eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

STRUCTURAL STUDIES OF FILAMENTOUS PLANT VIRUSES BY CRYO-ELECTRON MICROSCOPY AND CRYO-ELECTRON TOMOGRAPHY

REBECA CUESTA POU

DOCTORAL THESIS

JUNE 2021

STRUCTURAL STUDIES OF FILAMENTOUS PLANT VIRUSES BY CRYO-ELECTRON MICROSCOPY AND CRYO-ELECTRON TOMOGRAPHY

Memory presented by

Rebeca Cuesta Pou

Work supervised by:

Dr. Mikel Valle

Acknowledgements

I would like to thank the “Ministerio de Economía y Competitividad” for the funding to conduct my doctoral thesis. Thanks to Mikel Valle, for taking a chance on me and accepting me into his group. Thanks to my friends and co-workers here at CIC-bioGUNE. Thank you all for these amazing four years.

GENERAL INDEX

FIGURES INDEX

ABBREVIATIONS

1D = 1 dimension

2D = 2 dimensions

3D = 3 dimensions

aa = amino acid

AFM = Atomic force microscopy

AltMV = Alternanthera mosaic virus

ATP = adenosine triphosphate

BaMV = Bamboo mosaic virus

BYV = Beet yellow virus

CaMV = Cauliflower mosaic virus

CCD = Charge coupled device

CMV = Cucumber mosaic virus

CP = Coat protein

CryoEM = Cryo-electron microscopy

CryoET = Cryo-electron tomography

CSP = Chemical shift perturbation

CTF = Contrast transfer function

CTV = Citrus tristeza virus

DDC = Direct detection camera

DMSO = Dimethyl sulfoxide

DNA = Deoxyribonucleic acid

EM = Electron microscope

EMDB = Electron microscopy data bank

ER = Endoplasmic reticulum

FAO = Food and Agriculture Organization of the United Nations

FDMV = Foot-and-mouth disease virus

FT = Fourier transform

GFP = Green fluorescent protein

GUI = Guided user interface

HSQC = Heteronuclear single quantum coherence

ICTV = International Committee for the Taxonomy of Viruses

IPTG = isopropyl β-d-1-thiogalactopyranoside

NMR = Nuclear magnetic resonance

OD = optical density

ON = Over night

ORF = Open reading frame

PapMV = Papaya mosaic virus

PD = Plasmodesmata

PDB = Protein data bank

PepMV = Pepino mosaic virus

PPV = Plum pox virus

PVA = Potato virus A

PVX = Potato virus X

PVY = Potato virus Y

RNA = Ribonucleic acid

RT= Room temperature

SMV = Soybean mosaic virus

SNR = Signal-to-noise ratio

SPA = Single particle analysis

TEM = Transmission electron microscope

TMV = Tobacco mosaic virus

TuMV = Turnip mosaic virus

TYLCV = Tomato yellow leaf curl virus

VLP = Viral-like particles

VRC = Viral replication complex

WMV = Watermelon mosaic virus

1 letter code **Amino acids**	**3 letter code**	**Amino acid**
A	**Ala**	Alanine
R	**Arg**	Arginine
N	**Asn**	Asparagine
D	**Asp**	Aspartic acid
C	**Cys**	Cysteine
Q	**Gln**	Glutamine
E	**Glu**	Glutamic acid
G	**Gly**	Glycine
H	**His**	Histidine
I	**Ile**	Isoleucine
L	**Leu**	Leucine
K	**Lys**	Lysine
M	**Met**	Methionine
F	**Phe**	Phenylalanine
P	**Pro**	Proline
S	**Ser**	Serine
T	**Thr**	Threonine
W	**Trp**	Tryptophan
Y	**Tyr**	Tyrosine
V	**Val**	Valine

Chapter 1:

INTRODUCTION

1.1 GENERAL INTRODUCTION TO VIRUSES

Viruses are the most diverse and abundant microorganisms on earth. There are more than 10^{31} viral particles on our planet [1] infecting the three domains of life, *Bacteria, Archaea and Eukarya*, and covering even the most extreme ecosystems on the planet. They were discovered in the late XIX century by Dimitri Ivanovsky and Martinus Beijerinck, two scientist whose independent research on an "entity" that was causing a disease in tobacco plants established the basis for virus research, virology and the discovery of the first named virus, Tobacco mosaic virus (TMV) [2].

The Nobel Prize for Medicine and Physiology Peter Medwar described viruses as "a piece of nucleic acid surrounded by bad news"[3] meaning that viruses are composed basically of a nucleic acid that can be ribonucleic acid (RNA) or deoxyribonucleic acid (DNA) surrounded and protected by a proteinaceous capsid constructed by capsid proteins (CP). Viruses hijack the host's cellular machinery to make copies of themselves. Depending on the nature of the host and the virion, there are different methods for entering the cell. In animal viruses the virion must recognize receptors on the surface of the host cell which initiate viral entry by two different processes: receptor-mediated endocytosis and receptor-mediated activation of signalling pathways, both leading to virus internalisation via membrane fusion [4]. The rest of eukaryots (plants, algae and fungi) are covered by a cell wall, same as bacteria and archaea. Passing through the cell wall comprises a more complicated task and different viruses have their own mechanisms. Plant viruses can be transmitted by seeds or enter directly into the cytosol via insect or invertebrate vectors. Also, the cell wall can be broken mechanically and the plasma membrane perforated [5]. Bacteriophages inject the genome into the cytoplasm of the host cell, leaving the empty viral shell outside the host [6]. Once the viral particles are inside the host cell, the genetic material must be uncoated from the coat protein for the initiation of translation and transcription. Viruses take advantage of the replication and translation machineries and the newly synthesized genome and proteins assemble again into new viral particles. These particles can exit the cell by rupturing their membrane (lytic cycle) or forcing out through the cell membrane (budding) [7]. Some viruses, as herpesvirus and retroviruses, can undergo a phase of latency in the host cell, remaining silent inside the cytoplasm or integrating their genetic material into the host cell chromosomes [8].

The origin of viruses is not known and three different hypotheses have been considered, the “primordial virus world”, the “reductive virus origin” and the “escaped genes”. The first one says that viruses descend directly from the first forms of life. The second one stands for the opposite postulating that viruses are degenerated ancestral cells that developed to be intracellular parasites. In the last hypothesis, viruses may have been evolved from different organisms becoming parasites of those same organisms [9]. Interaction between virus and host’s genome can lead to functional changes that may be beneficial to the infected organisms. Viruses have also contributed to the actual genetic diversity due to their role in horizontal gene transfer between organisms [10].

1.1.1 Viral morphology

Viruses are highly structured microorganisms. They are composed of nucleic acid and capsid proteins (also referred as “coat proteins”) that protect the genetic material. Some viruses possess more than one type of capsid protein but others, as helical viruses, contain many copies of just one capsid protein. In addition, some viruses present a lipid membrane surrounding the capsid that comes from the host cell membranes. Those viruses are called enveloped. Viruses that aren’t enveloped are called naked viruses. Some viruses can come in two shapes defined by their highly ordered symmetry: helical and icosahedral. Helical viruses are composed of multiple copies of a single coat protein that protects the genome (Fig 1.1, A). Helical viruses can be enveloped or not. Enveloped helical viruses only infect animals, as Ebola or rabies virus, and helical viruses that infect plants aren’t enveloped. Enveloped viruses seem to infect predominantly organisms that do not possess a cell wall, while not enveloped viruses can infect hosts with or without it, so the absence of the envelope of plant viruses may be an evolutionary adaptation to cell walls, since viruses cannot enter the cells by endocytosis or exit plant cells by budding [11]. Icosahedral morphology is present both in plants and animals and their virions can be enveloped or not (Fig 1.1 B, C). Icosahedral morphology acts as a protein shell and the genetic material lays inside the protective shell [12]. This is a restriction for the viral genome, since its size is limited, and the nucleic acid can't be very large in order to fit inside. However, helical viruses can extend as long as the genetic material is.

Apart from these two main morphologies, there are also complex viral structures. The most known is the one composed by head and tail, typical of bacteriophages and archaeal viruses. This structure shows an icosahedral head attached to a helical tail (Fig 1.1 D). There are other complex morphologies as the ones found in Poxviruses that show large oval or brick-shaped particles (Fig 1.1, E) [12].

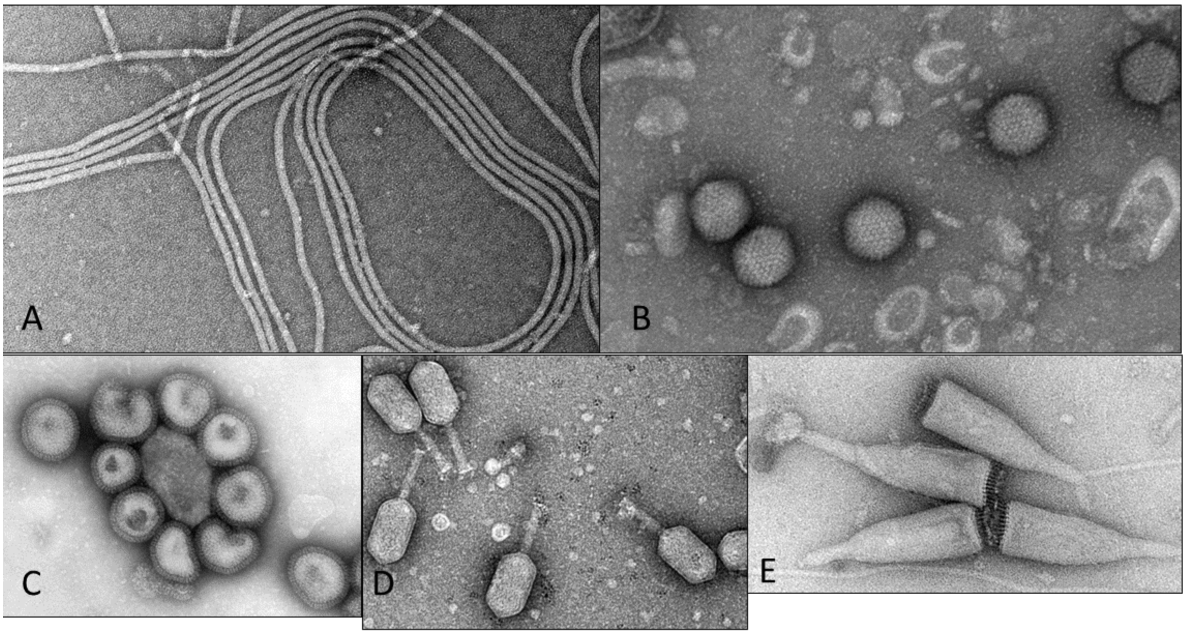


Fig 1.1 Electron microscope images of viruses with different morphologies. A) Helical naked virus, watermelon mosaic virus **B)** Small naked icosahedral virus, poliovirus. Image taken from Goldsmith and Miller [13] **C)** Enveloped virus, influenza B virus. Image taken from Goldsmith and Miller [13] **D)** Phage φpp2 particles. Image taken from Lin and Lin [14] **E)** Pleomorphic bottle-shaped virus, poxvirus ABV particles. Image taken from Häring et al [15].

1.1.2 Viral classification

Classification methods are important for scientists in order to understand and categorize the newly discovered species of biological entities. Viruses are classified following different criteria[16]. Thanks to the technological advances that are being made daily, more viral characteristics are being discovered, such as their genomic or structural relationships. For that reason, some classification methods are continuously under revision. There are two main classification methods: Baltimore classification and the taxonomical classification proposed by the "International Committee for the taxonomy of viruses" (ICTV).

Baltimore classification was proposed by the biologist David Baltimore and it divides viruses on seven groups depending on the nature of their genetic material; double-stranded DNA (dsDNA), single-stranded DNA (ssDNA), double-stranded RNA (dsRNA), plus sense single-stranded RNA (+(ss)RNA), negative sense single-stranded RNA (-(ss)RNA), ssRNA with a DNA intermediate and dsDNA with an RNA intermediate (Fig 1.X) [17]. The last two groups utilize the enzyme reverse transcriptase in their multiplication cycle [1].

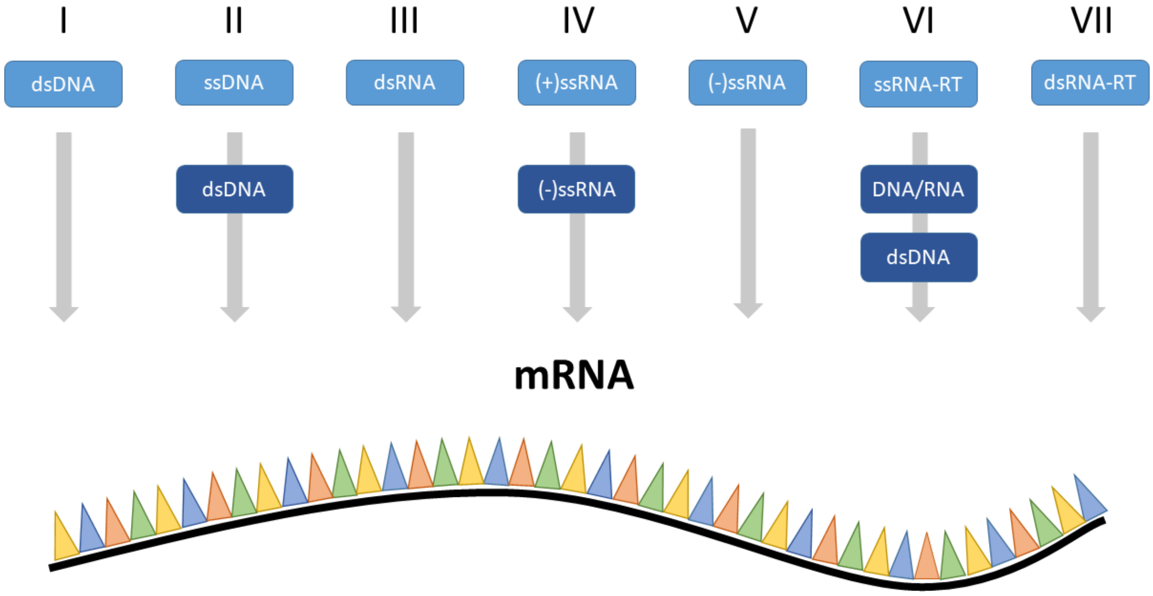


Fig 1.2 Baltimore classification scheme. The image shows the seven groups of the Baltimore classification, based on the nature of the genetic material of the virion and how it is transcribed inside the invaded cell.

ICTV classification organizes viruses following taxonomic criteria with the levels of order, family, genus and species. All cellular organisms share a last common ancestor, and that means that it's genetically possible to trace them back to their roots and "see" their universally conserved genes, allowing for their taxonomic classification. On the contrary, viruses do not have any set of universally conserved genes what makes it impossible to construct a phylogeny based on the evolution, so the taxonomy applied is not based on a unique and specific criterion. On one hand, some orders are based only on the morphology of the viruses, as the order *Herpesvirales.* On the other hand, the order *Tymovirales*, for example, is based on the related RNA polymerase genes that all the members share [18]. To date, there are 55 orders, 168 families, 1421 genera and 6590 species [19] of viruses. In the last years, thanks to the improvements in the field of structural biology, the structure for the coat protein of plenty of different viruses has been resolved. In 2012, Abrescia and collaborators proposed a structure-based classification method. This classification was divided into four main classes being; Picornavirus-like, HK97-like, PRD1/adenovirus like and BTV like. Picornavirus-like lineage is characterized by the single β-barril motif or “jelly-roll” fold, HK97-like for the swan-like motif, PRD1-like by the double β-barrel motif and BTV-like for the right-foot motif [1]. Nevertheless, this method does not include all viral families and they mainly include icosahedral viruses. However helical and enveloped viruses are also believed to fall into a limited number of lineages [1]. Recent computational studies had proposed four extra lineages that include viruses that didn’t fit in the categories proposed by Abrescia. This additional lineages are the retrotranscribing-like, leviviridae-like, inoviridae-like and nidovirales-like [20]. There are still viruses that do not fall into any of those categories and further studies are needed to complete this task.

1.2 PLANT VIRUSES

Plant viruses are a group of pathogens that infect plants, and although they are not a threat for human health, they cause considerable economic losses in agricultural and ornamental crops worldwide. To date, plant viruses account for 25 families and 2284 species characterized [21]. Plant viruses appear in all the aforementioned viral morphologies, but the filamentous architecture is highly prevalent. Their genetic material can be presented in a unique single or double stranded form, or it can be multipartite or segmented. There are not dsDNA plant viruses reported to date and most of plant viruses have (+) ssRNA as their genome. Multipartite genomes are packed into different viral particles while segmented genomes are packed in the same viral particle. Almost all viruses with multipartite genomes infect plants but it remains unclear why [22]. In terms of transmission, plant viruses can be transmitted horizontally or vertically. The vertical transmission is the one that is passed to the progeny of the infected plant, and the horizontal transmission requires external agents for bringing the viral particles to other plants via insects, nematodes, arachnids or fungi. The majority of plant viruses are transmitted by insects, in particular by those belonging to the order Hemiptera, such as aphids, whiteflies, leafhoppers and cicadas [23]. These insects are well suited for transmitting viruses thanks to their needle-like mouthparts that can break the cuticle and cell wall of the plant tissues and insert the viral particles inside [21]. Aphids are the most successful among insect vectors in viral transmission. The transmission through insects can follow different strategies depending on the time that the vector can transmit the viral particles to another plant. It can be non-persistent, semipersistent, circulative non-propagative or circulative and propagative transmission.

In the non-persistent manner, the insect vector feeds itself from the infected plant and keeps the viral particles during a time range between a few seconds and a few minutes. The viruses stay mainly in the insect's stylet and digestive tract [24]. In the semipersistent transmission manner the viral particles remain also in the mouthparts and the foregut, and the insect play as viral vector for a few hours to a few days until the virus is cleared from the vector system [23]. Similar to the semipersistent transmission, the persistent or circulative transmission requires acquisition periods from hours to days but the viral particles are ingested by the vector and then transmitted through the salivary glands. Depending if the virus replicates or not inside the vector, then the transmission method is called "circulative, propagative" or "circulative, non-propagative". In the non-propagative manner, viruses circulate the food system of the vector and then are transmitted again through the saliva canal of the vector [25] and they can remain infective through weeks. In the propagative manner, the virus replicates inside the vector cells and later exits through the salivary glands. This association remains for life and can be passed to the offspring through a transovarial route [26].

Since plant viruses do not affect animals or humans, from an anthropocentric point of view the plant viruses that preoccupy the most are those that produce great economic losses in food and ornamental crops. The are several examples of plant viruses from different families that suppose a great threat worldwide such as: TMV (*Virgaviridae)*, tomato yellow leaf curl virus (TYLCV) (*Geminiviridae*), cucumber mosaic virus (CMV) (*Bromoviridae*), potato virus Y (PVY) (*Potyviridae*), cauliflower mosaic virus (CaMV) (*Caulimoviridae*) or pepino mosaic virus (PepMV) (*Alphaflexiviridae*) [27].

