is oriented at. As the simulation allows every penton to rotate, it is expected to change its position as the calculations take place. Created with PyMOL, an open source molecular visualisation system [28].

After that, it is important to establish the bounds of the coordinates. For each rotational symmetry (*C5*, *C2* and *C3*) two extremes will be chosen [Figures 9, 10 and 11]. Those pentons will then be constrained to their current coordinates, although changes in rotation are allowed. For the simulation to replicate an adiabatic force of indentation, one of the extremes will then be made to approach the other one by a minute distance, to then calculate the displacement of the other pentons as a result.

Is in this part of the project where the usage of the SciPy open-source library of scientific computing is key. Specifically, the function of **optimize.minimize**, as its objective is to find the minimum value of any scalar function with one or more variables [29]. For it to work properly, it need certain parameters to be inputted first:

- The **function** to be minimised. In this case, V.
- An **initial set of variables**. The coordinates of the twelve pentons.
- All the **parameters** the function needs: p_1, p_2, m, and m'.
- The **preferred method** to solve the minimisation. There are numerous to choose from, but as advised by the supervisor of this project, the Limited-memory BFGS with bound constraints (L-BFGS-B) will be the one used. It utilises both the bounds of the edge pentons' coordinates and the jacobian matrix dV, and thus generates a better result in significantly less time.

- Either the **jacobian matrix** of the function, or the preferred method to calculate it. To save time and reduce the computing load of the data, it is already calculated in the function *dV*.

- If the method requires them, the **bounds** (or lack thereof). The program needs it as a sequence of *n* elements, *n* being the number of variables of the function. Each element consists of a pair of numbers: (*p*, *q*). If no bounds are needed for a specific variable, both *p* and *q* have to be equal to *None*, and if there are, *p* represents the lowest possible value we want it to strive for, and *q* the highest one.

- If the preferred method allows it, it is possible to input some **options** to control to which extent the algorithm is allowed to proceed, as the maximum number of interactions or the minimum value the function has to achieve respecting the prior iteration of the minimiser to allow the algorithm to take the new value as the new minimum.

At all times, the function will be *V* with its jacobian *dV*, and the preferred method L-BFGS-B and its options will not be tweaked.

The **scipy.optimize.minimize** will then be used for its first calculation: based on the initial coordinates we had, and taking the necessary precautions to make sure the symmetry of every system remains the same (such as, creating functions to recentre and rotate the system if a displacement from the origin of the coordinate system has occurred), we obtain the energy configuration of the relaxed regular icosahedron for *C5*, *C2* and *C3*, whose energy configuration is slightly lower that the initial value. This is explained easily, as the distance between each penton has shifted slightly to adjust the effect created by the interaction of pentons beyond closest neighbours.

Now that an ideal configuration for every symmetry has been generated, the creation of a Python routine that simulates an indentation process can start. It is built in a simple **while** loop, and in each iteration, a downward displacement of a six thousandth of the distance between two opposing vertices of the relaxed icosahedron is forced on the uppermost penton or pentons (depending on the symmetry in question). Then, the scipy.optimize.minimize algorithm utilises this new set of coordinates, producing as a result another array of coordinates (respecting the bounds

fixing the position of the uppermost and bottommost pentons) with an energy bond V which is expected to be marginally higher than in the previous run of the algorithm, due to the new coordinates being nearly identical to the previous ones. The only instance when this will not turn out true is when the coordinates of some pentons abruptly shift their positions, finding a new minimum configuration outside their vicinity. This happens because the method we have chosen, L-BFGS-B, searches for the minimum energy configuration in its vicinity, only not doing so if it is not viable to find one.

As we have settled that this would be an unnatural behaviour for a virus in a practical experiment and only feasible in the realm of a computer simulation, it is safe to assume this force would result in the deformation or rupture of the capsid [30]. Therefore, the algorithm is made to break the loop at that point, having saved every value of the energy bond and its respective indentation for the next step of the project: The graphic representation and calculation of the energy bond V, the force of indentation (the first derivative of V) and the elastic constant of the capsid (the second derivative of V) over its deformation.