1.2.1 Filamentous plant viruses

Filamentous plant viruses are so called because of their filament-shaped structure. There are 38 genera of viruses that share the filamentous architecture, distributed in 6 different families; *Alphaflexiviridae, Betaflexiviridae, Virgaviridae, Closteroviridae, Potyviridae* and *Ophioviridae* [28]. All plant filamentous viruses are non-enveloped and are composed of multiple copies of a single coat protein that arrange in a helical manner protecting the viral genome. Filamentous viruses can be distinguished between rigid rod-shaped or flexible (Fig 1.3).

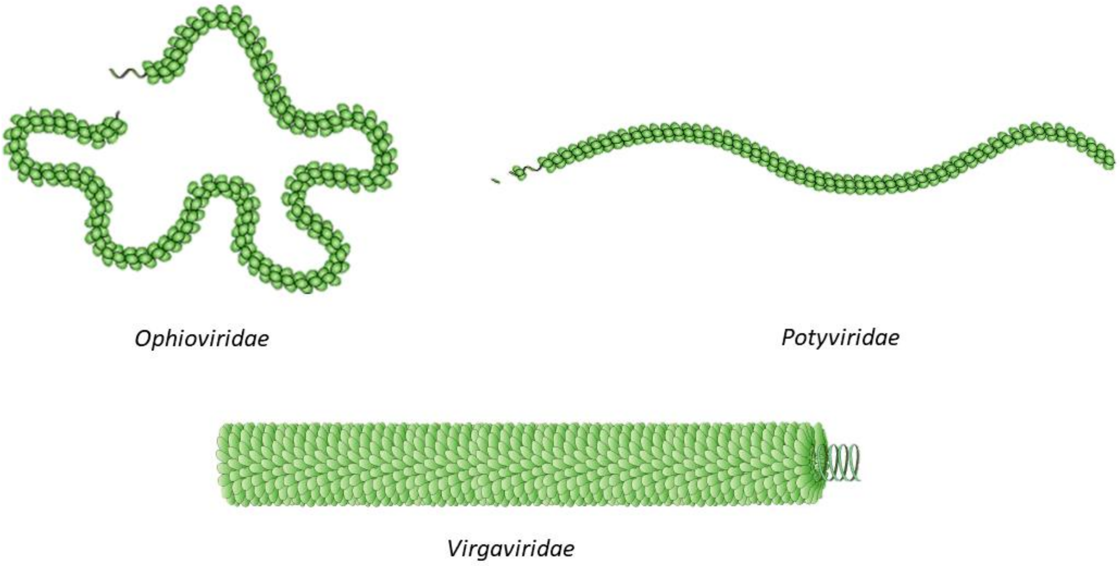


Fig 1.3 Cartoon representation of the different types of filamentous plant viruses. On the upper left there is a representation of a member of the family *Ophioviridae*, which present a highly flexible architecture. The cartoon on the upper right is a member of the family *Potyviridae* with

a less but still flexible architecture, and finally, at the bottom there is a representation of the family *Virgaviridae* which present a very rigid rod-shaped organization.

Tobacco mosaic virus (TMV) is a rigid rod-shaped filamentous virus and it was the first filamentous plant virus which structure was resolved by fiber diffraction methodology [29,30]. Flexible filamentous viruses cannot be crystallized due to their flexibility and it wasn't until the development of high resolution cryo-electron microscopy that the first near atomic structures for flexible filamentous viruses, Pepino mosaic virus (PepMV) at 3.9 Å and Bamboo mosaic virus (BaMV) at 5.6 Å, were reported [31,32].

Flexible filamentous plant viruses are distributed all around the globe and infect a wide variety of crops and ornamental plants. From the 6 families, *Potyviridae* is the one that comprises the largest number of species.

1.2.2 Economic importance of flexible filamentous viruses

Filamentous plant viruses infect plant species all around the globe and they have a huge economic impact, producing losses of several billion dollars every year [33]. The number of agricultural crops and yields is growing due to the necessity to produce enough food for the rapidly increasing human population. In addition, international trade and climate change are contributing to the arising of new viral species resulted from the synergistic interactions between different viruses and their high mutation rates [34]. The viral epidemics are more drastic in tropical and sub-tropical regions with continuous vegetation round the year [35] but it is a problem that affects all countries around the world. There are many examples of plant viruses that cause devastating diseases, as the Citrus tristeza virus (CTV) a member from the family *Closteroviridae* that affects mainly citrus and related species and provoked the death of almost 100 million orange trees [36] in Brazil. Potato virus Y (PVY) from the *Potyviridae* family affects potato crops all over the world and it can interact with PVX (*Alphaflexiviridae*) causing a more severe and acute disease in potato crops. The Plum pox virus (PPV) is another member of the family *Potyviridae* and it is considered one of the most economically important virus diseases of stone fruit worldwide. It produces a disease known as sharka disease. PPV is endemic in most of Europe, Mediterranean basin and Middle East, but it has also spread to America and Asia. It affects plants of the genus *Prunus* and it is very damaging for apricots reaching losses in crops of almost 100% [37].

All of the above are a few examples of filamentous viruses that affect different plant species and crops worldwide and that have a direct economic impact on the countries where they cause the disease. There are several ways to handle viral infections in plant crops. Management of viral diseases involve a series of phytosanitary control measures such as: hygienic measures, chemical control against pathogenic vectors, use of resistant plants against the vector or the virus, use of certified plants that guarantee the absence of the pathogen; virus detecting measures; and the elimination of infected plants and ground decontamination. The measure that has proven to be the most efficient is the use of virus-free certified plants [38]. The use of certified plants is often achieved by the *in vitro* culture of plant tissues. Complete plants can be grown in a controlled environment, starting from different vegetal tissues in a medium containing nutrients and hormones such as auxins and cytokines [39]. This controlled environment allows plants to grow, theoretically, without viral infections. Combination of *in vitro* culture and production of pathogen-free plants is an essential biotechnological tool for plague control, but in the case of viral infections, effective and full control does not exist. The first steps of viral infection are asymptomatic and there is no efficient antiviral treatment. For virus elimination, the propagation of tissues with a minor amount of viral load and thermal treatments are combined but they require several cycles to be effective. For all those reasons, there is a clear need to develop effective chemical treatments for the *in vitro* production of plants free of viruses.

1.2.3 *Potyviridae* family

The *Potyviridae* family is composed of ten genera: Bevemovirus, Brambyvirus, Bymovirus, Ipomovirus, Macluravius, Poacevirus, Potyvirus, Roymovirus, Rymovirus and Tritimovirus. All potyviruses are transmitted by insects, mostly aphids, except of bymoviruses, which are transmitted by amoebae [40]. Potyviruses are non-enveloped, flexuous and filamentous viruses that can vary from 680 to 900 nm long and 11-20 nm in diameter. The genome is a 8-11kb (+)ssRNA that can be monopartite or bipartite as in the case of genus Bymovirus [41].

Potyvirus genus include most of the viruses of the *Potyviridae* family.

1.2.4 Potyvirus genus

The genus Potyvirus comprises 176 described species [42] what makes it the largest and economically most important genus of all plant viruses, producing important economic losses in a wide range of crops. They are distributed all around the world but the majority of species have been reported in the United States, China and Australia, followed by Brazil, India, France, Italy and England [42] (Fig 1.4).

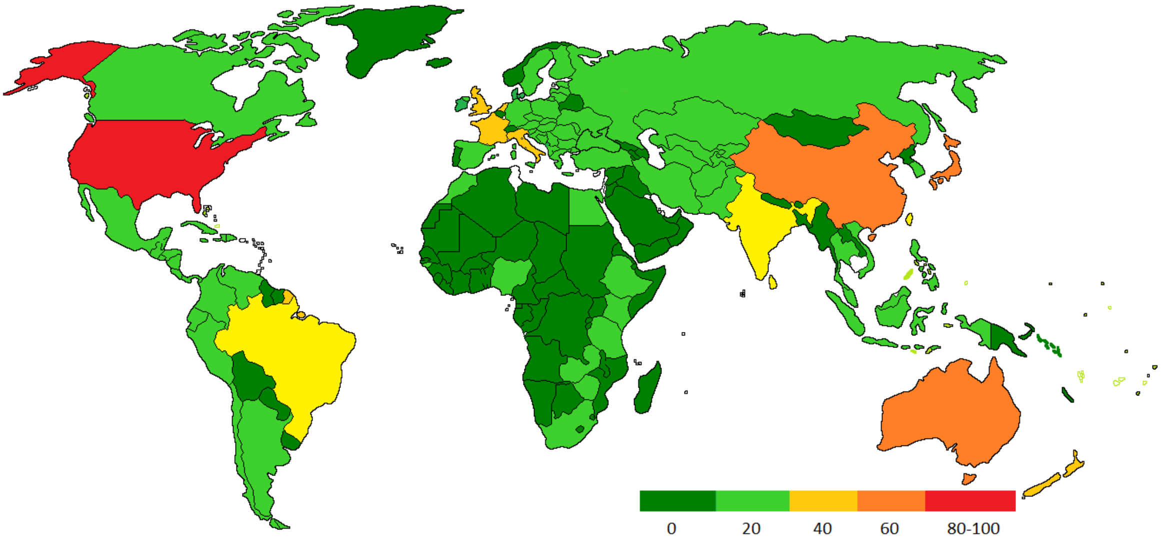


Fig 1.4 World map showing in a colour code the quantity of potyviruses that have been described in each country. The map shows all the countries coloured depending on how many species of potyviruses have been described in each one of them. The colour bar below relates each colour with a range of numbers from 0 to 100. (Adapted image from Gadhave et al [42]).

Potyviruses have a (+)ssRNA genome which contains a long open reading frame (ORF). This ORF is a single polyprotein that is self-processed and encodes for 10 potyviral proteins. The 11th protein, P3N-PIPO, is produced by a frameshift into a short ORF included inside the P3 protein region [43] (Fig 1.5). The polyproteins are co- and post-translationally processed by 3 viral encoded proteases, P1, HCPro and NIa-Pro to generate the mature proteins.

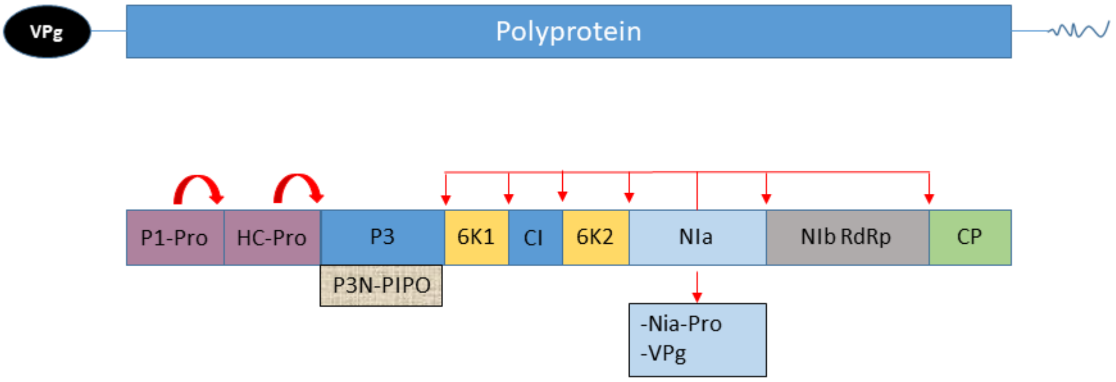


Fig 1.5 Scheme of the organization of potyvirus genome. The upper line is showing the long ORF that encode the polyprotein that will give rise to 10 potyviral proteins as well as the two terminal ends with the VPg attached at the 5' end and the poly A tail at the 3' end. The line above is a scheme of the order in which proteins are processed in the genome.

P1-Pro is one of the least studied potyviral proteins. It is a serine protease that self-cleavages at its own C-terminus [44]. It's known that it's cleavage is essential for viral viability and for the correct function of HC protein [45,46]. The HC-Pro is a cysteine protease that is also cleft at its C-terminus. This is, contrary to P1-Pro, one of the most studied potyviral proteins [44]. Its name comes from "Helper Component (HC) for aphid transmission" and it has been related to have several functions as RNA silencing suppressor, helper in polyprotein maturation and it is also involved in plant to plant transmission [47]. In addition, HC-Pro has been detected at the end of potyvirus virions and it is known to be required to stabilize CP and for the proper yield and infectivity of potyviral progeny [48]□P3 is another poorly studied protein and its biochemical function is not known. It interacts with CI, NIb and NIa [44]. It has been reported that P3 protein contributes to the regulation of virus replication, movement and pathogenesis, although the mechanisms are still unknown [49]. P3N-PIPO has an essential role in virus movement between the host cells [50]. Protein 6K1 has been proposed to have an important role in viral replication [51]. CI protein is a multifunctional cylindrical inclusion protein and it contains ATP binding and RNA helicase activities [44]. It is also implicated in the cell-to-cell movement of the virus, forming cone structures at the plasmodesmata (PD) and interacting with the CP [52–54]. 6K2 is an integral membrane protein and it is role is associated with viral replication by inducing the endoplasmic reticulum-derived replication vesicles [54]. NIa is the protease responsible for the cleavage of the proteins at the central and C-terminal domain of the potyviral polyprotein [44]. NIa is self-cleft in two proteins: VPg and NIa-Pro. The VPg (potyvirus-derived viral genome linked protein) is covalently attached at the 5' terminus of the potyviral genome and it is known to be implicated in the interaction with the eukaryotic translation initiation factor 4E (eIF4E) [55–57], which plays an important role in post-transcriptional control in plants and it is crucial for infection. VPg acts as a cap-structure to initiate viral translation [55]. Furthermore, VPg interacts with the majority of the potyviral proteins and with many other host factors [44,55]. NIa-Pro has DNAse activity and it has been speculated that it is implicated in the degradation of host's cellular DNA , regulating host gene expression [44]. NIb protein is a RNA-dependent RNA polymerase and it is the responsible for potyviral genome replication [44]. Finally, the coat protein (CP) is the responsible for the encapsidation of the viral genome [44] and is involved in cell-to-cell movement [58].

1.2.4.1 Potyviral infection cycle

Virions from the genus potyvirus are transmitted mostly by aphids in a non-persistent and non-circulative manner [42]. The infection process starts when the viral particles are introduced in the host cell cytoplasm, either by vectors or by seed transmission. Once inside, viral particles must, at least partially, disassemble to allow access to the viral genome (Fig 1.6). VPg interacts with the translation initiation factor eIF4E and recruits the translation machinery [55,59]. VPg has also been related to the suppression of host's RNA silencing pathway by mediating degradation of SGS3 (suppressor of gene silencing 3) and SGS3 interaction partners such as RDR6 (RNA-dependent RNA-polymerase 6) [60]. After that, and for the replication process, potyviruses take advantage of the host's secretory pathway and of the cytoplasmic membranes of the endoplasmic reticulum (ER) and the chloroplasts [43,61,62]. Several viral and host's proteins form the so-called viral replication complex (VRC) which are motile vesicles where replication occurs. The VRCs move to the cell periphery and to plasmodesmata (PD) using the vesicle trafficking machinery of the cell secretory pathway [63]. PD are channels present in plant cells that connect the cytoplasm and the ER of two neighbouring cells [58]. The potyviral protein CI forms pinwheel structures at the PD in order to enlarge it and help the newly synthetized viral particles pass through and reach the adjacent cell [63]. The complexes that go through the PD must be assembled virions or at least ribonucleoprotein complexes containing the viral genome associated with the CP [58] .The CP is not only and structural protein but is also essential in virion transport locally, systemically and between hosts. Finally, viral particles are spread from cell to cell and at one point they might be degraded by host defence mechanisms (Fig 1.6).

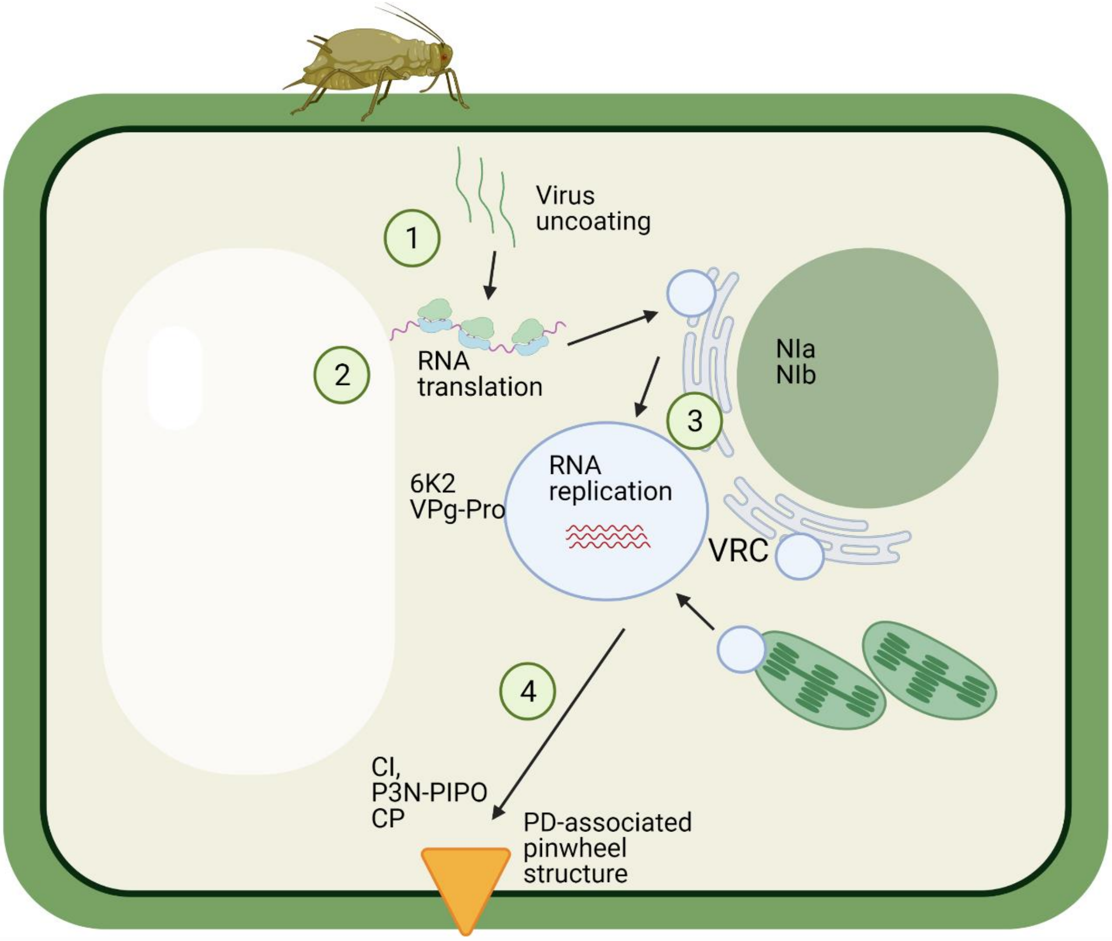


Fig 1.6 Schematic representation of the infection cycle of potyviruses. The aphid introduces the viral particles inside the host cell where viral particles are uncoated. Later, the genome is translated and replication is associated with membranous structures of the cell. Finally, vesicles are transported to the cell wall where the cell-to-cell movement occurs via PD. The proteins implicated in each step are named. Image created with BioRender.com.