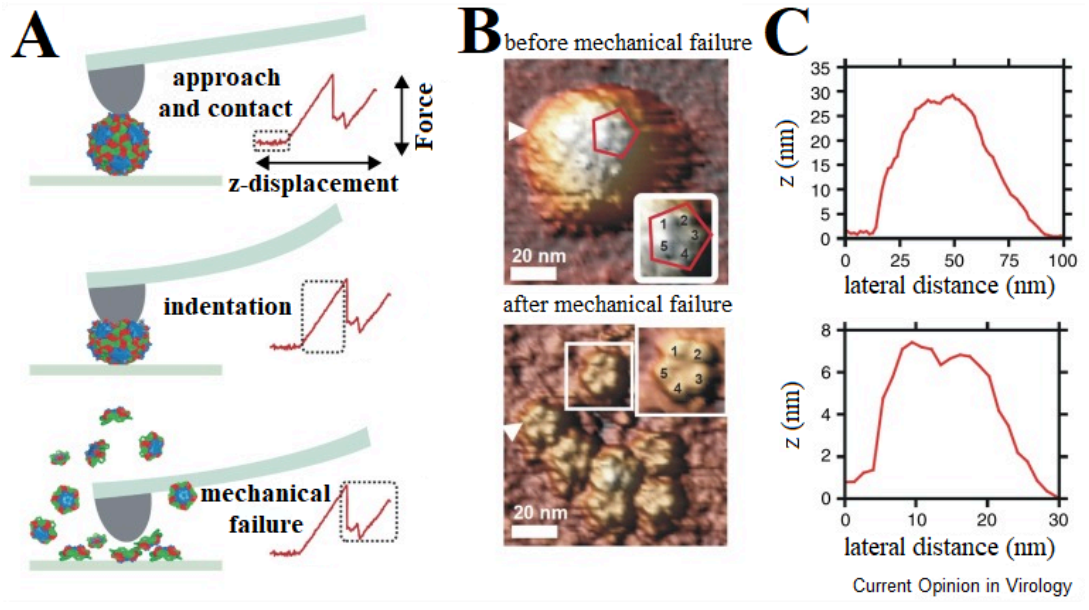
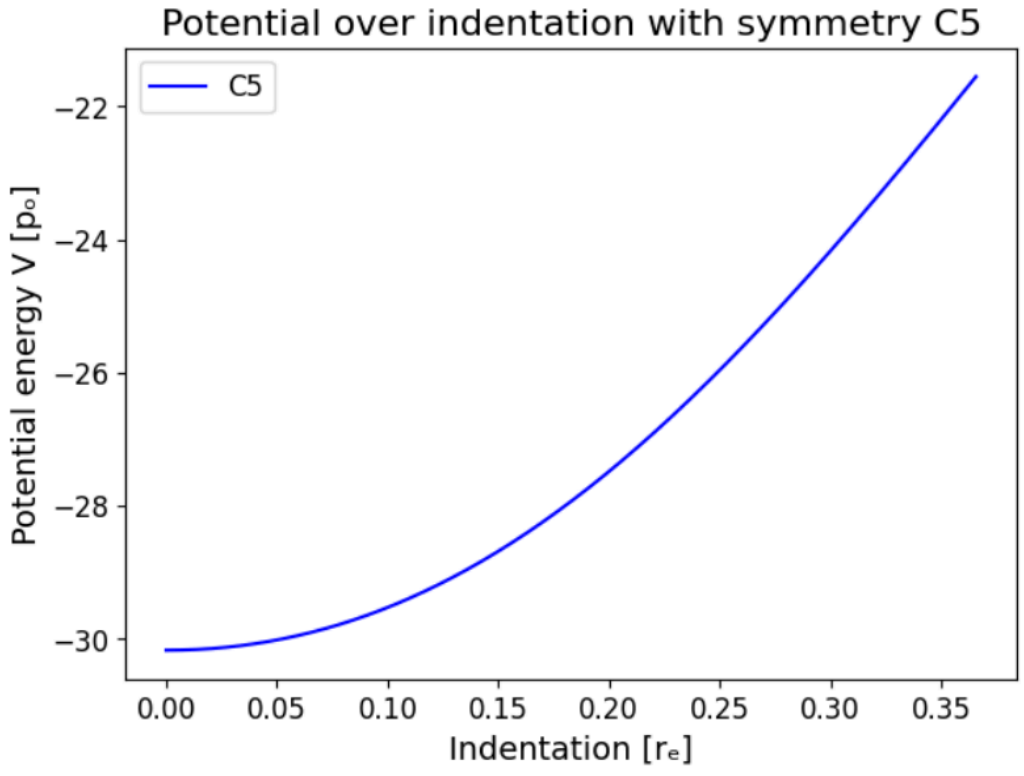
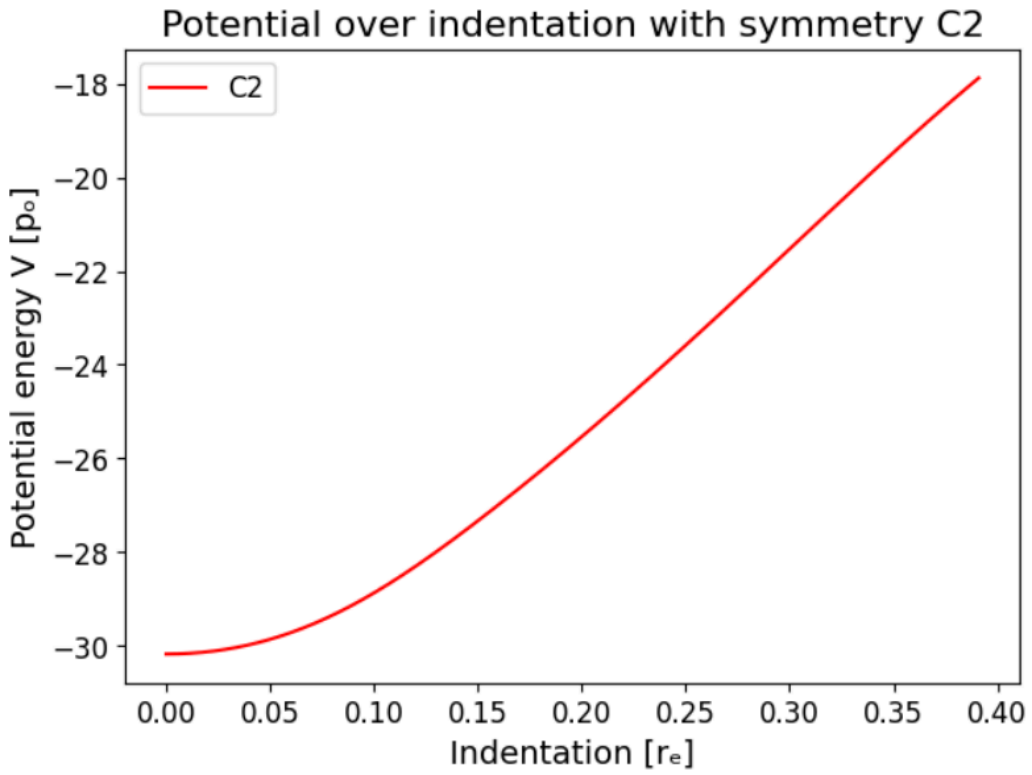