1.2.4.2 Potato virus Y

Potato is one of the major crops cultivated globally and viral diseases can reduce potato yields beyond 50% [35]. Potato virus Y (PVY) is a member of the genus Potyvirus and it is transmitted by more than 40 aphid species. It is distributed worldwide and it does not infect only potatoes but also more members of the Solanaceae family as tomatoes, peppers and tobacco [35]. The symptoms that host plants develop when infected with PVY are dark spots or streaks on infected leaves. Those leaves end up falling and leave a bare steam. Infected plants are pale, mottled and present rugose leaves (Fig 1.7, A, B). Moreover, tubers are small and cracked [64] (Fig 1.7, C,D,E,F). Currently, the main strategy to avoid infection of potato crops with PVY is the use of virus free certified tubers but there are many variants of the virus that make the infection very difficult to detect [35].

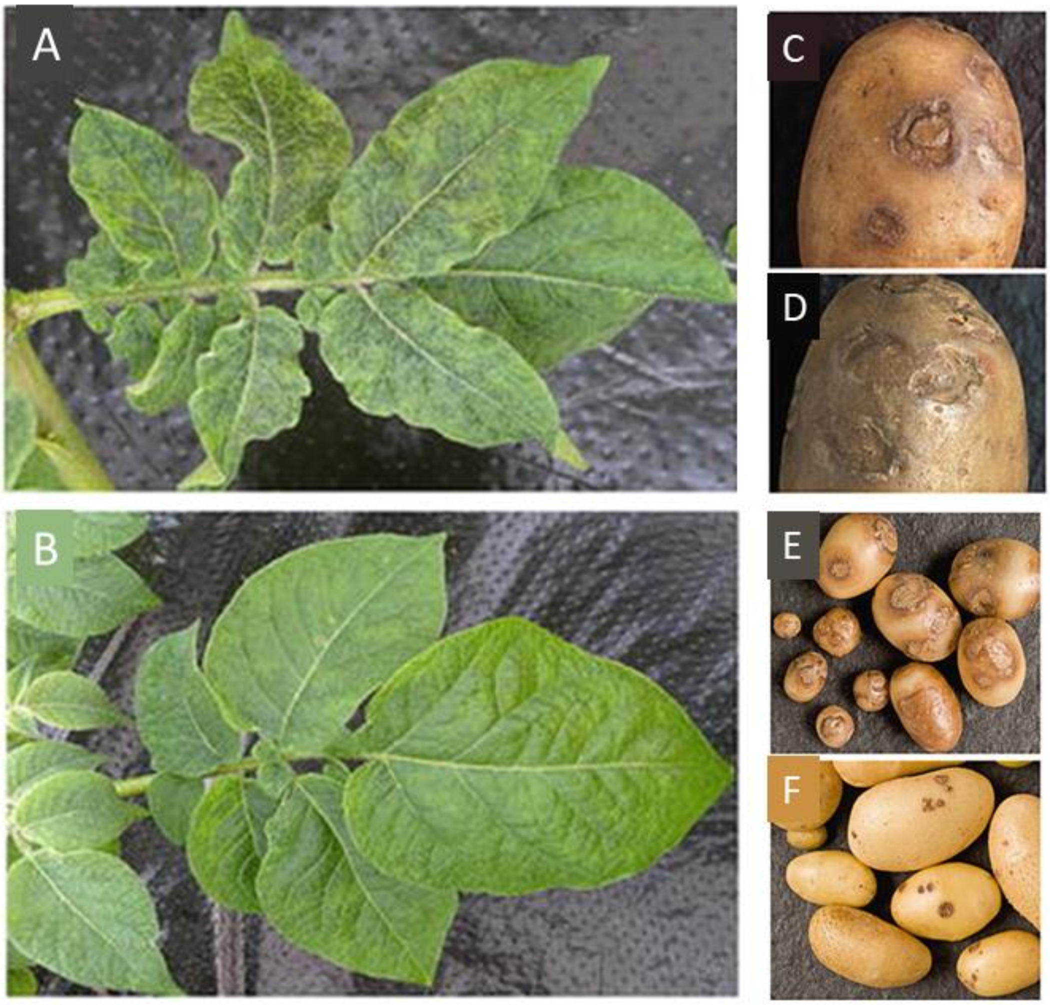


Fig 1.7 Symptoms in potato plants infected with PVY. Images A and B show the symptoms of the infection with PVY on the leaves of infected plants. Images from C to F display the symptoms on the potato tubers. Image taken from Lacomme and Jacquot [65].

1.2.4.3 Turnip mosaic virus

Turnip mosaic virus (TuMV) is also a Potyvirus and it causes damage in a wide range of different plants all over the planet infecting members of the genus *Brassica* among others [66]. It is transmitted by more than 40 aphid species and the symptoms depend on the strain of TuMV and several other factors. The symptoms that it produces are variable, but the most commons are vein clearing, mosaic, necrosis and plant death (Fig 1.8).

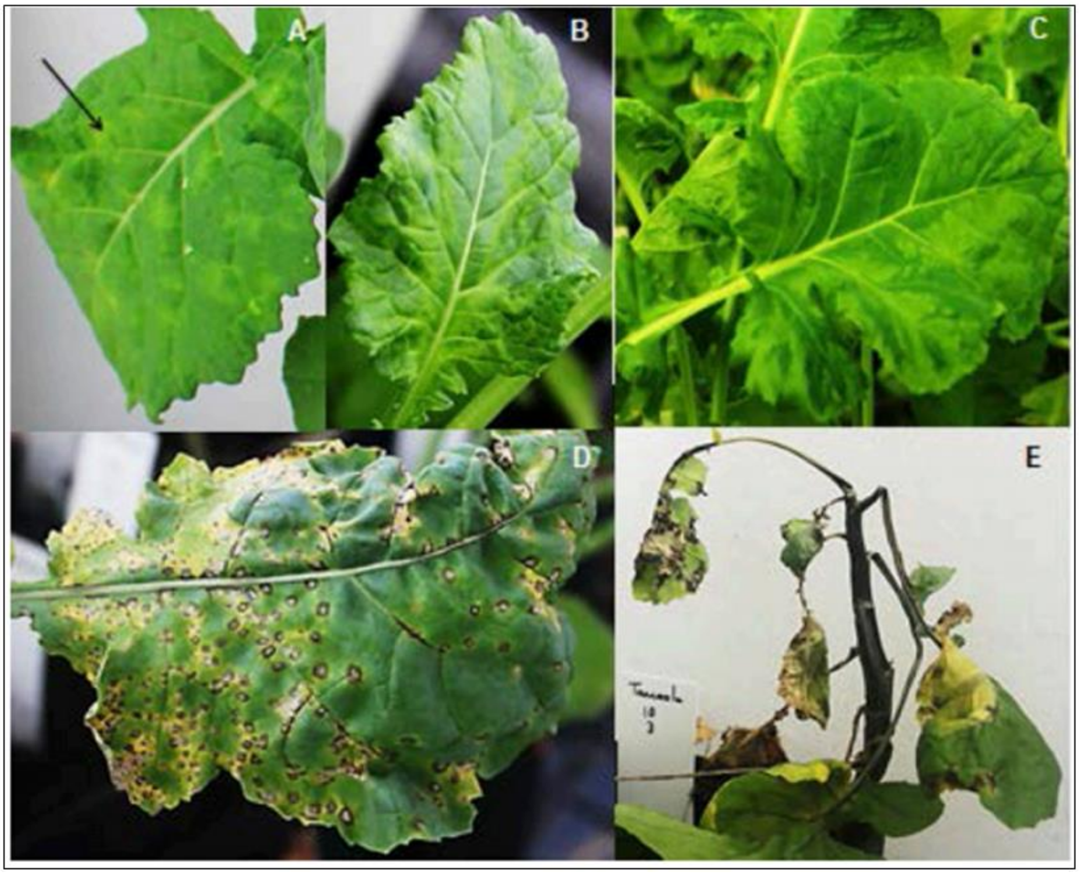


Fig 1.8 Symptoms of TuMV infection in *Brassica napus* plants. A to E show different symptoms of TuMV infection, as chlorotic spots in young leaves (A), mosaic symptoms (B), leaf deformation (C), necrotic spots (D) and systemic necrosis (E).

1.3 VIRUS LIKE PARTICLES

Virus-like particles (VLPs) are multiprotein structures that resemble the virus but lack the viral genetic material required for infection [67]. Viral VLPs have been widely studied for their applications in nanobiotechnology and nanomedicine. For a particle to be suitable for nanobiotechnology it must adopt a series of characteristics as the capacity to introduce additional functional groups in their structure, to have a consistent size, to be environmentally friendly, and to be easy and cheap to produce [68]. Viral VLPs fulfil all the above and, moreover, plant viral derived VLPs have some additional characteristics that make them even better scaffolds for their use in biomedicine as the fact that plant viruses are non-pathogenic to animals and humans. Furthermore, plant derived VLPs can be produced easily and in great amounts *in planta* or even in bacterial, insect or mammalian cells [68,69]. One of the useful applications of VLPs is their use as nanocarriers of therapeutics or bioimaging dyes inside the human body. Foreign molecules can be inserted inside the body of the VLPs and use them as vehicles for drug delivery or for the imaging of specific tissues [68]. One of the most interesting applications of plant derived VLPs in nanomedicine is their use in vaccine development. The high structural organization of helical plant viruses makes them very good candidates for this task and VLPs can be modified to express immunogenic epitopes on their surface, present antigens in the human or animal body and enhance the immune system thanks to their repetitive architecture [68–71]. Plant virus derived VLPs have low probability of pre-existing immunity and also have the ability to adapt to a great variety of different conditions [68].

BaMV is a member of the genus Potexvirus and its particles are flexible and rod-shaped. Epitopes of the capsid protein of foot-and-mouth disease virus (FMDV) have been expressed tagged to the coat protein of BaMV, and inoculation of swine with the chimeric virion produced neutralizing antibodies [72].

Papaya mosaic virus (PapMV) is another member of the genus Potexvirus for which its potential as vaccine carrier has been explored. The CPs of the VLPs derived from PapMV virions were fused with epitopes from Hepatitis C virus and their inoculation in mice produced long-lasting humoral immunity [73]. PapMV is also studied for cancer vaccines as it has been seen that PapMV activates antitumor immune responses in mice [70].

These are just a few examples of how plant virus derived VLPs can serve as drug carriers or as scaffolds for vaccine development, but a lot more have been and are being studied, including cowpea mosaic virus, alfalfa mosaic virus, TMV, potato virus C, PVY, PVX, tomato bushy stunt virus, zucchini yellow mosaic virus, plum pox virus, cucumber mosaic virus and cowpea chlorotic mottle virus among others [68–71,74–76].

Structurally, viral VLPs haven't been studied and knowing the atomic structure of VLPs CPs would entail a great advantage in antigen presentation systems where structure-guided designs could be developed.

1.4 ELECTRON MICROSCOPY

The first electron microscope was built in 1931 by two electrical engineers, Ernst Ruska and Max Knoll for which Ruska was awarded the Nobel Prize in Physics in 1986. The power of the light microscope is limited by the wavelength of light, electron microscopes, on the other hand, can produce much more highly magnified images because the beam of electrons has a smaller wavelength which creates images of higher resolution [77]. An electron microscope is composed of an electron source and a battery of electromagnetic lenses. The electrons source is called the electron gun. Inside the electron gun there is a bent wire and an accelerator stack. The current of electrons passes through the wire and eventually some electrons will have enough energy to escape the tip and go into the microscope column. Those electrons that escape the tip of the filament should be temporally and spatially coherent, this means that they should be escaping on the same direction, with the same speed, and same wave properties with their phases in register. Once the electrons have escaped the source, they pass through a series of lenses that control the beam. The first set of lenses are the condenser lenses that focus the electrons and direct them to the sample. The sample holder is placed right below the condenser lenses, so the electrons hit the sample at this stage. Next, the electrons that have interacted with the sample are scattered and will pass through the objective lenses responsible of the image formation. Later on, a set of projector lenses will augment the magnification and will sent the image onto a fluorescent screen at the bottom of the column or to a detector that will produce the final image [78].

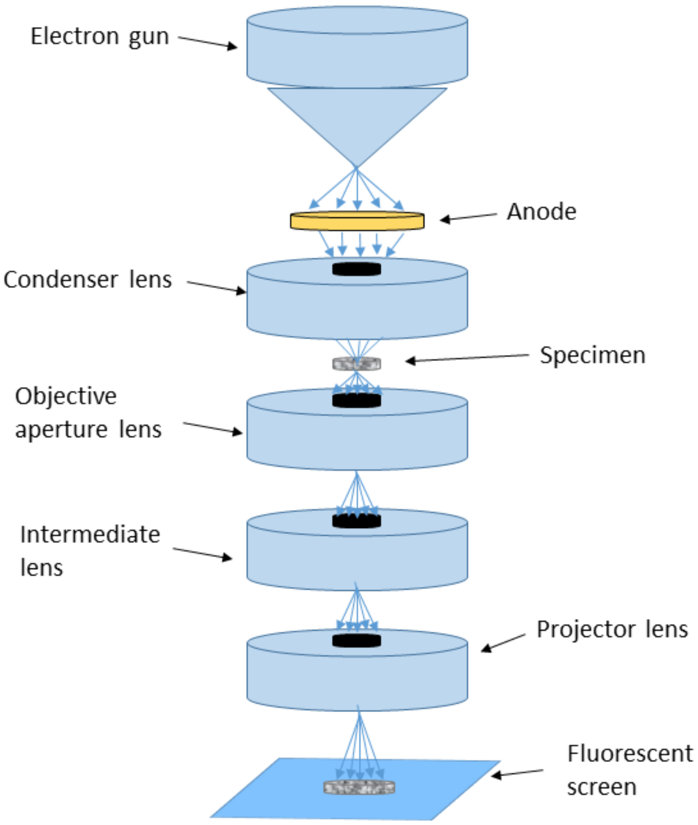


Fig 1.9 Schematic representation of the interior of a transmission electron microscope. Different lenses and objects are presented schematically in the order that those elements are placed inside the column of the TEM.

There are different kind of detectors; photographic films, “charge coupled device” (CCD) cameras and direct detection cameras (DDC). The first microscopes used photographic films as detector system. The images obtained in photographic films had to be processed and digitalized for later image processing. CCD cameras take digital images automatically, but they cannot detect electrons directly and follow a two-step method to create the image. They are composed of a scintillator that transforms the electrons into photons. Below the scintillator there is a bundle of fibre optic where the photons are transferred to a CCD camera where the light is converted to charges which are read out and a digital image is formed. The disadvantage of CCD cameras is that they are thick devises and the electrons that hit the scintillator can be scattered in very different directions, giving noisy and degraded images at high resolutions. In direct detector cameras, the electrons hit the detector and are directly converted into digital signal minimizing the spread of electrons and preserving the information at high resolution.

1.4.1 Image formation in the electron microscope

Electrons are charged particles which have properties of particle and wave at the same time. Electrons are fired in a coherent beam by the electron source at a desired voltage. Being coherent means that the electron waves in the beam have the same wavelength, amplitudes and phases (Fig 1.10, A). The wavefront is focused to the beam axis of the microscope by the series of lenses and deflectors. Biological samples have very low contrast and usually have very few scattering points (particles) so, when the wavefront hits the sample, most of the electrons pass through it without interaction (unscattered electrons). A smaller fraction of electrons interact with the sample and are scattered in an elastic or inelastic way. Inelastic scattering means that there is an energy transfer between those electrons and the atoms of the sample. Elastically scattered electrons are those that interact with the sample but do not lose their energy (Fig 1.10, B). The final image will be formed as an interference between all the scattered and unscattered electrons. Inelastic scattered electrons will contribute to the noise in the image. Scattering produces a phase shift in the phases of the wavefront (Fig 1.10, A).

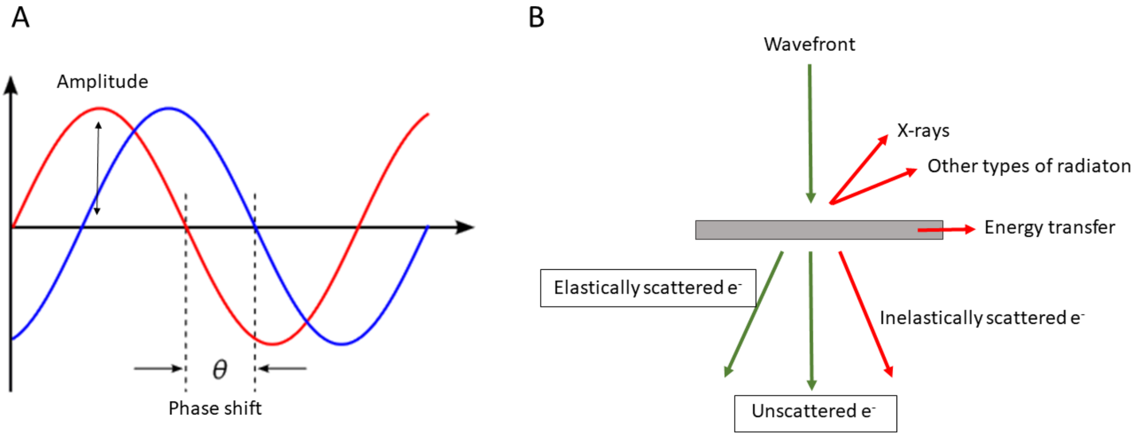


Fig 1.10 A) Representation of the amplitude and phase parameters in sine waves and B) Schematic representation of how electrons interact with matter. In **A)** two sine waves are

plotted in blue and red. The amplitude parameter is displayed for the red wave and the phase shift between both waves is also displayed in discontinuous lines. **B)** This image shows a scheme of how the electron wave interacts with the sample and the distinct types of scattering that happens.

The final image is obtained using the information of the amplitudes and phases, and taking into account the modulation of the signal produced by the microscope with a mathematical function called the contrast transfer function (CTF).

1.4.2 Cryo-electron microscopy

Structural biology is the branch of biology that studies the 3D structure of macromolecules. There are three main techniques for resolving macromolecular structures: X-ray crystallography; nuclear magnetic resonance (NMR); and EM. X-ray crystallography usually reaches atomic resolution, but it needs a huge amount of sample and the crystallization can be a tough task since not all proteins crystallize well and the improvement of the crystals can take a very long time. NMR has the advantage that molecules can be in solution and allows to study the dynamics of the sample. Nevertheless, molecules are limited in size and usually they need to be smaller than 50 kDa. In contrast, EM needs little amount of sample, it is suitable for proteins larger than 50kDa, and it does not require crystallization.

Classic electron microscopy has several limitations that make it impossible to achieve the reconstruction of macromolecules at high-resolution from its micrographs. First, the column of the microscope must be in high vacuum what is counter-productive for biological samples because the vacuum causes dehydration. Second, a biological specimen cannot be irradiated with a high dose of electrons because the radiation damage will also destroy it, so images have to be recorded at low electron doses what results in a poor signal to noise ratio (SNR), what makes difficult to get information at high resolution [79]. However, it was proposed that hydrated biological samples could be maintained in vacuum if they are previously frozen [80]. To address this issue, Jacques Dubochet developed a protocol in which samples were rapidly plunged and frozen into liquid ethane previously cooled with liquid nitrogen. This rapid freezing results in the formation of non-crystalline ice (vitreous ice) that maintains the sample embedded inside [81,82]. Frozen samples can tolerate better the dose of electrons, augmenting the SNR. CryoEM imaging is made in classic transmission electron microscopes that have been adapted for cryo samples. These microscopes have a chamber where the sample is kept frozen during the imaging and data collection. In the recent years, several technical advances, as the introduction of DDC, development of better and more powerful computers and improvements in reconstruction algorithms, have been made, and cryoEM has passed from being a technique that could resolve structures at medium-low resolution, to being capable of achieving high and atomic resolution, what has been called as the “resolution revolution” [83]. The majority of structures deposited in the protein data bank (PDB) data base were resolved using X-ray crystallography and some by NMR very few structures were resolved by EM, but after the “resolution revolution” the quantity of structures resolved by cryoEM has grown exponentially almost reaching the numbers of X-ray crystallography (Fig 1.11). The highest resolution achieved by cryoEM is the one for apoferritin at 1.25 Å [84].

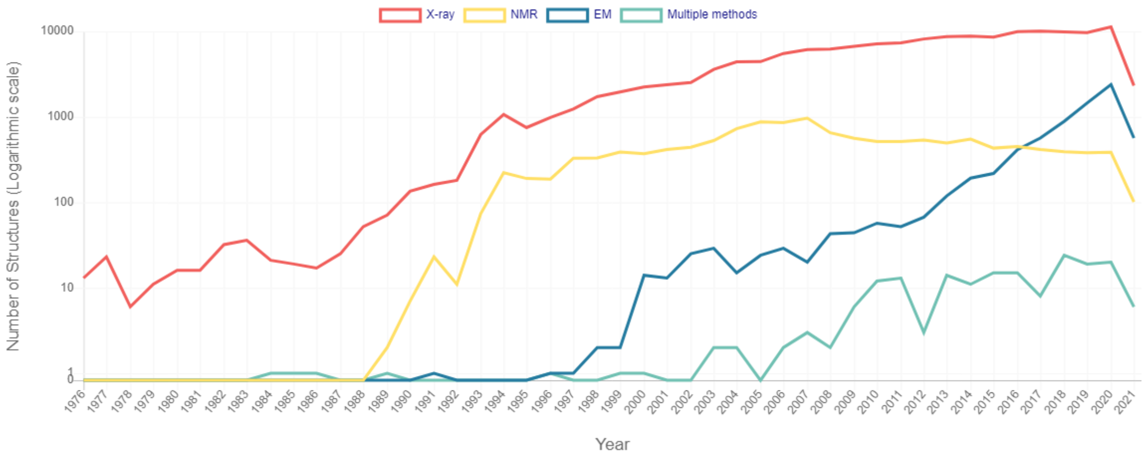


Fig 1.11 Graphic showing the amount of structures deposited per year by different structural methods from 1976 to 2021. The graphic shows the number of structures uploaded per year in the PDB data base by X-ray crystallography in red, NMR in yellow, cryoEM in deep blue and multiple methods in light blue. Image taken from the PDB database.

CryoEM is a powerful technique that is continuously growing and improving, and it has been stablished as the main method for resolving the structure of big and flexible macromolecules.

1.4.2.1 Image processing of cryoEM data

Electron micrographs obtained in the TEM have to be computationally processed in order to get the structural information from the thousands of particles that will be extracted. Particles obtained from the micrographs are very noisy and the aim is to determine their relative orientations in order to average them together and improve the SNR to recover the high-resolution information [85]. This method is called the single particle analysis (SPA).

Image processing is based on the extraction of the particles from the micrographs. Each particle is a 2D projection of the object of interest defined by the three Euler angles and two translation parameters with respect to a fixed reference system. In SPA, the relative orientations of the particles must be determined and later the particles are aligned together in order to reconstruct a 3D model of the object of interest. There are two different reconstruction scenarios depending if the structure of the molecule is known or a similar structure is available or, on the contrary, no previous structural data is available.

On one hand, if the structure is known, a 3D density map of a similar structure can be given as initial reference, as well as a 3D map with the overall shape of the structure (a cylinder for helical objects etc…). In this scenario, projection matching algorithms are used. The initial reference will allow to determine the Euler angles of the sample in an iterative manner. For that, the 3D reference is projected in multiple 2D projections. Each experimental particle image is compared with each of the 2D projections of the reference. The correlation coefficient from each comparison will designate the angles of the 2D projection to the experimental images. Later on, that experimental images are back-projected to generate a new 3D model that will serve as reference for the next iteration process. This will improve the assignment of the angles until the process converges. Finally a 3D density map of the object of interest is generated [86] (Fig 1.12, B,C,D)

On the other hand, if a "de novo" structure is being reconstructed, then the first 3D model must be generated from the same experimental set of single particles. One of the possibilities is to use the so-called "common lines" procedure which is based in the Fourier central section theorem. This method is based on the principle that the fourier transform (FT) of a pair of 2D projections of the same object shares one 1D projection in the 3D FT of the object [85]. Common lines have the same amplitudes and phases in the FT of two separate class averages. With three projection images, there are three common lines and it is possible to find the angles between two common lines with respect to the third one [85,87] (Fig 1.12, A).

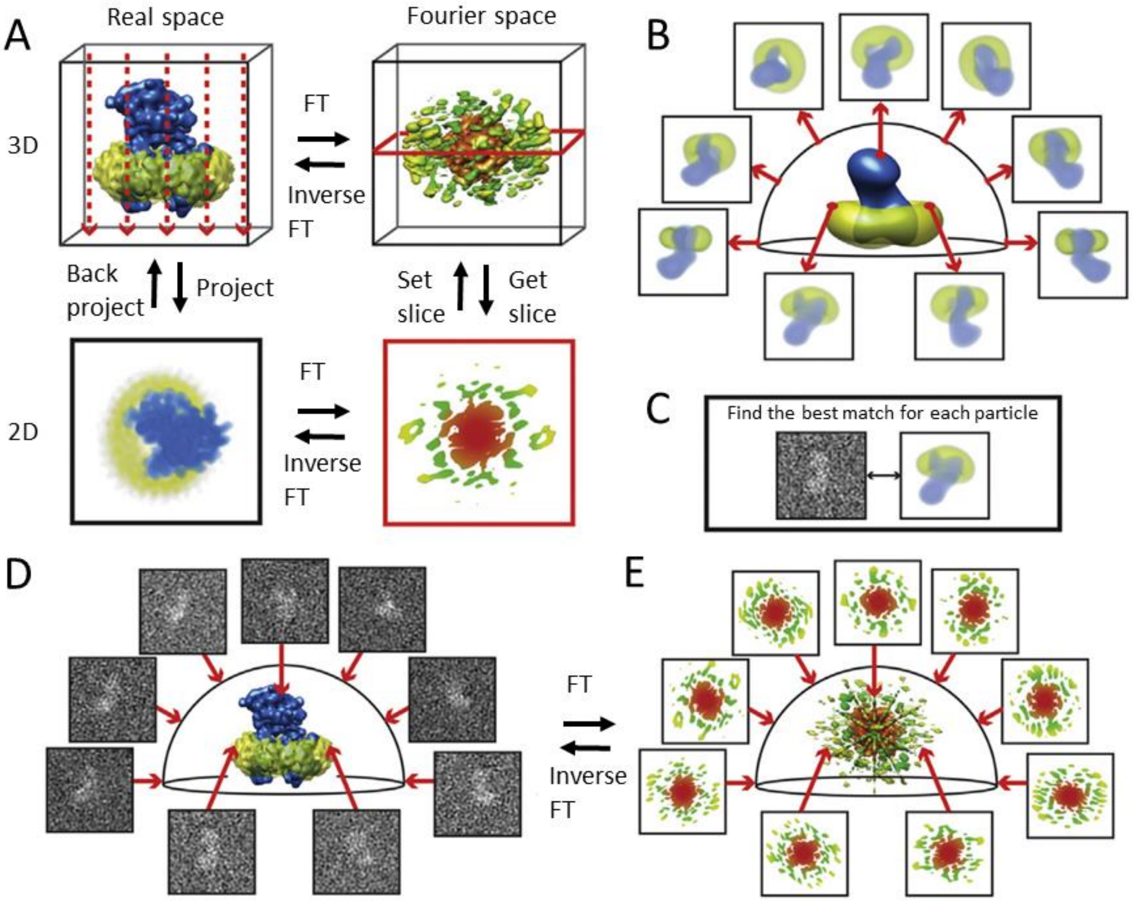


Fig 1.12 Diagrams showing the different strategies for single particle image reconstruction. A) Schematic representation of the central section theorem in which 2D projection of a 3D image is related to the central slice of the 3D FT of the object. **B), C) and D)** are and explanation of the projection matching strategy. This scheme takes an initial 3D reference and projects it in 2D (B). Those projections are cross-correlated with the experimental 2D projections (C) and then the

angles of the 2D initial reference projections are assigned to the experimental 2D projections that cross-correlated with them. Finally, those projections are back-projected and a 3D model is reconstructed with the experimental 2D projections (E). Image taken from Nogales and Scheres [88].

Samples with helical symmetry follow a different reconstruction approach due to their special features. Helical specimens have the same asymmetrical unit repeated along the z-axis of the helix [89]. Helical structures are characterized by the radial localization of the subunits that conform it with respect to the central helical axis, defining the number of subunits per turn of the helix, equivalent to divide 360 by the rotational angle per subunit. The helical rise describes the distance between two adjacent subunits of the helix along the same helical axis. The twist is the angle of that rise and goes between -180˚and +180˚ being negative when the helix is left handed and positive when it is right handed [89] (Fig 1.13).

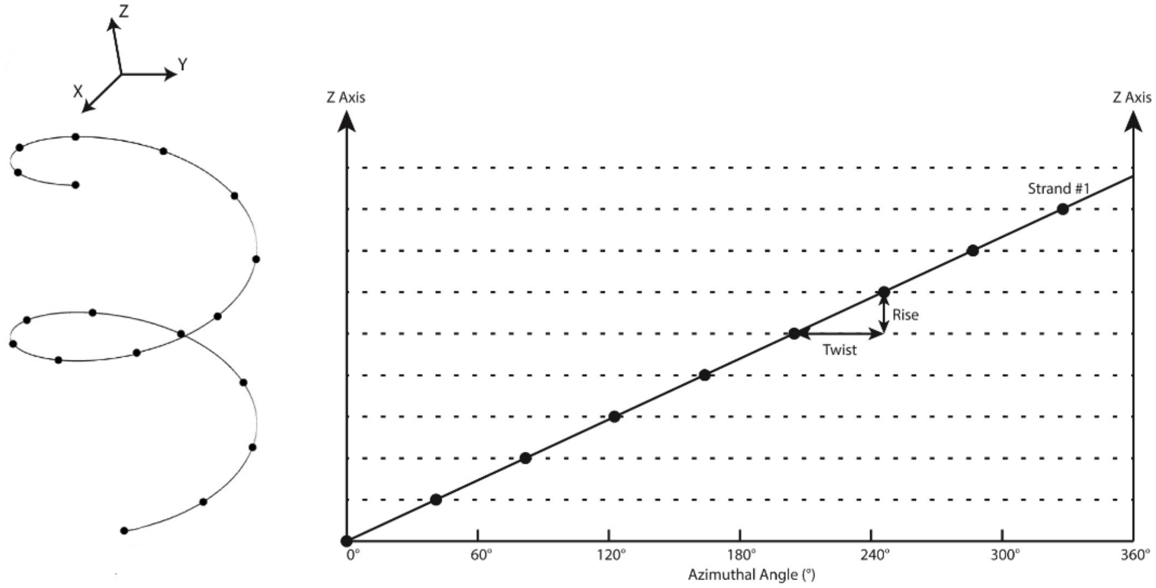


Fig 1.13 Representation of a helix and its principal symmetry parameters. The image on the left shows a folded helix with different subunits in the form of dots along the Z-axis. The image on the right is a linear representation of the helix with the subunits that form it at different levels in the z-axis. The rotation angle (azimuthal angle) is shown in the x axis and it goes from 0˚ to 360˚. The helical rise and twist are represented between two subunits. Image adapted from He and Scheres [89].

The rest of the image processing for helical specimens is the same as for SPA but taking into account the helical parameters of the sample in the extraction, 3D refinement and 3D classification steps.

1.4.3 Cryo-electron tomography

CryoEM has proved to be a strong technique to deliver atomic resolution structures of proteins and other macromolecules, but it is not very suitable for highly heterogeneous and pleomorphic samples [90]. For this reason, electron tomography (ET) is the most suitable methodology available. In ET the sample is tilted inside a TEM in order to get image projections from different angles. Those projections are later aligned and a 3D density map, a tomogram, is obtained. The resolution that ET usually achieves is in the range of nanometers [91]. The technical advances in cryoEM have also been applied to ET leading to cryo-electron tomography (cryoET), where the sample is plunge frozen upon vitrification the same way as in cryoEM. For cryoET the sample is vitrified and then inserted into the TEM. Then the stage where the sample is placed is tilted over ranges of -60° to +60°. There are different schemes for acquiring tilted images, unidirectional where it goes from -60° to 60° directly, bidirectional goes from 0° to -60°, goes back to 0 ° and then goes to +60°, and the dose symmetric where it goes from -2° to +2°, -4° to +4° etcetera… Usually images are taken after 2° or 3° degrees of tilting [92]. The dose of electrons has to be low as for cryoEM but in the case of cryoET the total dose is distributed along all the angular views resulting in a poorer SNR, for that cryoET has become more popular with the introduction of direct electron detection cameras that can reduce the noise and lead to better resolutions [90]. Once the tilted images are acquired and aligned together, we can reconstruct a 3D tomogram (Fig 1.14).

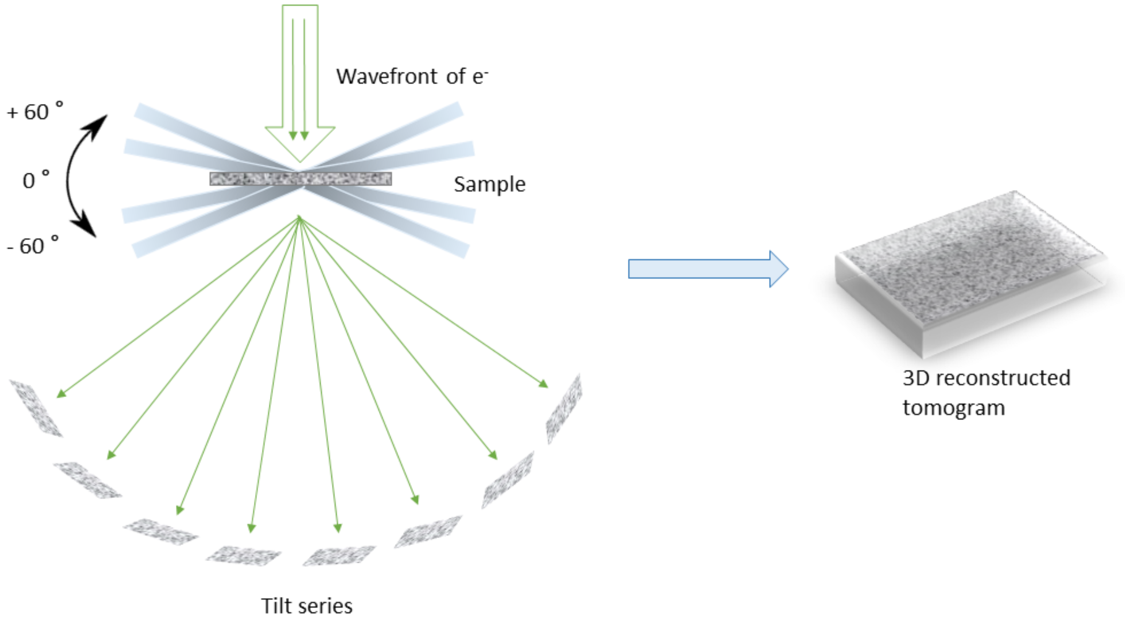


Fig 1.14 Scheme of tomogram data acquiring in the TEM. The image is displaying how the sample is tilted in a range of angles from -60° to +60°. As the wavefront of e$^-$ hit the sample, images are acquired at different degrees of tilting, what is called a tilt series. From the tilt series, a 3D tomogram can be reconstructed after aligning the tilt series between each other.

1.4.3.1 Subtomogram averaging

Subtomogram averaging is a quite new methodology for performing something similar to single particle cryoEM but using cryoET tomograms. 3D volumes of single particles are extracted from the tomograms. Each 3D volume has information of the particle at all the angles that the sample has been tilted, not as in cryoEM where for each particle we only have information of one view. Later, those 3D volumes can be averaged together in order to reduce the noise of the images and try to improve the resolution to observe structural details (Fig 1.15).

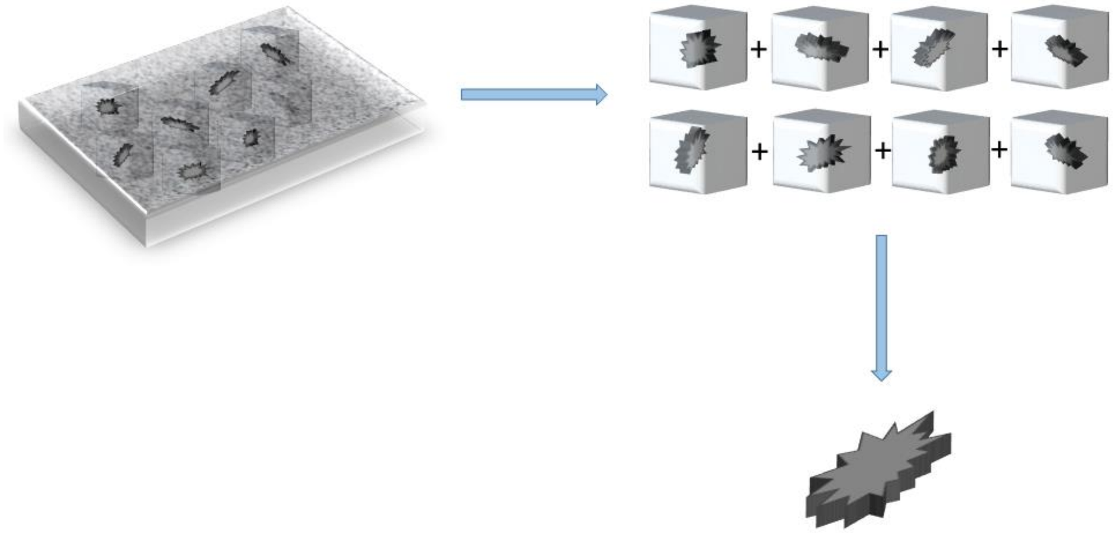


Fig 1.15 Simplified representation of sub-tomogram averaging methodology. The particles of interest are extracted in 3D from the tomogram in boxes. Those 3D particles can be averaged together to improve the SNR and give rise to a 3D map of the sample of interest.

Image processing of subtomogram particles extracted from 3D tomograms is done very similar to the SPA method but it comprises a series of particularities due to the image acquisition procedure. As it was explained before, the sample is placed in the sample holder, and it is rotated between a series of given angles. The geometry of the holder makes it impossible to tilt it at angles in the 70-90° range, so there is a cone of missing information. This is called the missing wedge [93]. The missing wedge appears as a deformation of the reconstructed volume along the direction of the missing information. Moreover, the tilting procedure incorporates drifting and motion in the images, the defocus is not uniform and it varies along the tilting angles, and the dose of electrons used has to be fractionated among the images taken at different tilting what increases the noise and decreases the signal of the recorded images, making the SNR lower.