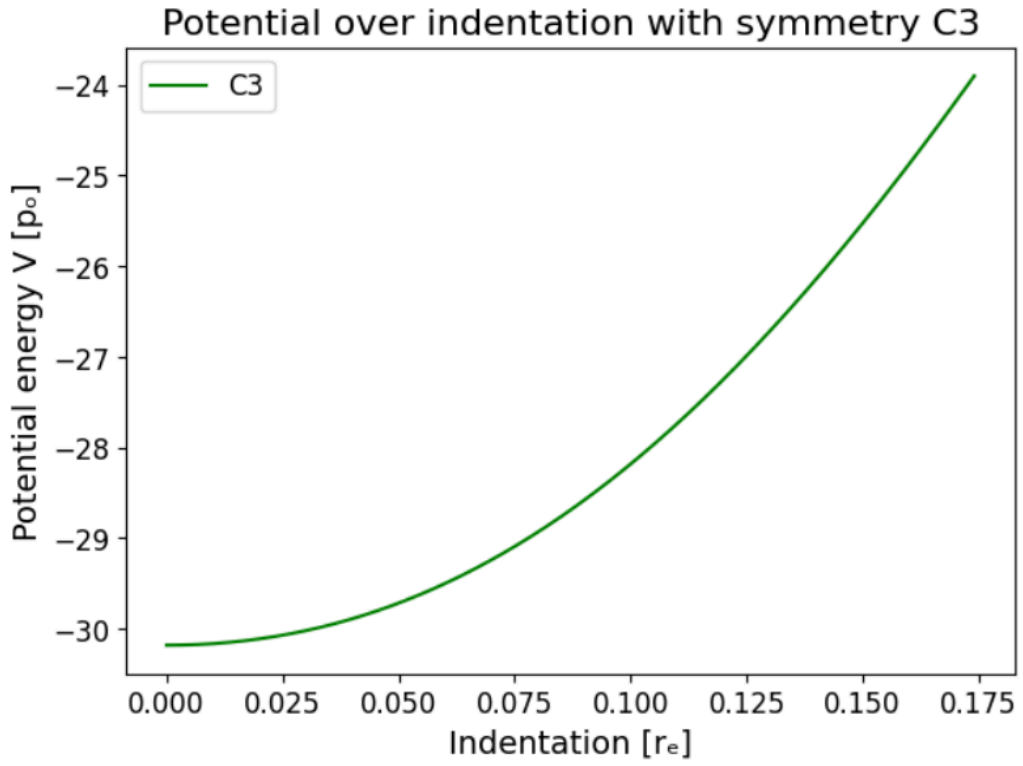
Figure 12: Results acquired from laboratory experiments of virus indentation. They are **A**, a diagram of how the indentation takes place, showing a representation of what happens to the capsid over the gathering of data of force over z-displacement; **B**, two images, the before and after of

a viral capsid of the triatoma virus which underwent the atomic force microscope (AFM) indentation process, where a penton subunit can be observed as part of the capsid and as a subunit, respectively; and **C**, two graphs representing the results of height of a raster scan made for the states of the images in **B**. Taken from [30].

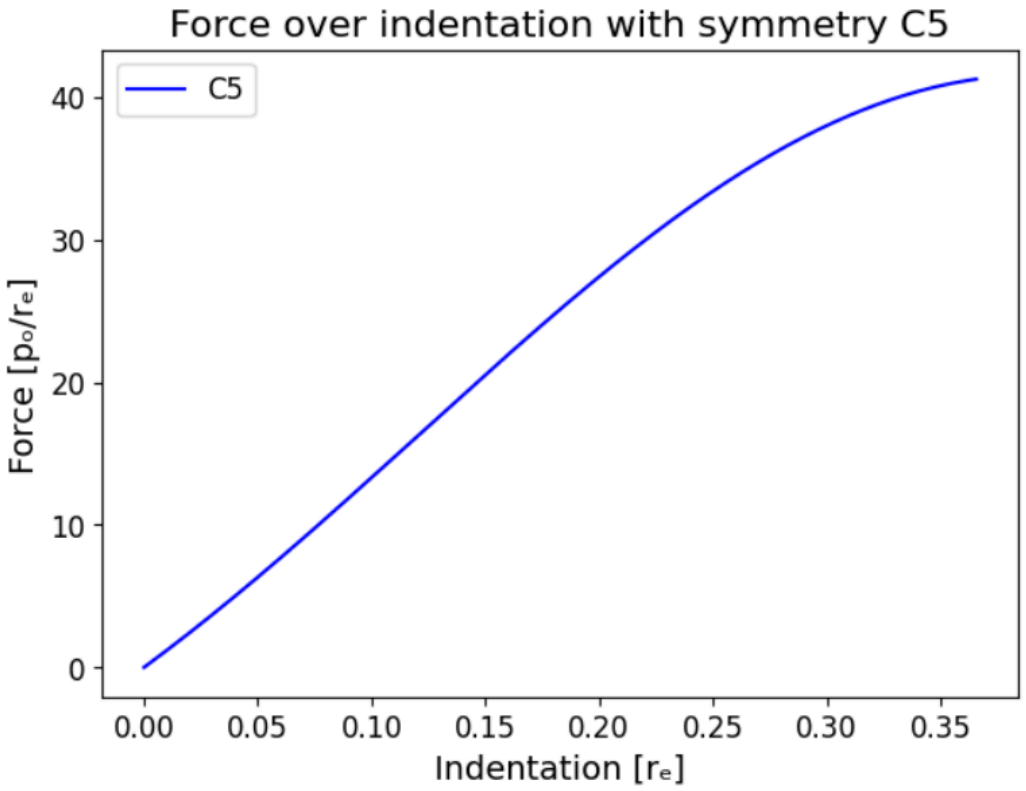
It is possible to do so thanks again to the Scipy open-source library of scientific computing. This time, we are going to make use of the algorithm known as **scipy.interpolate.InterpolatedUnivariateSpline** [29], whose goal is to create a spline that interpolates a set group of (x, y) data points. Doing so with our array of capsid displacement as x and their respective energy bond as y might seem unnecessary at first, as the data could be represented in a graph with not such help, but one of the submethods of the algorithm, **derivative**, has the ability to calculate (as his name states) the derivative of one coordinate imputed. Just a couple of **for** loops are required to obtain the first and second derivative of the spline previously mentioned, effectively creating the simulated curve of force applied to the capsid and elastic constant of the capsid (respectively). These results are given in the following figures:

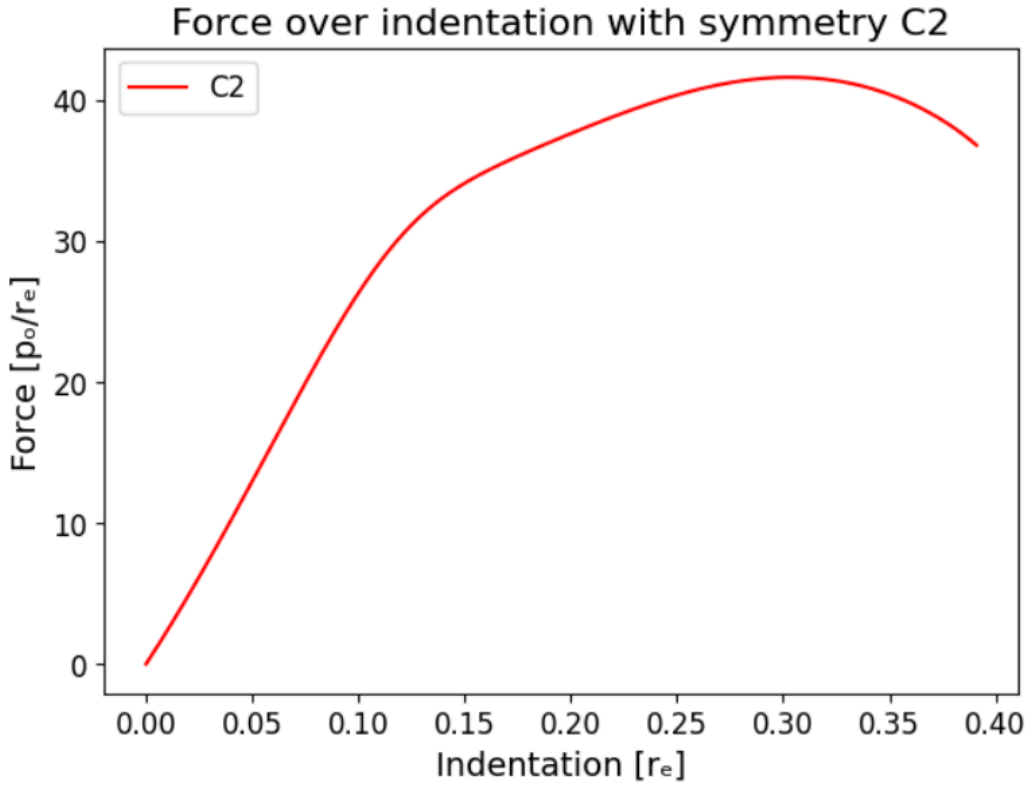


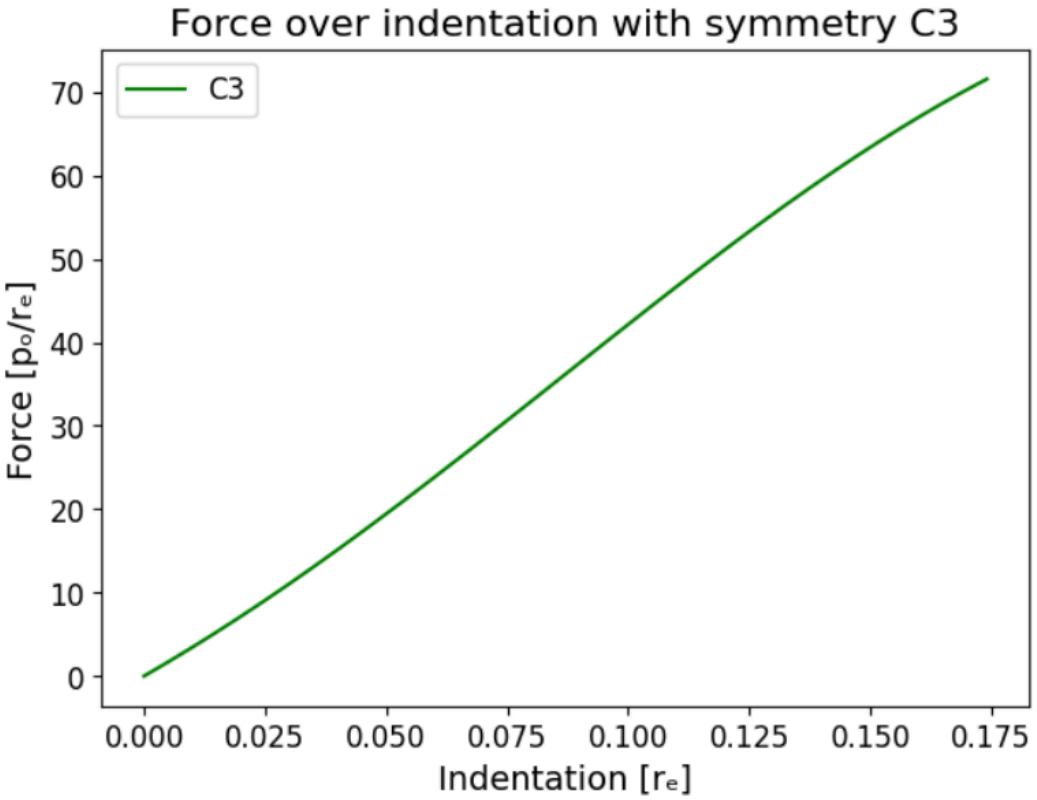




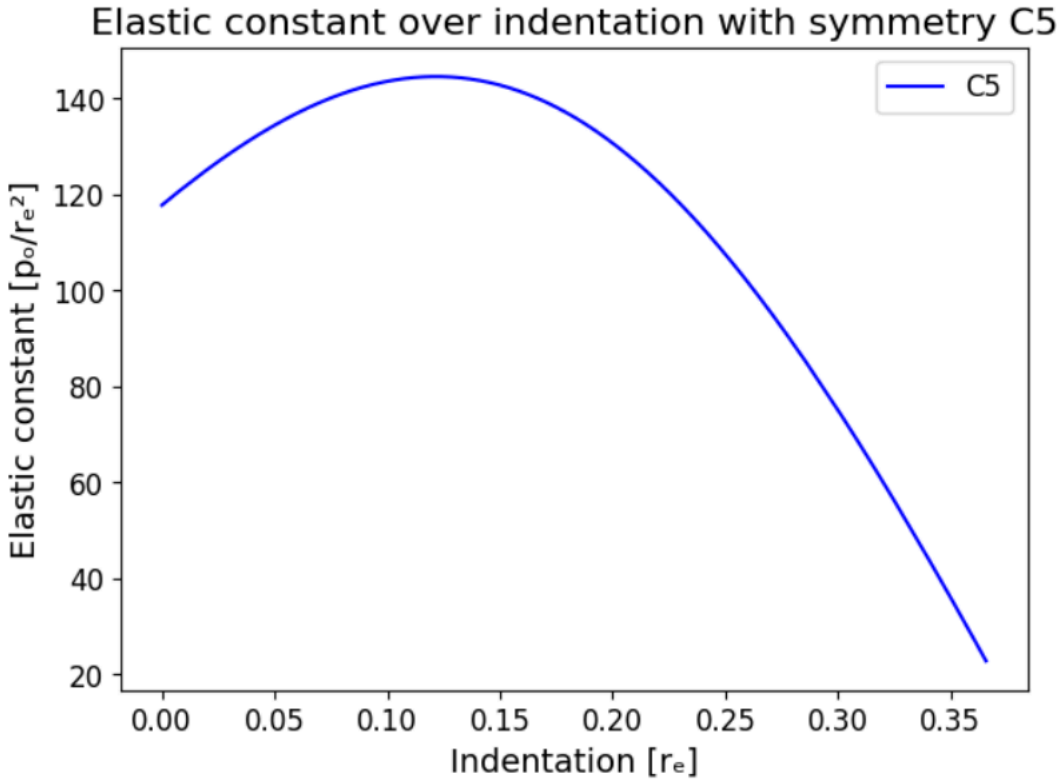
Figuress 13-15: Potential over indentation representations for symmetries *C5*, *C2* and *C3*, respectively. Their indentation values range from zero to their respective critical indentation, when the capsid ruptures.

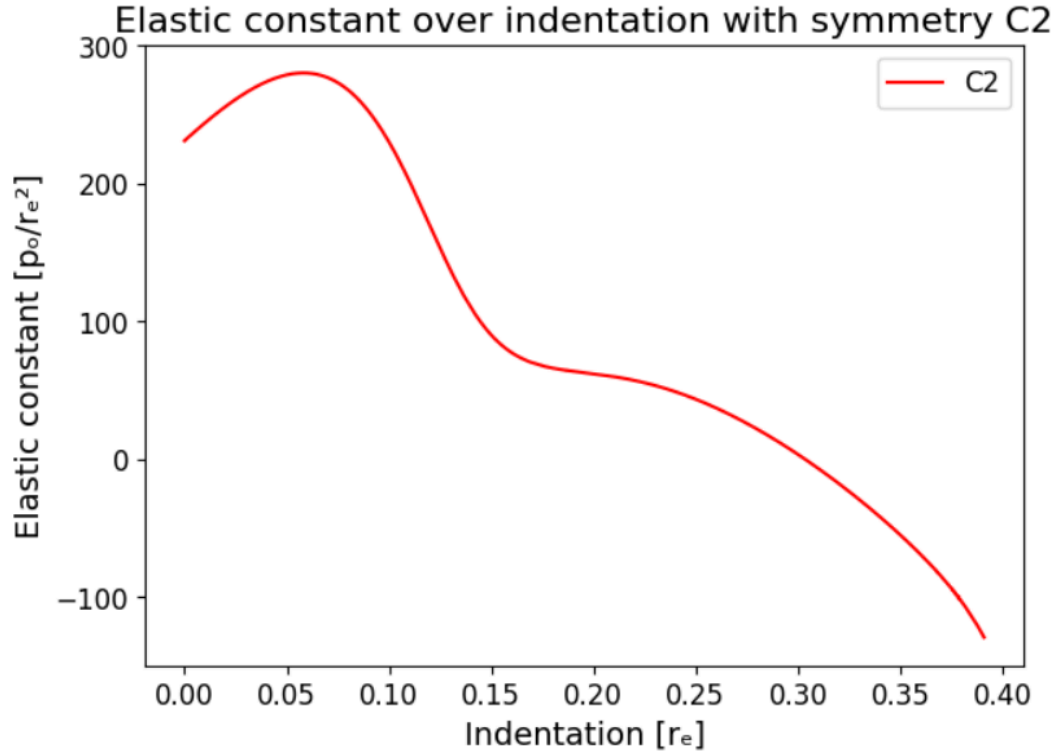

Force over indentation with symmetry C5
C5
Force [p₀/rₑ]
Indentation [rₑ]
0
10
20
30
40
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35