This technique is rather new and the number of structures deposited in the PDB coming from subtomogram averaging represent the 6% from the total of those coming from EM methods, but it is expected to grow in the upcoming years.

1.4.4 Previous knowledge on the structures of flexible filamentous plant viruses

Until few years ago, only the structure for rigid rod-shaped filamentous viruses had been solved thanks to their high symmetry and rigid architecture. TMV was the first reconstructed structure of a virus, at 2.9 Å resolution [29]. Since then, more structures of members belonging to the genus Tobamovirus have been reported, such as the structures for Barley stripe mosaic virus (BSMV) [94], Ribgrass mosaic virus (RMV) [95], Hibiscus latent Singapore virus (HLSV) [96] and Cucumber green mottle mosaic virus (CGMMV) [97]. However, the intrinsic flexibility of flexible filamentous viruses makes them very difficult to crystallize and study them by the techniques used for rigid rod-shaped viruses, as fiber diffraction. It wasn't until the improvements in cryoEM and reconstruction algorithms that the first structure of a flexible filamentous virus was resolved. Since then, the structures of two members of the familiy *Potyviridae* (genus Potyvirus), WMV [98] and PVY [99], and three members of the family *Alphaflexiviridae* (genus Potexvirus), BaMV [100], PepMV [31] and PVX [101] have been resolved at near-atomic or atomic resolutions. All of them present a very similar overall structure of the whole filament. The filaments are left-handed helices with approximately 8.8 subunits per turn, and a pitch of around 34-35 Å (Fig 1.16).

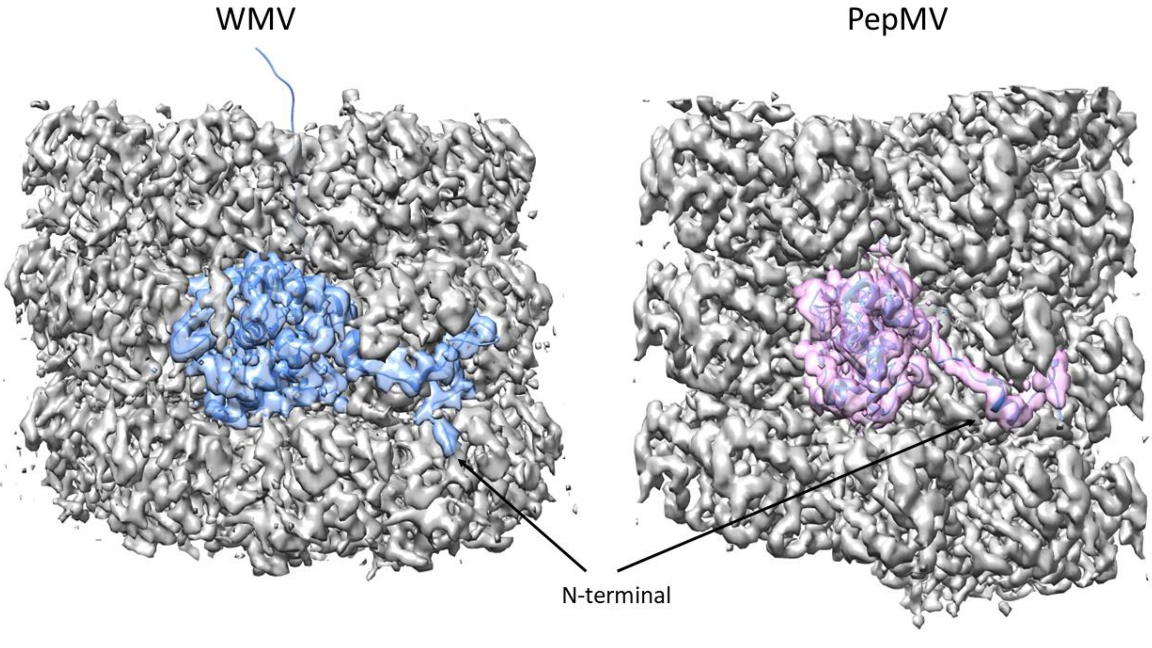


Fig 1.16 CryoEM maps of WMV and PepMV displaying the atomic models for their CPs, respectively. Image on the left shows the density map for WMV in grey (EMDB code: EMD-3785), with the density map for the CP in blue and the atomic model of the CP (PDB code: 5odv) fitted inside in dark blue. The image on the right is showing the density map for PepMV in grey (EMDB code: EMD-3236), with the density map for the CP in pink and the atomic model of the CP (PDB code: 5fn1) fitted inside in dark blue. Both N-terminal arms are labelled.

The near-atomic resolution allowed the modelling of atomic structures for their CPs. Although potyviruses and potexviruses do not share sequence homology between their CPs, it have been demonstrated that these CPs from different families display a clear structural homology. Both genera CPs have a well folded helical core and two long and unstructured terminal arms (Fig 1.17).

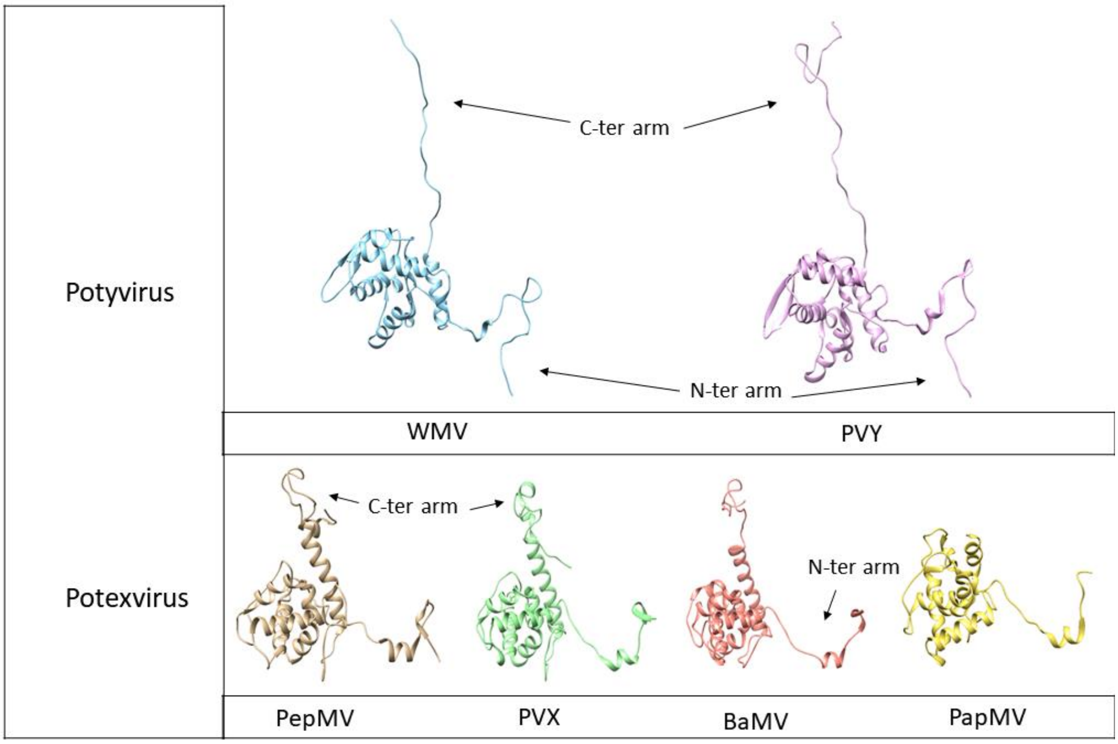


Fig 1.17 Comparison between the CPs of members from the genus Potyvirus and members from the genus Potexvirus. In the upper row, the two atomic models for the CP of members of the genus Potyvirus, WMV and PVY, are shown. Both share the same core folding and two long unstructured terminal arms. Both N-terminal arms are incomplete. The bottom row is displaying the four atomic models of the CPs of different members of the genus Potexvirus: PepMV, PVX,

BaMV and PapMV. The four of them present a very similar architecture with a well folded helical core and two terminal arms. The C-terminal of PapMV isn't present in the atomic model.

The study of flexible filamentous viruses CPs has brought to light a conserved region at the ssRNA binding site. CPs from both families present identical ssRNA binding sites, constituted by a binding pocket where the genomic RNA is accommodated orienting a nitrogenous base towards the inner site of the pocket (Fig 1.18, A). Several residues in the binding pocket have been found to be conserved in the same positions in members belonging to several families of flexible filamentous plant viruses. Those conserved residues are; Ser (S), Arg (R) and Asp (D) (Fig 1.18, A, B)[31,32,98,99,101,102] . Moreover, functional studies with PepMV virions demonstrated that point mutations in this conserved region block viral assembly and multiplication *in vivo* [31], what reinforces the crucial role of those three residues for the binding of the ssRNA to the CP.

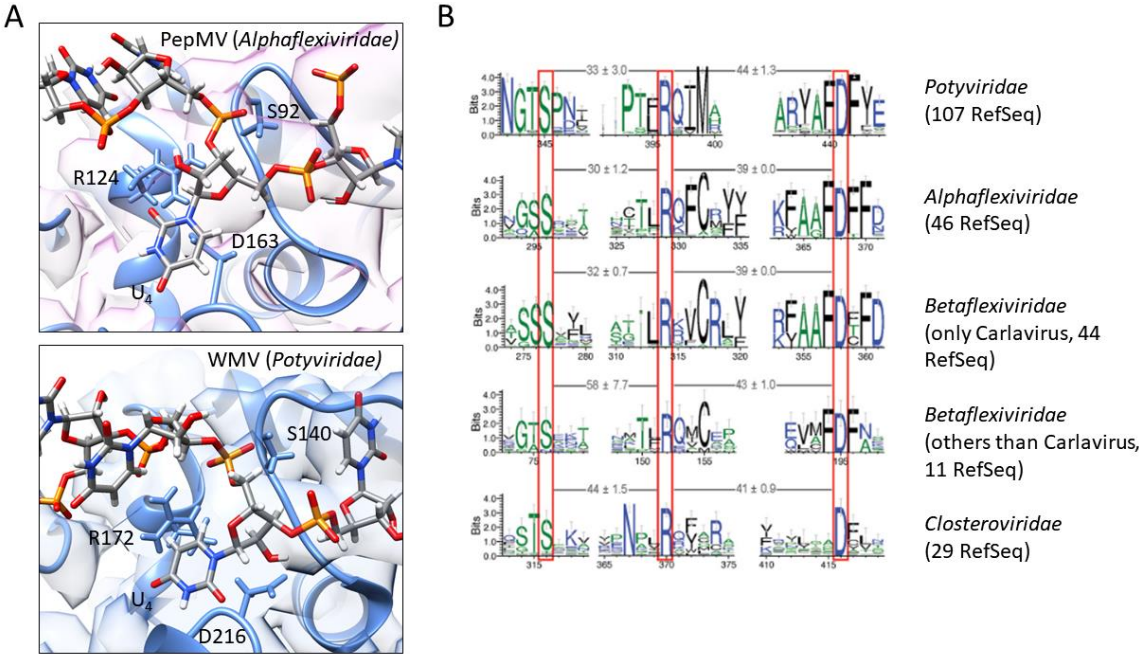


Fig 1.18 ssRNA binding site in PepMV and WMV and conserved residues along flexible filamentous viruses. A) Image above shows the binding site of PepMV with the three conserved residues, S92, R124 and D163 highlighted. The U4 from the ssRNA that is placed towards the binding pocket of the CP is also highlighted. Image below, shows the same elements but for WMV. In this case, the residues are S140, R172 and D216. **B)** Alignment between the consensus sequences for nucleoproteins of several flexible filamentous plant viruses grouped by different families, image taken from Zamora et al [98].

Chapter 2:

OBJECTIVES

1) Structural characterization of TuMV virions and its VLPs by cryo-electron microscopy.

2) Characterization of TuMV virion viral ends by cyro-electron tomography and subtomogram averaging.

3) Evaluation of PVY VLPs as platform for the structural characterization of the binding of selected chemical compounds to viral CPs by cryo-electron microscopy.

Chapter 3:

MATERIALS AND METHODS

3.1 PRODUCTION AND PURIFICATION OF PLANT VIRUSES AND VLPs

3.1.1 TuMV virions and VLPs

For TuMV virions and VLPs, the process of production, purification and measurement of sample concentration were conducted at the "Center for Biotechnology and Plant Genomics" by Dr. Fernando Ponz's group.

For the production of viral particles, several *Brassica juncea* plants were infected with TuMV virions. For that purpose, a preinocule from 5 *Nicotiana benthamiana* plants was prepared. Inoculation was made up mechanically, spraying carborundum at them in order to produce microinjuries on the tissue. Later on, 5 µl from the inoculum were deposited and extended on the surface of selected leaves. After 2 weeks, leaves from those 5 specimens presenting symptoms of infection were collected and triturated in 5 mL of 50 mM pH 7.5 dipotassium phosphate buffer. With this extract *B. juncea* plants were inoculated, spraying them again with carborundum. Infected plants grew for 4 weeks before purification.

For VLP production, agroinfiltration of at least 30 specimens of *N. benthamiana* is required. *Agrobacterium tumefaciens* media was transformed with the CP construction via pEAQ-HT-Dest1 vector. This vector allows the expression of recombinant proteins in agroinfiltrated plant tissue [103]. Media was grown to an optical density at 600 nm wavelength (OD_{600}) of 1.2, pelleted at 2000x g and resuspended with MMA buffer (10 mM MES, pH 5.6; 10 mM magnesium chloride; 450 µM acetosyringone). Resuspended media was incubated for 24h and agroinfiltrated into the leaves of five week old plants, using a blunt-ended 2 mL syringe. Tissue was harvested 10-12 days post-agroinfiltration.

For TuMV virions and VLPs purification, plant tissue was grinded with an electrical tissue grinder (1:2 w/v) at 4°C in potassium phosphate buffer 0.5 M pH 7.5. Resulting suspension was extracted with one volume of chloroform at 4°C. Phases were separated by centrifugation and the liquid phase was separated through Miracloth. Next, virions were precipitated for 90 minutes at 4°C with 6% of PEG 6000 (w/v) and 4% NaCl (w/v). Particles were recovered by 10 min centrifugation at 12.000x g. Sediment was resuspended overnight (ON) in potassium phosphate buffer 0.5 M pH 7.5 with EDTA 10 mM. Solution was clarified by 10 minutes centrifugation at 9.000x g and viral particles were deposited with 2h ultracentrifugation at 80.000x g. The sediment was resuspended in potassium phosphate buffer 0.25 M pH 7.5, EDTA 10 mM and CsCl was added to a final density of 1.27 g/cm^3. Resulting solution was centrifuged at 150.000x g for 18h at 4°C. The visible band of the gradient which included the particles was recovered perforating the tube with syringe and needle. It was diluted in potassium phosphate buffer 0.25 M pH 7.5, EDTA 10 mM and sedimented by 2h centrifugation at 80.000x g. Finally, sediment was resuspended in 50% glycerol, Tris-HCl 5 mM pH 7.5, EDTA 5 mM and stored at -20°C

3.1.2 PVY VLPs

Escherichia coli cells were transformed with Pt7-7 vector (Addgene) containing the gene fragment encoding PVY CP [99]. This construct was kindly handed by Dr. Marjetka Podobnik from the National Institute of Chemistry in Ljubljana, Slovenia. *E.coli* cells were grown to an OD_{600} of 0.6-1.0 in LB buffer. Protein expression was induced using 0.2 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) for 20h at 20°C. Media culture was centrifuged in a Beckman JLA-8.1000 rotor at 8.000 rpm for 30 minutes. Resulting pellet was resuspended and sonicated with PBS buffer (1.8 mM KH_2PO_4; 10.1 mM Na_2HPO_4; 140 mM NaCl and 2.7 mM KCl). Sonication product was centrifuged in a Beckman Type 70.Ti Rotor at 30.000 rpm for 30 minutes. Lysate was incubated for 20 minutes in a mixture of 4% PEG 8000 and 0.5 M NaCl for VLP precipitation. VLP suspension was centrifuged for 20 minutes at 14.000x g and the resulting pellet was resuspended ON in PBS buffer. Resulting sample was centrifuged at 35.000x g for 20 minutes in order to discard remaining solid materials. Supernatant was loaded onto a 15%-60% sucrose gradient and centrifuged for 6h at 117.000x g in a swinging rotor SW40. After centrifugation, fractions of the gradient were analysed by SDS-PAGE to identify fractions containing VLPs. Those fractions were collected and stored at 4°C.

3.1.3 WMV virions

WMV virions were purified directly from infected leaves. Inoculation of the virions was made by the group of Dr. Miguel Aranda at CEBAS-CSIC in Murcia. *Cucurbita pepo* young leaves were inoculated through abrasion with carborundum with a homogenizate made up of dry leaves infected with WMV in a sodium phosphate buffer 30 mM pH 8.0. Inoculated plants grew for 4 weeks and young leaves presenting symptoms of infection were collected for purification. The purification process was made by me at the CEBAS-CSIC facilities.

For virion purification, leaves were blended manually in a mortar with the help of liquid nitrogen. 5.38 mL/gr of homogenizer buffer solution (0.5 M K_2HPO_4; 5 mM EDTA; 10 mM DIECA; 20 mM Na_2SO_3) was added to the resulting leave dust and mixed for 15 minutes at 4°C. Then, resulting solution was centrifuged at 7.500x g for 10 minutes and the supernatant was filtrated through five layers of muslin. 1% w/v of Triton X-100 was added to the filtrate and mixed for 1h at 4°C. The virions in solution were precipitated by ultracentrifugation at 300.000x g for 90 minutes at 4°C. The supernatant was discarded, and the pellet was resuspended with citrate solution pH 7.5 (sodium citrate 0.05 M; Na_2SO_3 20 mM; Citric acid 0.25 mM). 1% (w/v) Triton X-100 was added and the resuspension was mixed ON at 4°C. Following day, the resulting solution was centrifuged at 9.000x g for 10 minutes. Supernatant was collected and stored at 4°C. The pellet was again resuspended with citrate solution pH 7.5, added 1% w/v of Triton X-100 and mixed for 10 minutes at 4°C. Once again, the solution was centrifuged at 9.000x g for 10 minutes. Supernatant was collected and mixed with the supernatant stored before at 4°C. 10% w/v chloroform was added to the supernatant and mixed gently. Solution was centrifuged at 15.000x g for 10 minutes. The aqueous phase, the one containing the virions, was collected and centrifuged for 1h and 50 minutes at 245.000x g with 5 mL of 30% sucrose cushion. Resulting pellet was resuspended at 4°C ON in 1 ml of citrate solution and resulting solution was centrifuged for 10 minutes at 9.000x g. 0.26 gr/ml of Cs_2SO_4 were added to the supernatant for creating a density gradient in the tube. Gradient was centrifuged at 28.200 rpm in a swinging rotor SW40 for 30h and 30 minutes at 4°C. WMV virions were concentrated on a dense band in the gradient, which was collected with a fractionator. The fraction containing the virion sample was diluted with citrate solution and centrifuged at 108.000x g for 90 minutes at 4°C. Resulting pellet was resuspended ON at 4°C in 200 µl of citrate solution and later stored at 4°C until further use.

3.2 PROTEIN PURIFICATION METHODS

3.2.1 Purification of viral CPs labelled with 15N

For NMR studies, proteins must be labelled with a radioactive isotope as 15N. For this purpose, *E.coli* cells were grown in what is called minimal media (M9). Minimal media is a culture media which contains the indispensable minimal nutrients.

E.coli cells were transformed with pET-28a(+)-TEV vectors containing the gene fragments encoding PepMV CP, PepMV R124Q, WMV CP 71-247, WMV CP R172A and CTV CP. pET-28a(+)-TEV vectors were bought from GenScript and carry a N-terminal His-tag and a T7 promoter. Vector containing the gene for PepMV was handed to us by Dr. Miguel Aranda from CEBAS-CSIS. This vector is a pET His6 TEV LIC (2-BT) vector which also has a His-Tag at the N-terminal and a T7 promoter. Cells were grown in minimal media until an OD_{600} of 0.1. After that, for 2 L of culture we added 200 mL of minimal media: 0.1 mM $CaCl_2$; 1 mM Mg_2SO_4; 10 mg/L thiamine; 10 mg/L biotin; 100µg/mL ampicillin; 3 gr/L glucose; 1 gr/L of ammonium chloride labelled with N15 and 4 mL of trace elements. Culture was grown at 37°C until an OD of 0.6-0.7. Protein expression was induced using 2 mL of IPTG and left ON at 20°C. For protein purification, media culture was centrifuged in a Beckman JLA-8.1000 rotor at 8.000 rpm for 15 minutes. Resulting pellet was resuspended with lysis buffer (50 mM TRIS-HCl; 150 mM NaCl; 2 mM $MgCl_2$; RNAse) and sonicated for rupturing the cells. The lysate was clarified by centrifugation in a Beckman Type 70.Ti Rotor at 30.000 rpm for 30 minutes. Resulting supernatant was analysed by SDS-PAGE to check the presence of the protein of interest.

All proteins were affinity purified using a His-Trap column (GE Healthcare) followed by a size exclusion chromatography using a Superdex 75 10/300 column (GE Healthcare) in buffer 20 mM Tris-HCl, 150 mM NaCl pH 8.0 and 1 ml DTT. Purified proteins were stored at -20°C until use.

3.3 SAMPLE PREPARATION FOR CRYOEM AND CRYOET

3.3.1 Negative staining

Negative staining method was used for visualizing our samples to determine their concentration and morphology in the TEM. Samples were prepared upon copper grids coated with a thin layer of carbon. Grids were subjected to an ionic discharge process were the surface of the grid was hydrophilized in order to allow uniform spread of aqueous solutions. After hydrophilization, grids were placed on top of a 10 µL drop of sample and incubated for 1 minute. Grids were washed with 10 µL of mili-Q water and sample excess was cleared from the grid with filter paper. Finally, grids were dyed with uranyl acetate at 2% for 45 seconds and dried at room temperature (RT).

For grid visualization, we used in house 120kV JEOL-1230 (JEOL) transmission electron microscope operated at 100kV, equipped with an Orius SC1000 cooled slow-scan CCD camera (GATAN).

3.3.2 Grid preparation for cryoEM and cryoET

Vitrification process was in general the same for all samples with small differences between them. For vitrification, Quantifoil 2/2 holey carbon grids coated with a thin layer of carbon were used.

3.3.2.1 Carbon film preparation

The layer of carbon upon the grid helps the sample to distribute better upon the grid. For making the carbon film, carbon is evaporated over a piece of mica in a BAL-TEC MED 020 high-vacuum coating system. Later the carbon is floated in water and fished with the grids (Fig 3.1) resulting in a holey carbon film that has been sandwiched between the grid and the thin continuous carbon film.

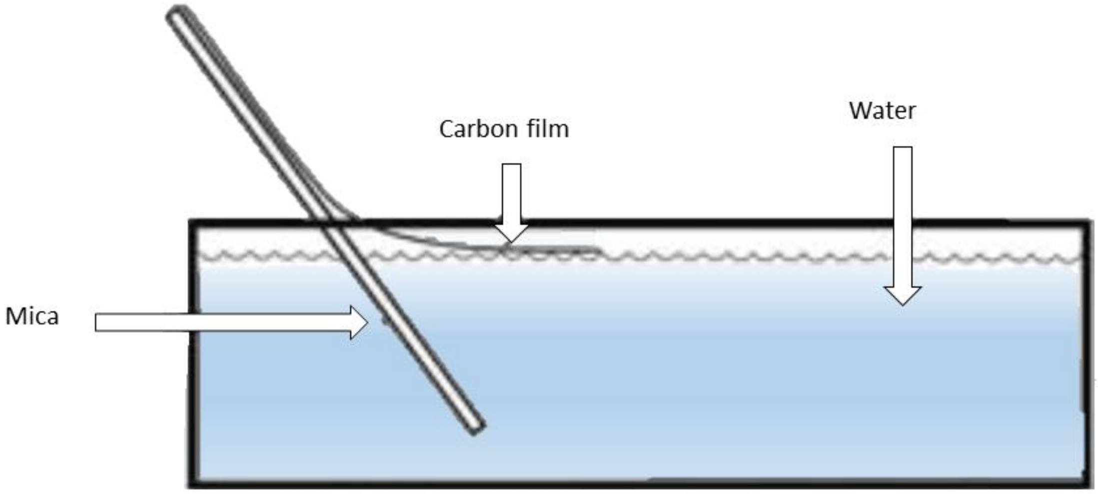


Fig 3.1 Carbon film fishing. The scheme shows hoy the thin layer of carbon upon the mica surface is placed in water.

3.3.2.2 Vitrification

All grids were vitrified using a Vitrobot (FEI). We applied a second carbon layer to all the samples. This extra layer helps to evenly distribute the filaments and the aqueous solutions on the grids.

For TuMV and VLPs we applied 4 µL of sample to the grids with an incubation time of 1 minute. After incubation, excess of sample was removed from the grids with a filter paper and rapidly plunged in liquid ethane at -180°C.

For tomography, gold beads were added to the sample for their use as fiducials in tomogram reconstruction. We used the BSA Gold Tracer, 6nm EM Grade (Electron Microscopy Sciences) just before vitrifying in a 1:4 ratio of virion sample and beads, respectively. 4 µL of the mixture were used for each grid.

For PVY VLPs in the presence of a selected chemical compound we tried three different strategies for vitrification. The compound (25.D8) had to be diluted in dimethyl sulfoxide (DMSO). DMSO is a cryoprotector and it interferes with the vitrification, so we needed to reduce the percentage of DMSO in the sample without compromising the amount of compound present. 85 µl of PVY VLPs were incubated with 15 µl of compound (2.3 mM) ON at 4°C. First approach was a serial dilution of the sample. Grids were incubated for 30 seconds in drops containing different solutions. The first one contained the PVY VLP + 25.D8 sample and the three remaining drops were only buffer + 25.D8 (without DMSO). Second approach was doing a dialysis in a petri dish, where 50 µL of the sample was placed upon a Whatman filter paper (Schleicher & Schuell) floating on a solution free of DMSO for 40 min. Third approach was a dialysis ON in a dialysis bag ON at 4°C. In the three type of samples the vitrification was performed as after an incubation time of 30 seconds on the grids. Dialysis in petri dish with a filter membrane resulted to be the method that gave us the better looking filaments (Fig 3.2). Filaments were not broken as in the other methods and their distribution was right for data collection.

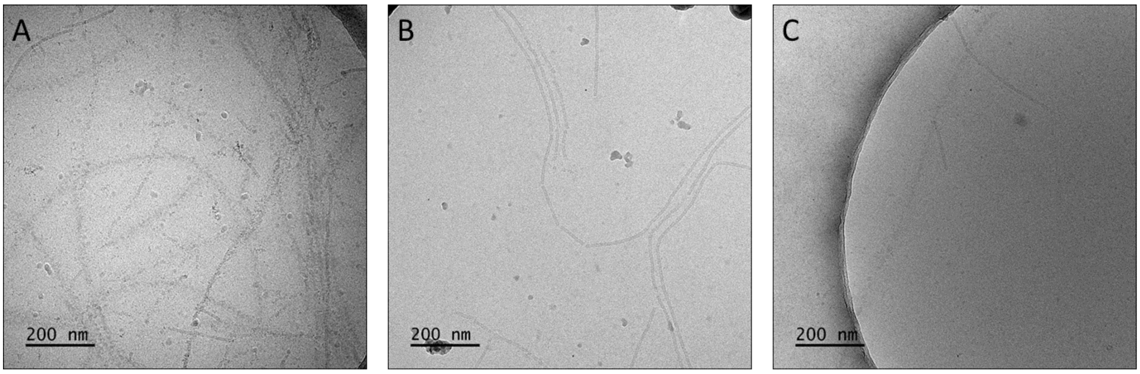


Fig 3.2 Electron micrographs for the three strategies of vitrification. A) First approach showing the result of a serial dilution of the sample, filaments appear broken and poorly defined **B)** Second approach, dialysis in petri dish where filaments present good definition and distribution **C)** Third approach, dialysis in dialysis bag where very few filaments can be seen.

3.4 DATA COLLECTION OF CRYOEM AND CRYOET IMAGES

We checked our vitrified samples in a 200kV JEM-2200FS/CR (JEOL) transmission electron microscope, equipped with an UltraScan 4000 SP cooled slow-scan CCD camera (GATAN).

For cryoEM high-resolution data collection, we sent our samples to “The Netherlands Centre for Electron Nanoscopy” (NeCEN). The microscope used was a Titan Krios FEI electron microscope with 300kV accelerating voltage using a K2 direct detector camera (DDC) (GATAN).

For cryoET data collection, samples were sent to “Centro Nacional de Biotecnología” (CNB-CSIC). The microscope used was a Talos Artica electron microscope operated at 200kV using a Falcon III direct detector camera. Tilt series were collected with a tilt range from -60° to +60°, with an angular step size of 3 degrees, starting from 0 to -60° and then coming back to +3° and going from +3° to +60°.

All data collections were performed using EPU software for automated data acquisition from Thermo Fisher Scientific.

3.5 IMAGE PROCESSING OF CRYOEM DATA

3.5.1 Motion correction

Aligning the individual frames of movies is necessary to correct beam induced image blurring and restore important high resolution information. For that reason, each image is recorded into several frames or movies so that the motion between frames can be corrected. MotionCorr2 program [104] and the Relion’s own implementation of the MotionCorr2 [105] was used to align the frames for correcting the drift and beam-induced movement.

3.5.2 CTF estimation

The CTF is the function in reciprocal space that modulates the signal recorded in the electron microscope that depends on several parameters such as the defocus setting, the accelerating voltage, spherical aberration, amplitude contrast and astigmatism. All these parameters modulate the image and these modulations need to be corrected if we want to restore the correct structural information of the sample. Defocus and astigmatism do not depend on the microscope and are the only values that vary for each individual image taken under the same conditions. The CTF is manifested in the power spectra of the images as a series of concentric black and white circles named “Thon rings”. These circles represent the information available in the sample, the black ones do not have any information and they are called “zeros”. For this reason, images are collected at different defocus values, in order to cover all the frequencies. There are several software packages that can estimate the CTF of the images and we used CTFFIND4 [106].

3.5.3 Particle picking

Once we have our motion corrected images and after CTF estimation, it's time to pick and extract the particles from the micrographs. In our case, we followed helical processing in Relion, which is based in single particle method but implemented for helical samples [89]. Relion is a stand-alone computer program for refinement of 3D reconstructions or 2D class averages in cryoEM microscopy. It provides a GUI (guided user interface) which makes it very approachable and user friendly. Once the movie frames are aligned all the micrographs are imported to Relion. Regions of straight filaments were selected, and overlapping segments of the helical filaments were extracted from the micrographs as single particles. The amount of overlap between adjacent segmented boxes will define how many new asymmetrical units (X) are contained in each new box [89]. Helical specimens possess one critical advantage over other symmetries, because, in contrast to single particles, vitrified filaments do not assume random orientations and segments along the same helix keep relative orientations related to the helical parameters. Furthermore, helical filaments vary only around their azimuth angle with little out-of-plane tilt, and a single helix provides many different orientational views of its asymmetric unit.

3.5.4 2D Classification

The next step in the image processing is the classification of all the selected particles according to the similarity between them. 2D classification follows the maximum likelihood method. It is an iterative process in which all particles are aligned to a defined number of templates that are randomly selected from the particle set. Each particle is translated and rotated until it fits with one of the references. In each iteration, the next generation of templates is calculated by averaging the clustered particles. The 2D classification procedure results in the assignment of all the particles to different 2D averages. All particles that are grouped in such classes can then be aligned and averaged to create class averages. The calculated class averages can be used to assess the heterogeneity of the sample and to get rid of those particles that appear broken, blurred or do not display high-resolution features. What follows is a selection of the averages that present good features, and the creation of a selection file with the particles associated to those averages.

3.5.5 3D refinement of filaments

3D refinement follows the template matching procedure by iterating through cycles of projection, alignment and reconstruction until convergence. For 3D refinement it is necessary to give an initial reference map. If we know nothing about the shape and symmetry of our sample, then an initial 3D model must be created using Relion. If density maps of similar samples are available, then those can be used as references with a large low pass filter in order to avoid model bias. In the case of filaments, a featureless cylinder with the diameter of the filament can be used as initial reference. Relion uses projections of this initial model and compares them with the extracted experimental images. The critical step of 3D refinement is the determination of the orientation parameters of the particles. For filamentous particles with helical symmetry there might be some prior knowledge about the parameters of the helix (Euler angles; tilt, psi and rot). The Euler angles are three angles that describe the orientation of a rigid body with respect of a fixed coordinate system. Filaments tend to lie horizontally inside the ice layer so an angle of 90° for the tilt angle is used as default with a local search limited to an interval of +/- 15°. When particles are extracted from micrographs, the program assigns a rotational angle for the particles, which is used as starting point for search of the psi angle in an interval of, also, +/- 15°. For rot angle Relion does a search spanning all the space, from -180° to +180°. The symmetry parameters must be given for the reconstruction, as twist and rise along the z-axis of the helix, and the number of true unique asymmetrical units contained in each segment (number of new asymmetric units per step size during the segmentation). During 3D refinement we can perform local searches of symmetry. Flexible helices can vary their parameters of twist and rise along the filament. Giving the program initial search ranges and step size parameters allows to find variances between the averaged sections of the map and polish the helical parameters for the reconstruction.

During 3D-refinement, Relion separates the data set into two non-overlapping half sets with the same number of particles and performs two independent reconstructions. This method prevents the iterative build-up of noise getting rid of the noisy pixels that do not cross-correlate between both reconstructions. The two resulting 3D maps are correlated in reciprocal space shell by shell to yield a 1D curve of "Fourier Shell Correlation" (FSC) coefficients as a function of the modulus of spatial frequency. High values indicate consistent signal but those close to zero mean that there is no reliable signal. FSC is also used for the estimation of the resolution of the reconstructed map. Such a number is determined from the FSC curve as the spatial frequency where the curve drops below 0.143 threshold (gold-standard method) [107].

Postprocessing of the maps may be used to sharpen the refined maps and to calculate their true resolution after solvent masking. For this a tight mask is used, slightly larger than the map itself so that the features of the densities inside the mask are not affected by the solvent background. In this step, a b-factor is applied to the map. High-resolution data of the map appears underrepresented by effect of the CTF's envelope, which lowers the amplitudes at higher frequencies. The b-factor applied increases the weighting of the higher resolution data and applies a sharpening to the map.

3.5.6 3D classification

3D classification allows to explore the heterogeneity of the sample. It is based on the maximum likelihood approach. Here, multiple projections are reconstructed and refined simultaneously starting from a unique model. The user provides a number of classes and in the first iteration particles are assigned randomly to each class. The angular searches of the orientations can be performed exhaustive and coarse if we haven't got any knowledge about them, or fine and local if we come from a 3D-refinement where the angles have been already set. In these asymmetric classifications one can remove junk and bad particles. Suitable particles tend to group together in one class. Coarse angular sampling is preferred when we want to separate large conformational differences between the particles in the data set. For local and small conformational changes, one may want to do finer angular searches. Masked 3D classification is also an option if we want to focus on one specific part of our molecule.

3.5.7 Multibody refinement

Multibody refinement is a tool available in Relion that allows the characterization of continuous movements that cannot be described just as discrete conformational changes in 3D classification [108]. A consensus structure is divided into a discrete number of individual moving bodies that are masked during the refinement. The program subtracts the signal outside each mask and each single body will be aligned with projections of itself. This approach improves the alignment of regions that display relative movements and provides information about their motions[109].

3.5.8 Local Resolution

Although classification separates the populations of particles that have changes in the structure, the final refined map may still have some structural heterogeneity that can lead to local variations in resolution. Local Resolution estimates the resolution variations along the 3D density map. Local Resolution estimation was done within Relion software. Relion uses an own implementation of the local resolution software ResMap [110].

3.6 IMAGE PROCESSING OF CRYOET DATA

3.6.1 Motion Correction

In cryoET, data acquisition suffers from the same technical problems as in cryoEM data collection but with some particularities. Frames acquired at higher tilts present large variations on their optical parameters, as their defocus values, what leads to variations on their CTFs [111]. Moreover, the drifting between frames is much bigger than in cryoEM since in tomography the sample stage is physically moving. Besides, the SNR in tomography is way smaller because the dose per frame used is significantly lower than in high-resolution cryoEM. Motion correction of the tilt series was done using MotionCorr2 [112].

3.6.2 Tomogram reconstruction

For tomogram reconstruction we need to align all the 2D tilt series acquired at different angles in order to make a 3D volume. There are several programs that can perform this task, and we used IMOD software. This package contains a set of programs for tomographic reconstruction and a graphical user interface (etomo) that manages the whole process [113].

Tomographic reconstruction starts with a pre-processing step. X-rays can hit the CCD camera during collection and this cause extreme pixel values that can affect the contrast and reconstruction of the tomogram. These pixels are erased with a program called CCD eraser. After this step we perform the first coarse alignment of the reconstruction. Tiltxcorr program uses cross-correlation to determine X and Y translations on each tilt in order to align all of them and produce an aligned stack. Once we have a pre-aligned tomogram, we can go a little bit further and create a fiducial model. Here, gold beads are selected as fiducials for image alignment through the tilts. First, gold beads are selected at the first image acquired, in our case at 0°, and a seed model is created. The seed model can be created manually, selecting the gold beads one by one, or automatically, where the program Autofidseed selects a well-distributed set of fiducials. Beadtrack program finds the beads in all other sections. Usually this computer-generated model is not perfect, and we need to fill the gaps or fix the errors manually. It is important to make sure that the large majority of the fiducials are tracked all the way to the two ends of the tilt series. Next step entails an iterative process of fixing the fiducial model. Here, the prealigned stack and the fiducial model file are opened in 3dmod and gaps in the trajectories are fixed manually. Once the fiducial model is generated, tracked and fixed, then tilt images can be aligned in a smoother way. The fine alignment is done by the program Tiltalign. This program uses the position of the gold beads to find the best fit. In this step, IMOD log window will display a value of residual error mean. If the error is too large, we can fix it by going back to the reconstruction and fix the fiducial points in an iterative mode until the error of the alignment does not improve.