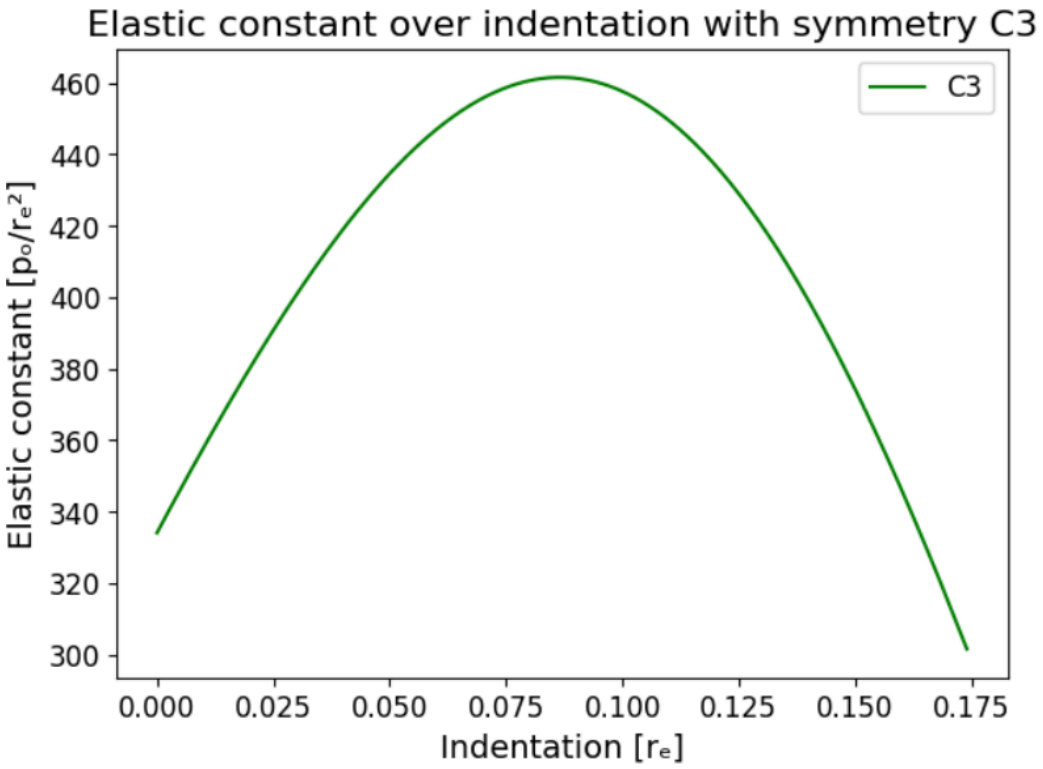

Force over indentation with symmetry C2
C2
Force [p₀/rₑ]
Indentation [rₑ]
0
10
20
30
40
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40



Figures 16-18: Force over indentation representations for symmetries *C5*, *C2* and *C3*, respectively. Their indentation values range from zero to their respective critical indentation, when the capsid ruptures.


Elastic constant over indentation with symmetry C5
C5
Elastic constant [p₀/rₑ²]
140
120
100
80
60
40
20
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
Indentation [rₑ]


Elastic constant over indentation with symmetry C2
C2
Elastic constant [p₀/rₑ²]
300
200
100
0
−100
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
Indentation [rₑ]



Figures 19-21: Elastic constant over indentation representations for symmetries *C5*, *C2* and *C3*, respectively. Their indentation values range from zero to their respective critical indentation, when the capsid ruptures.

After these results, it is right to ponder how the parameters p_1, p_2, m and m' affect the final representation. For that to be clarified, a total of four extra simulations were made, following every step as it has been described, except for a considerable change in one of the parameters. We will now compare the critical indentation (distance of deformation which causes the simulated rupture of the capsid) and maximum elastic constant (k) value of the original simulation and the extra four ones, for the three symmetries of the capsid: *C5*, *C2* and *C3*. The results are given in the following tables:

(p_1, p_2, m, m')	Maximum k for *C5*	Critical indentation *C5*
(10, 10, 12, 1.5)	144.5679792 [p_0/r_e^2]	0.3661166832 [r_e]
(2, 10, 12, 1.5)	102.033926 [p_0/r_e^2]	0.3593131211 [r_e]
(10, 2, 12, 1.5)	109.8612288 [p_0/r_e^2]	0.3682359236 [r_e]
(10, 10, 6, 3)	147.9958576 [p_0/r_e^2]	0.3565750764 [r_e]
(10, 10, 12, 0.75)	130.608214 [p_0/r_e^2]	0.4077445134 [r_e]

Table 1: Maximum k and critical indentation for the system in *C5* symmetry.

(p_1, p_2, m, m')	Maximum k for *C2*	Critical indentation *C2*
(10, 10, 12, 1.5)	280.3344078 [p_0/r_e^2]	0.3913766825 [r_e]
(2, 10, 12, 1.5)	212.1968476 [p_0/r_e^2]	0.5364247876 [r_e]
(10, 2, 12, 1.5)	226.2561997 [p_0/r_e^2]	0.5161590595 [r_e]
(10, 10, 6, 3)	228.4942697 [p_0/r_e^2]	0.2441625499 [r_e]
(10, 10, 12, 0.75)	262.7621211 [p_0/r_e^2]	0.343951764 [r_e]

Table 2: Maximum k and critical indentation for the system in *C2* symmetry.

(p_1, p_2, m, m')	Maximum k for *C3*	Critical indentation *C3*
(10, 10, 12, 1.5)	461.5460866 [p_0/r_e^2]	0.1743590128 [r_e]
(2, 10, 12, 1.5)	331.456205 [p_0/r_e^2]	0.164897671 [r_e]
(10, 2, 12, 1.5)	346.7161069 [p_0/r_e^2]	0.1730041633 [r_e]
(10, 10, 6, 3)	381.125277 [p_0/r_e^2]	0.2676120914 [r_e]
(10, 10, 12, 0.75)	416.050141 [p_0/r_e^2]	0.1870363764 [r_e]

Table 3: Maximum k and critical indentation for the system in *C3* symmetry.

After a quick glance at the tables it can be noted that, while not always having the uppermost values, the parameters p_1, p_2, m and m' prove to have the most consistent results of the pack.

In all of the representations, a coefficient r_d has been utilised to adjust all values to those of a similar experiment [27], where the same equations and simulations were made to a virus with the same characteristics that the MMV, being the only difference with this project that, instead of pentons, they acknowledged that the experimental results of the MMV indentation proves that the subunits of its capsid are formed by trimers, and thus their calculations revolve around that assertion. The coefficient r_d is the result of dividing the distance between two nearest penton neighbours and the distance between two nearest trimer neighbours.