The next step is the tomogram generation, where the goal is to reconstruct the 3D data that contains our sample in the smallest possible volume. Here we can navigate our tomogram in the X, Y and Z axis and manually select the top and bottom boundaries of our tomograms, called the model contours. Now the tomogram thickness can be set so that it contains all of our sample. Lastly, we can create the final aligned stack. In this case, we binned the tomograms by a factor of 2, resulting in augmenting the pixel size from 1.83 Å to 3.66 Å. Finally, we generated the full tomogram. We used the weighted back-projection (WBP) reconstruction algorithm [114].

3.6.3 Subtomogram averaging

3D particles can be extracted from tomograms and averaged together for particle reconstruction as in single particle cryoEM in a process called subtomogram averaging. For this aim we used Relion software following a protocol for subtomogram averaging by Scheres and Bharat [115]. Our tomograms generated in IMOD are in ".rec" format, which is the image format that IMOD handles. Tomograms must be converted into ".mrc" format, the one that Relion can read. Moreover, sub-tomogram coordinates must be given. For this purpose, we opened our tomograms in the 3dmod visualization software and selected our particles as individual models. This created a file with the coordinates of the particles in the tomogram. Then we converted the models into text formats with the model2point command, with ".coords" suffix. The aligned tilt series are also essential for CTF estimation, 3D alignment and classification, they must be saved in an individual file for each tomogram with the format ".mrcs". The final tilt series angles used have to be in a separate text file with the suffix ".tlt". For each aligned tilt series, we have to create a text file that list the tilt angles and the accumulated radiation for each image in the tilt series. This file has the suffix ".order".

3.6.3.1 CTF estimation

The step of CTF estimation and CTF correction is as important as for cryoEM. It depends on the defocus values for each image of the tilt series. Here we used a python script written by Bharat and Scheres, where CTFFIND2 [116] is called and all the tomograms are CTF estimated. Each CTF is weighted by the accumulated dose of each tilt. In the end, this step will create files with the corresponding 3D CTF model for each subtomogram.

3.6.3.2 Particle extraction

All the image processing from now on can be carried out in the Relion GUI. Particle extraction is performed as for classical single particle cryoEM data. The only difference here is that the extracted particles are 3D volumes, not 2D projections.

3.6.3.3 3D refinement and classification

The steps of 3D refinement and classification are performed exactly as for single particle based processing.

3.7 ATOMIC MODEL BUILDING

Atomic model building is critical for the understanding and interpretation of the high resolution density maps obtained from cryoEM image processing. We only could model the CP for TuMV virions as is the only map that reached near-atomic resolution. For this purpose we extracted the density for the CP from the whole virion density map following the Segger method [117] in the visualization program Chimera [118]. The segmentation allows us to extract individual parts from the whole density map. The atomic model structure for WMV CP was used as a template. The amino acid sequence of WMV CP was mutated manually using Coot [119] to match the sequence for TuMV CP. Further modelling of the atomic structure was carried out manually using Coot and the stereochemistry of the model was improved by real-space refinement in Phenix [120]. For the ssRNA, a modelled polyU was included in the refinement. To build up a multimer of CPs and the ssRNA a final refinement in Phenix was performed using noncrystallographic symmetry. Evaluation of the modelled atomic structure for TuMV CP was carried out in MolProbity [121]. To calculate the surface electrostatic potential for the atomic structure of TuMV CP, the atomic coordinates were loaded in Bluues server [122] ant the potential was determined based on generalized Born radii [123].

Chapter 4:

RESULTS

4.1 STRUCTURE OF TuMV AND ITS VLPs BY CRYO-EM

4.1.1 TuMV and VLPs image collection and processing

TuMV is a virus of the family *Potyviridae* that causes diseases in cruciferous plants, among others. At the time this work started, the structure for WMV was the only potyviral structure available at high resolution [98]. This structure, along with a potexvirus, PepMV, structure [31], brought to light a conserved region in the a binding pocket where the coat protein interacts with the ssRNA. The finding of this conserved region could open the doors to many bio and nano-technological applications for plant viruses. Moreover, for both virions, it was elucidated that the N-terminal arm acted as a bridge between adjacent subunits. In PepMV the N-terminal arm only supports side-by side contacts, but in WMV the N-terminal arm bridges the adjacent subunit and, by a 90° turn, it reaches a second subunit in the next turn of the helix. In this work we explored not only if TuMV virions share the same conserved region as other filamentous plant viruses at their binding pocket, but also if the configuration of the N-terminal arm was a structural characteristic of potyviral family. Parallel to our investigation, the structure for another potyvirus, PVY, was resolved at atomic level [99]. PVY presented the three conserved residues at its binding pocket and the N-terminal was acting exactly as in WMV virions, supporting the idea of a conserved characteristic along all potyviruses. Moreover, it is of high importance to know the role of the ssRNA in the maintenance of virion structure and stability. For these reasons, we characterized the structure of TuMV, and also its VLPs by cryoEM.

TuMV virions were extracted from Indian mustard plants (*Brassica juncea)* inoculated with an extract containing viral particles for 21 days approximately. For VLPs purification, *Nicotiana benthamiana* plants were agroinfiltrated with *Agrobacterium tumefaciens* constructs expressing TuMV CP. After several weeks of plant growth, young leaves presenting symptoms of infection were harvested and processed in order to purify TuMV virions, and for VLPs all leaves were selected. Young leaves haven't yet developed their defence mechanisms against external pathogens, as the production of jasmonic acid or silicic acid. For viral infection, the young stage of plants is the most susceptible and the maximum amount of virions will be present in this phase.

For both cases, leaves were homogenized, filtrated and centrifuged to discard large cellular products. Following steps of ultracentrifugation with organic and inorganic agents final isolation in cesium chloride (CsCl) density gradients resulted in two samples of 0.9 mg/ml of TuMV virions and 1.3 mg/ml of TuMV VLPs. The whole inoculation and purification processes were carried out by the goup of Dr. Fernando Ponz in the Center of Biotechnology and Plant Genomics (CBGP).

Negative stained grids with different dilutions of the purified virions and VLPs were prepared for their visualization in a 120 kV in house transmission electron microscope (TEM) although images were taken at 100 kV. This step is necessary in order to know if the concentration of virions in solution is suitable for the image acquisition in a cryo-electron microscope. The samples showed a good distribution and concentration as well as homogeneity, with few or no contaminants. The VLPs seemed to be more variable in length than TuMV virions (Fig 4.1).

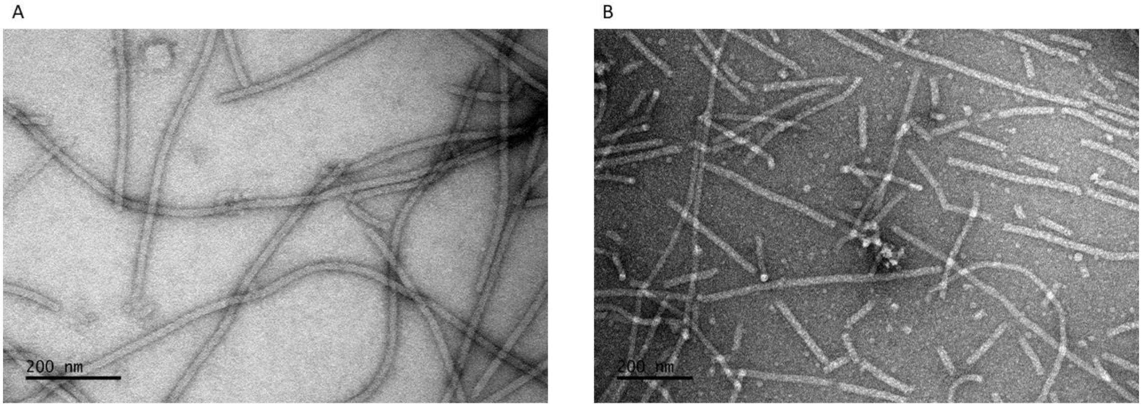


Fig 4.1 Electron micrographs of **negatively stained**. **A)** TuMV virions and **B)** TuMV VLPs.

For cryoEM imaging, 3-4 µl of the viral samples were placed on Quantifoil 2/2 holey carbon grids coated with a thin layer of carbon. This method has proven to have positive effects on sample distribution in other filamentous viruses [31]. Movie frames images were collected in a Titan Krios FEI electron microscope operated at 300kV using a K2 direct detector camera (GATAN) at “The Netherlands Centre for Electron Nanoscopy” (NeCEN). Images were taken at a nominal magnification of x130,000 for both TuMV and VLPs, with a corresponding pixel size of 1.1 Å. Micrographs were taken with time exposures of 9s in electron counting mode resulting in movies with 40 frames each and a total dose of 40 e^{-}/ $Å^{2}$. Motion between frames was corrected using Relion's own implementation of MotionCor2-like algorithms. Dose weighting was not used and we selected frames 3 to 31, resulting in accumulative dose of 31 e^{-}/ $Å^{2}$. 1265 and 1259 micrographs were obtained for TuMV virions and VLPs respectively. Images were motion corrected and their contrast transfer function was estimated using CTFFIND3 [116]. Filaments were manually selected resulting in 444,678 overlapped segments for TuMV virions and 307,333 for TuMV VLPs with a step size of 2 asymmetric units (8 Å) and images of a size of 200 pixels. Image processing of the cryoEM data was done using Relion2 [89]. For helical processing, Relion2 uses the single particle technique but taking into account the helical parameters of the sample like the twist and rise. In this case, as initial parameters, we used the helical parameters of watermelon mosaic virus (WMV) [98] as it is also a potyvirus and as it has been proposed in previous studies [98–100], potyviruses may have not only very similar helical parameters but also their CPs may have strong structural homology. Furthermore, Relion does twist and rise searches for local optimization of those parameters. Picked helical segments were used to do a 2D classification. In this step we can see the helical "skeleton" of our samples. This method creates groups of particles with similar features, creating an average of particles in each group. Overall, this method allows us to separate real particles from noise or particles that are damaged. We selected the groups that exhibited good characteristics and discarded the bad ones. That left us with 335,228 segments of TuMV virions and 118,597 of VLPs for 3D refinement (Fig 4.2).

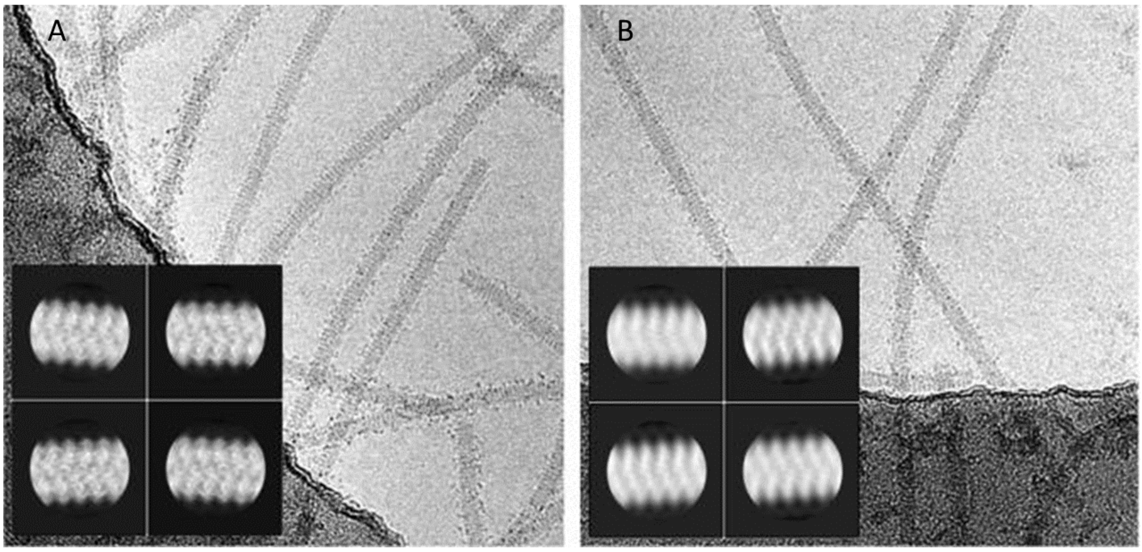


Fig 4.2 Electron micrographs corresponding to A) TuMV virions and B) TuMV VLPs. The black box at the left down corner of each image shows four of the selected classes from the 25 classes calculated in the 2D classification.

The following step was the 3D reconstruction of those projections. Relion allows us to do a 3D alignment of all the particles (3D refinement) or a 3D classification, where it aligns the particles and separates them into a given number of groups depending on local differences of the particles. For 3D alignment an initial reference must be given. Our initial reference for TuMV virions was a featureless cylinder with the diameter of the filaments and for TuMV VLPs the density map for TuMV virions low-pass filtered at 40 Å. In the case of TuMV virions, starting with such featureless cylinder requires to re-run the refinement using as reference the map obtained in the previous round. Helical symmetry was imposed during the 3D refinements and also to the final 3D cryoEM maps.

4.1.2 Analysis of TuMV virions density maps

The cryoEM 3D map for TuMV virions shows a left-handed helix with the coat protein arranged in an helical manner identical to that of earlier characterized flexible filamentous plant viruses [31,98–100]. As WMV virions, TuMV ones are composed of 8.8

subunits per helical turn. The helix has a diameter of 135 Å with a pitch of 35.2 Å and a twist of -40.9° (Fig 4.3).

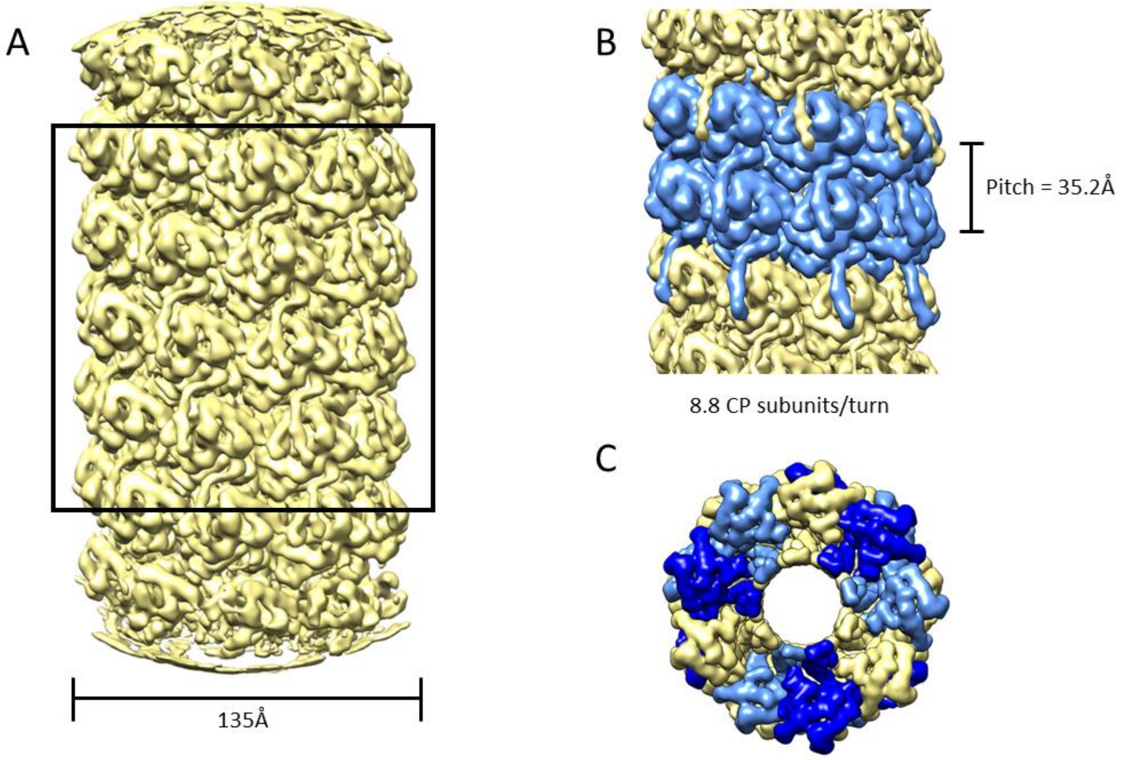


Fig 4.3 3D density map corresponding to TuMV virions A) 3D reconstruction of a segment of the whole virion showing the arrangement of CPs along the filament, which has a diameter of 135 Å. **B)** Closer view of a central section of the filament displaying the pitch between turns of the helix and the number of CPs per turn. **C)** Top view of the 3D density map with the CPs arranged in a helical manner.

Since helical values may vary in the places were the filaments bend, helical segments that are not perfectly straight can led to inaccuracies in the reconstruction. As a flexible helix can only be approximated with a straight helix, Relion uses the central part of the filaments (around 30%) because the top and bottoms parts are more blurred than the central part [89]. Postprocessing of our map gave us a resolution of 5.18 Å, calculated following the gold standard method in which the cutoff of the FSC curve is at 0.143 (Fig 4.4, B). Local resolution calculations (Fig 4.4, A) shows that our map was very uniform and almost all of it had values around 5.2 Å.

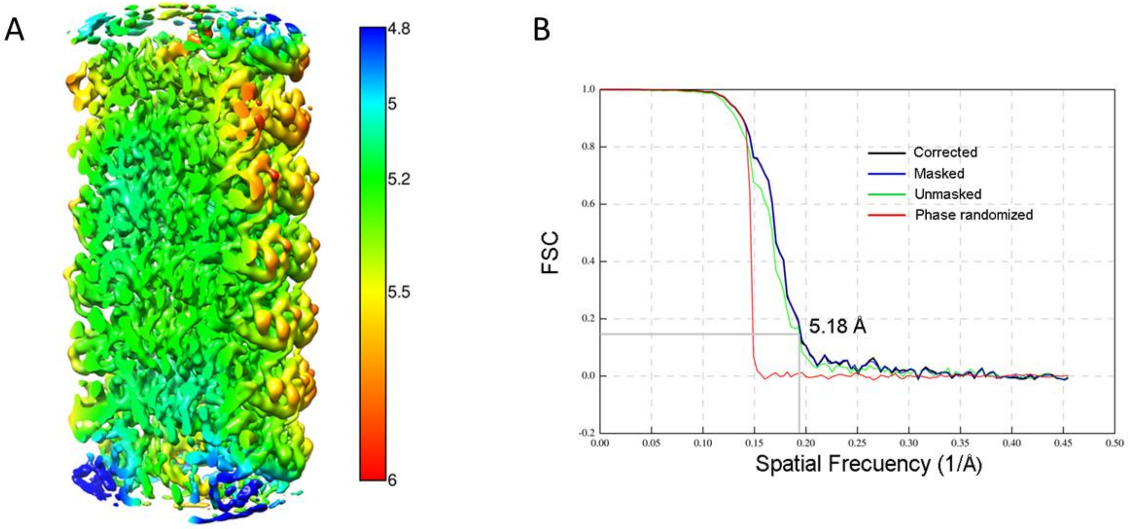
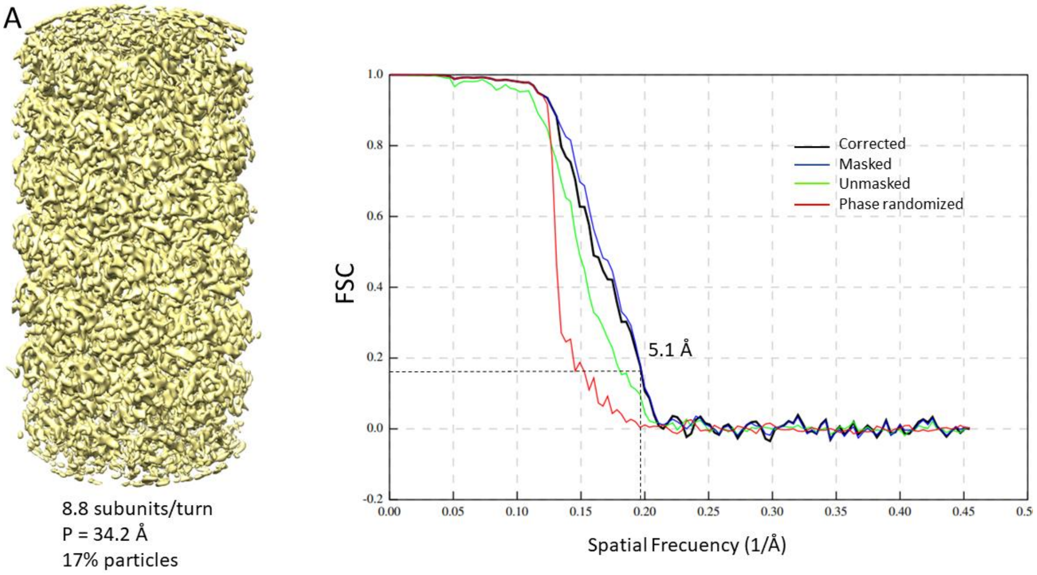


Fig 4.4 3D density map and FSC curve displaying resolution of TuMV maps. **A)** Local Resolution density map showing resolution values along the filament. **B)** FSC curve with cutoff at 0.143.