Results and conclusion

Abstract. *Los gráficos de energía sobre hendimiento muestran diferencias significativas en sus valores críticos, teniendo la simetría C3 el menor hendimiento crítico. Esto indica una mayor resistencia a la deformación o ruptura en comparación con C2 y C5, lo que queda confirmado por los gráficos de la constante elástica. Aunque los parámetros iniciales afectan los valores críticos, las constantes elásticas permanecen estables en todas las simetrías. Los resultados se comparan con un estudio que utiliza trímeros, mostrando tendencias similares que demuestran que ambos modelos son válidos para cápsides geométricamente similares. Concluimos que los modelos de grano grueso ofrecen información general aún necesitando los parámetros adecuados para mejores resultados. La experimentación práctica aclara la estructura de la cápside de MMV, resaltando la importancia de ajustar los parámetros para simulaciones precisas.*

The potential energy over indentation charts [Figures 13-15] show significant differences between them, being one of them their disparity in critical indentation values. They demonstrate that *C3* has, in theory, the lowest of the three (around 16% of capsid indentation), allowing less than half of the deformation of the capsid that the *C2* or *C5* symmetry permit (more than a 35% of capsid indentation) before being irreversibly deformed or ruptured. This translates accordingly to the force over indentation charts [Figures 16-18], where it is shown how the effort needed to damage a capsid in *C3* is significantly higher than its counterparts. It can be therefore argued that *C3* has a greater stiffness, which the charted elastic constant over indentation shows to be the case [Figures 19-21], at least on this experience.

This information raises a question: “Does this apply to every set of initial parameters?” It can be answered confidently that although some initial values of p_1, p_2, m, and m' allow significant variances in the critical indentation values, the maximum elasticity constant demonstrates to have a steady value for every

symmetry configuration: *C3* is consistently the hardest to deform, and *C5*, the easiest, according to the pentons model.

All along the experiment, another project with similar objectives has been referenced [27], as its main difference regarding the study of simulated capsid behaviour regarding an adiabatic indentation from different symmetrical perspectives lies on the choice of subunit for the MMV capsid structure: trimers. Its results can therefore be compared to those emerged from these calculations, and used to demonstrate if a different modelling of the structure translates into similar data.

Remarkably, the average of both the elastic constant and critical indentation are resemblant. The biggest variance between them lies in the individual values of critical indentation and elastic constant obtained in each symmetry configuration (*C5*, *C2* and *C3*), as in the trimer model the *C3* symmetry has instead the highest critical indentation of the three (24% of capsid indentation). This can be explained by the different choice of rigid units, as it is logical that the assumption of trimer rigid units will conclude in penton units being deformed, and vice versa, which translates to distinct breaking points for each of the symmetries depending on the capsid subunit model.

Observing their average behaviour we can establish that both capsid models reproduced similar elastic properties and characteristics in general, meaning that two separate viruses with geometrically similar capsids share those too, even if the organisation of its subunits are different; but each model has different specific symmetry properties that clearly differentiates one model from the other. This information can be very helpful to discern the kind of subunit the capsid is formed by. This gives us the conclusion that both methods, albeit different depending on the initial parameters, portray the general behaviour of the capsid quite notably. The coarse-grained model works the same way grosso modo, which demonstrates that these simulations based on the potential energy equation here presented are only capable of distinguishing between capsids with subunits of different nature when studied not their average elastic properties, but each symmetry at a time.

For a more minute and exact understanding of the kind of subunits a virus uses to build its capsid, a better precision method is needed. For example, instead of a

calculation of parameters p_1, p_2, m, and m' made by trial and error, as this project has done, an exhaustive search of better values can be made with the proper subroutine, as in my tutor group's paper [20], where other ways to further expand this investigation relied on better methods of global optimization (like the Harmony Search [31]) to adjust the energy function automatically depending on the experimental data we want to study. While this method doesn't change the obtained results, the parameters now are better adjusted to them, easing the simulations to better recreate the capsid behaviour.

In the end, it was the practical experimentation of virus indentation with AFMs that clarified once and for all which subunit uses the MMV to create its capsid (trimers), albeit not an indentation experiment, as the incertitude derived by the error margin of the measurements makes it so the estimates of force of indentation in every symmetry are indistinguishable from each other. Only in virtual simulations, like those found in this project or in those where this hypothesis has been based on [26, 27], each symmetry can be observed separately. Not only that, but also their elastic constant and critical indentation values are unique to them, not sharing its characteristics between models, as it happened in this experience. With a precise enough laboratory experiment and this theoretical data, it could be established with certainty which type of subunits form any viral capsid that satisfies the $T = 1$ Caspar and Klug theory.

Data recollection in the laboratory proves to work wonders to gather the necessary information to create accurate simulations, even when using models such as the Coarse-grained model, which simplifies enormously the intricate structure of even the most simple viral capsid.

References

[1]. Lawrence, C M; Menon, S; Eilers, B J; Bothner, B; Khayat, R; Douglas, T; Young, M J (2009). "*Structural and functional studies of archaeal viruses*". The Journal of Biological Chemistry, **284**(19): 12599–12603.

[2]. "*Virus Taxonomy: 2023 Release*". International Committee on Taxonomy of Viruses. https://ictv.global/taxonomy

[3]. Buist, John B (1887). "*Vaccinia and variola, a study of their life history*". J. & A. Churchill.

[4]. Zaitlin, M (1998). "*The Discovery of the Causal Agent of the Tobacco Mosaic Disease*". Discoveries in Plant Biology, pág 105–110.

[5]. Dhama, K; Khan, S; Tiwari, R; Sircar, S; Bhat, S; Malik, Y S; Singh, K P; Chaicumpa, W; Bonilla-Aldana, D K; Rodriguez-Morales, A J (2020). "*Coronavirus Disease 2019-COVID-19*". Clinical Microbiology Reviews, **33**(4): e00028-20.

[6]. Brito, Anderson. Wikimedia Commons.

[7]. Hozáková, Lucie; Vokatá, Barbora; Ruml, Tomáš; Ulbrich, Pavel (2022). "*Targeting the Virus Capsid as a Tool to Fight RNA Viruses*". Viruses, **14**(2): 174.

[8]. Flurry, R L (1980). "*Symmetry Groups : Theory and Chemical Applications*". Prentice-Hall.

[9]. Baltimore, D (1971). "*Expression of animal virus genomes*". Bacteriol. Rev, **35**(3): 235–241.

[10]. Gorbalenya, A E; Krupovic, M; Mushegian, A; Kropinski, A M; Siddell, S G; Varsani, A; Adams, M J; Davison, A J; Dutilh, B E; Harrach, B; Harrison, R L; Junglen, S; King, A M; Knowles, N J; Lefkowitz, E J; Nibert, M L; Rubino, L; Sabanadzovic, S; Sanfaçon, H; Simmonds, P; Walker, P J; Zerbini, F M; Kuhn,

J H (2020). "*The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks*". Nature Microbiology, **5**(5): 668-674.

[11]. "*Eukaryotic microorganisms and viruses*" (WS 2010/2011). Institute for Chemistry and Biology of the Marine Environment, pág 9.

[12]. Buzón, Pedro; Maity, Sourav; Roos, Wouter H (2019). "*Physical virology: From virus self-assembly to particle mechanics*". WIREs Nanomedicine and Nanobiotechnology, **12**(4): e1613.

[13]. Yin J, Redovich J (2018). "*Kinetic Modeling of Virus Growth in Cells*". Microbiology and Molecular Biology Reviews, **82**(2): e00066-17.

[14]. Brownstein, David G; Smith, Abigail L; Johnson, Elizabeth A; Pintel, David J; Naeger, Lisa Kay; Tattersall, Peter (1992). "*The pathogenesis of infection with minute virus of mice depends on expression of the small nonstructural protein NS2 and on the genotype of the allotropic determinants VP1 and VP2*". Journal of Virology, **66**(5): 3118-24.

[15]. Pritchett-Corning, Kathleen R; Cosentino, Janice; Clifford, Charles B (2009). "*Contemporary prevalence of infectious agents in laboratory mice and rats*" Lab Anim, **43**(2): 165-73.

[16]. Baker, David G (1998). "*Natural pathogens of laboratory mice, rats, and rabbits and their effects on research*". Clinical Microbiology Reviews, **11**(2): 231-66.

[17]. Weaver, Justin; Husson, Scott M; Murphy, Louise; Wickramasinghe, S. Ranil (2013). "*Anion exchange membrane adsorbers for flow-through polishing steps: Part I. clearance of minute virus of mice*". Biotechnology and Bioengineering, **110**(2): 491-9.

[18]. Mice Minute Virus (Mvm), Strain I (VIPERdb)

[19]. Verdaguer, N; Garriga, D; Fita, I (2013). "*X-ray crystallography of viruses*". Subcell Biochem, **68**: 117-44.

[20]. Martín Bravo, Manuel; Gomez Llorente, Jose M; Hernández-Rojas, Javier (2024). "*Virtual indentation of the empty capsid of the minute virus of mice using a minimal coarse-grained model*". Physical review E, **109**: 024402.

[21]. Matsuura, Kazunori (2014). "*Rational Design of Self-assembled Proteins and Peptides for Nano- and Micro-sized Architectures*". ChemInform Abstract, **45**(6).

[22]. Caspar, D L; Klug, A (1962). "*Physical principles in the construction of regular viruses*". Cold Spring Harbor Symposia on Quantitative Biology, **27**: 1-24.

[23]. "*Viruses*" (2015). Cronodon.

[24]. "*Virus symmetry and T number*" (2011). ViralZone.

[25]. Kmiecik, S; Gront, D; Kolinski, M; Wieteska, L; Dawid, A E; Kolinski, A (2016). "*Coarse-Grained Protein Models and Their Applications*". Chemical Reviews, **116**(14): 7898-7936.

[26]. Martín Bravo, Manuel; Gomez Llorente, Jose M; Hernández-Rojas, Javier; Wales, David J (2014). "*Minimal design principles for icosahedral virus capsids*". ACS Nano, **15**(9): 14873-14884.

[27]. Martín Bravo, Manuel (2022). "*Simulación numérica de cápsides víricas con modelos de grano grueso*". [Doctoral thesis, Universidad de La Laguna].

[28]. Schrödinger, L; DeLano, W (2020). PyMOL.

[29]. Virtanen, Pauli; Gommers, Ralf; Oliphant, Travis E; Haberland, Matt; Reddy, Tyler; Cournapeau, David; Burovski, Evgeni; Peterson, Pearu; Weckesser, Warren; Bright, Jonathan; van der Walt, Stéfan J; Brett, Matthew; Wilson, Joshua; Millman, K Jarrod; Mayorov, Nikolay; Nelson, Andrew R J; Jones, Eric; Kern, Robert; Larson, Eric; Carey, C J; Polat, İlhan; Feng, Yu; Moore, Eric W; VanderPlas, Jake; Laxalde, Denis; Perktold, Josef; Cimrman, Robert; Henriksen, Ian; Quintero, E A; Harris, Charles R; Archibald, Anne M; Ribeiro, Antônio H; Pedregosa, Fabian; van Mulbregt, Paul; and SciPy 1.0 Contributors

(2020). "*SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*". Nature Methods, **17**(3): 261-272.

[30]. Marchetti, M; Lodewijk Wuite, Gijs Jan; Roos, Wouter H (2016). "*Atomic force microscopy observation and characterization of single virions and virus-like particles by nano-indentation*". Current Opinion in Virology, **18**: 82-8.

[31]. Geem, Z W; Kim, J H; Loganathan, G (2001). "*A new Heuristic optimization algorithm: Harmony Search*". SIMULATION, **76**(2): 60-68.