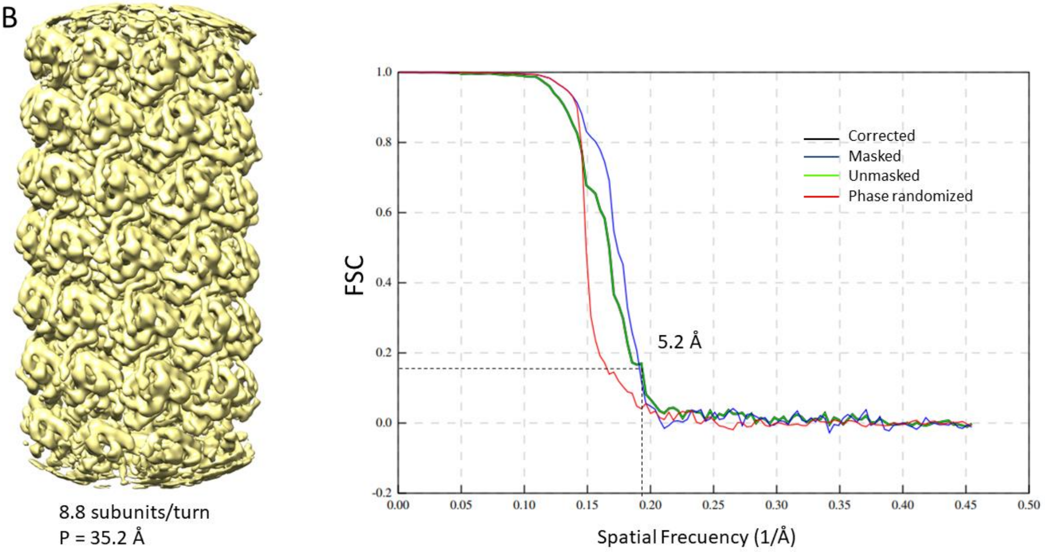
Later, unsupervised 3D classification of the total dataset gave us three classes, that after completing the process of 3D refinement and post-processing of each of them revealed that regions of the filaments stretch and shrink with an amplitude of around 2Å per turn. This feature could be noticed looking to the parameters of pitch, which went from 34.2 Å for the first class to 36.1 Å for the last one, showing the intrinsic flexibility of the virions (Fig 4.5). Between the three classes, two of them had few particles each one, and although their resolution values were good, the maps weren't well defined and had a lot of noise (Fig 4.5, A, C). The second one was the predominant class with the 58% of the particles. It was less noisy and better defined so we selected this class for further analysis of the coat protein (Fig 4.5, B).

A



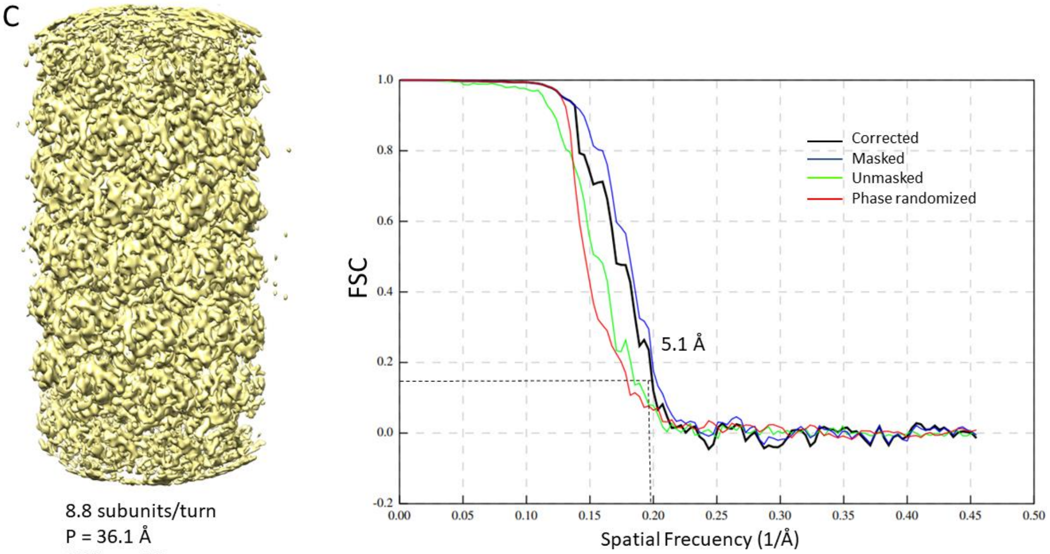
8.8 subunits/turn
P = 34.2 Å
17% particles

B



8.8 subunits/turn
P = 35.2 Å
58% particles

C



8.8 subunits/turn
P = 36.1 Å
25% particles

Fig 4.5 3D density maps corresponding to 3D classification of TuMV virions. Images on the left (A, B and C) correspond to the density maps for the three classes. They show how the filament changes its pitch value in 2Å between the **A** and **C** maps, how the subunits per turn stays the same and the distribution of the particles on each class. The images on the right for the three maps display the FSC curves with the cutoff at 0.143.

4.1.3 CP atomic model building and analysis of CPs and ssRNA interactions

We used the 3D map for the most populated group of the 3D classification for the calculation of the atomic model for TuMV CP. 3D fold of the CPs from flexible filamentous viruses of different families is highly conserved despite the absence of sequence homology between them. Within potyviruses the known CP structures for WMV and PVY [98,99] are almost identical.

The CP from TuMV shows high sequence conservation with these both CPs so we expect the structure of TuMV to be alike to the other two potyviral structures. Actually, the atomic models for WMV and PVY CPs superimpose in almost full agreement with the cryoEM density map for TuMV virions (Fig 4.6).

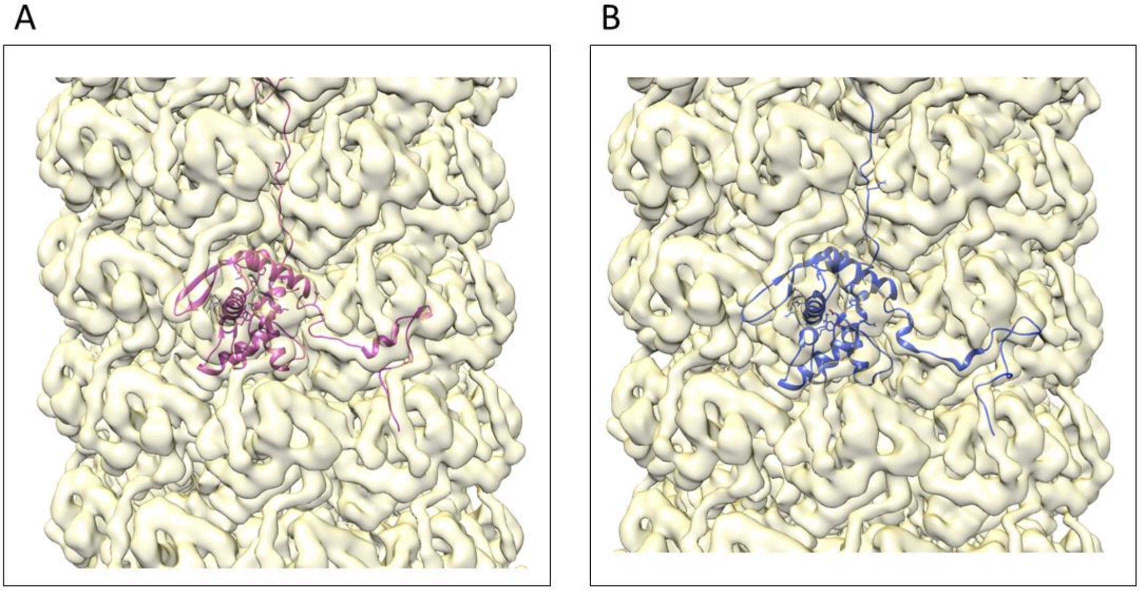
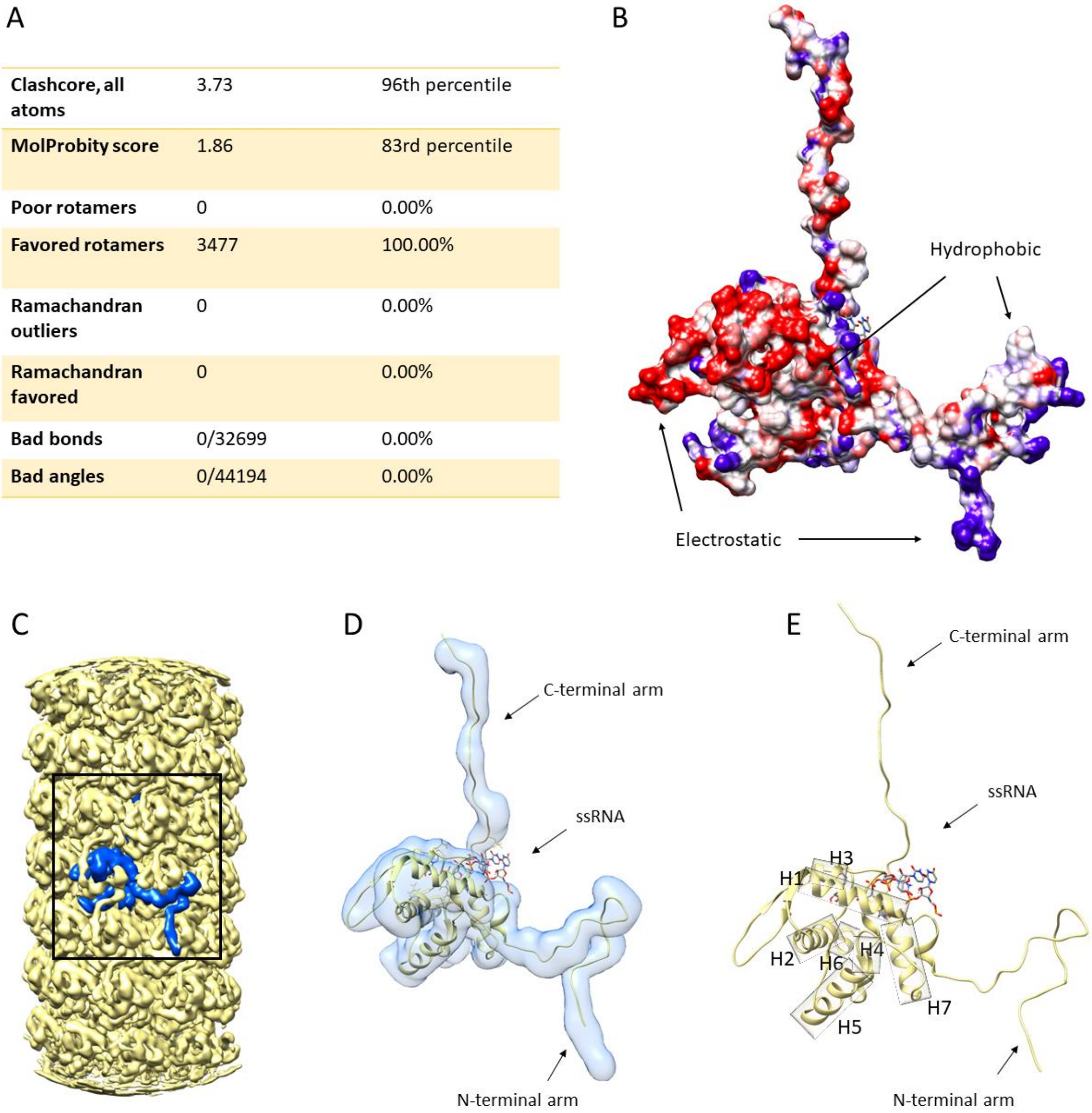


Fig 4.6 3D density map for TuMV virions with fitted atomic models corresponding to A) WMV and B) PVY, showing how they fit in the density almost perfectly.

Even though our cryoEM map for TuMV is limited to 5Å resolution, the high sequence homology and structural conservation allowed us to build an accurate atomic model for TuMV CP (Fig 4.7, E) based on the structure for WMV CP. The atomic coordinates for TuMV CP show a central alpha-helical core composed of seven α-helices and two long terminal arms (Fig 4.7, B, C). The cryoEM map does not show density for the first 65 amino acids at the N-terminal end, a flexible region exposed to the solvent (Fig 4.7, B). The last 16 residues at the C-terminus cannot be traced.



Clashcore, all atoms	3.73	96th percentile
MolProbity score	1.86	83rd percentile
Poor rotamers	0	0.00%
Favored rotamers	3477	100.00%
Ramachandran outliers	0	0.00%
Ramachandran favored	0	0.00%
Bad bonds	0/32699	0.00%
Bad angles	0/44194	0.00%

Fig 4.7 Coat protein atomic model analysis. A) Parameters that evaluate the stereochemistry of the atomic model for the TuMV CP obtained in MolProbity. **B)** Coat protein surface coloured by electrostatic potential. Regions that participate in electrostatic or hydrophobic CP-CP and CP-

RNA interactions are shown. **C)** 3D density map for TuMV virion in yellow, highlighting inside the box a single coat protein in blue. **D)** Atomic model for TuMV CP fitted inside the density map where the C and N-terminals are indicated, as well as the ssRNA. **E)** Atomic model for TuMV CP presenting the seven helices (H1 to H7) that form the helical core of the protein.

As shown before, the participation of flexible N- and C-terminal arms in the interaction between CP subunits is the structural basis for the flexible nature of the virions. The N-terminal of each CP interacts with other two subunits. It bridges the adjacent subunit and, by a 90˚turn, reaches a second subunit in the next turn of the helix. On the one hand, side-by-side interaction is mediated via hydrophobic interactions between the N-terminal arm and a groove in the adjacent subunit (Fig 4.7, B) and on the other hand interaction with the subunit in the next run of the helix is favoured by complementary electrostatic potentials (Fig 4.7, B). The dual role for the N-terminal arm that supports side and axial polymerization and the nature of the local interactions were also observed for WMV and PVY and seem to be a signature for potyviruses [98,99].

The density for the RNA is clear and each CP spans five ribonucleotides (Fig 4.8 A, B). It stands in a groove at the folded central domain, just next to the H7 helix, and the RNA binding site of TuMV CP includes the universally conserved pocket in flexible filamentous plant viruses formed by amino acids Ser, Arg and Asp (Fig 4.8, C).

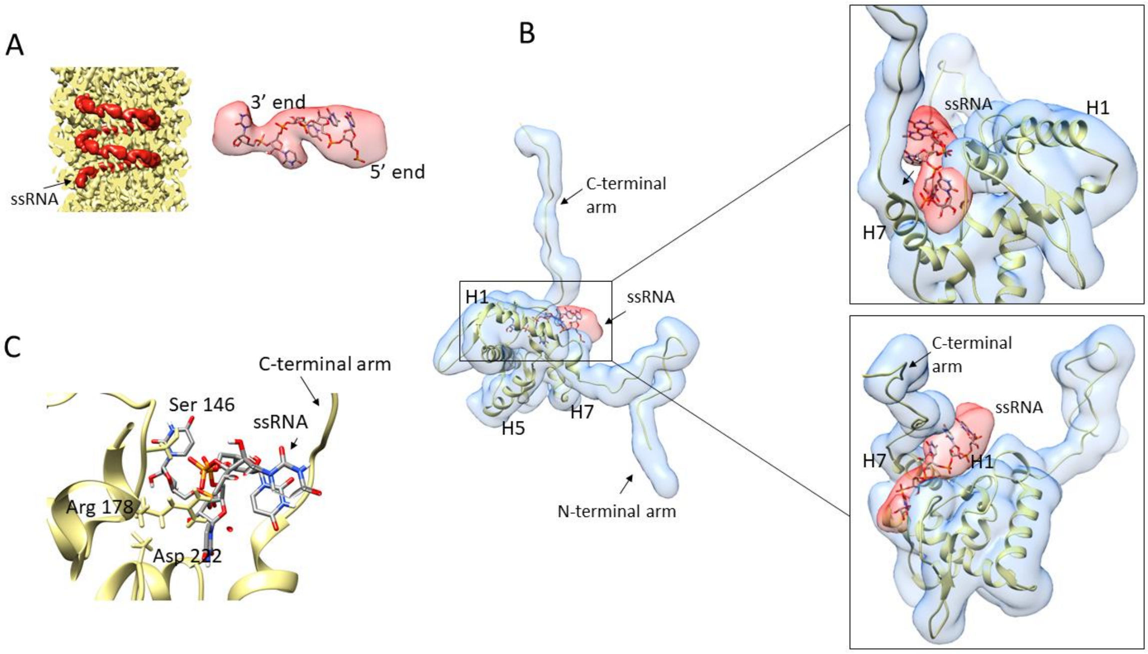


Fig 4.8 ssRNA density and placing. A) ssRNA density displayed in red along the filament and on the right, the transparent density with the 5 nucleotides that spans one coat protein subunit. **B)** Placing of the ssRNA between the H7 and the H1 helices of the coat protein **C)** The conserved aminoacid region in the binding pocket of the coat protein composed of S146, R178 and D222.

This 3D density map was uploaded in the "Electron Microscopy Data Bank" (EMDB) database (www.emdatabank.org) with the accession code EMD-10373 and the atomic model calculated for TuMV's CP and ssRNA was deposited in the "Protein Data Bank" (PDB) database (www.rcsb.org) under the accession code 6T34.

4.1.4 Analysis of VLPs maps and implication of ssRNA in virion structure

For TuMV virion cryoEM map reconstruction, helical symmetry was imposed but in the case of VLPs the imposition of the symmetry didn't converge into reproducible 3D maps, thus, a 3D classification of the filament segments was performed without any imposed symmetry. This classification showed that only 60% of the particles had a helical arrangement with defined CP subunits, while the others had very poor structural features (Fig 4.9).

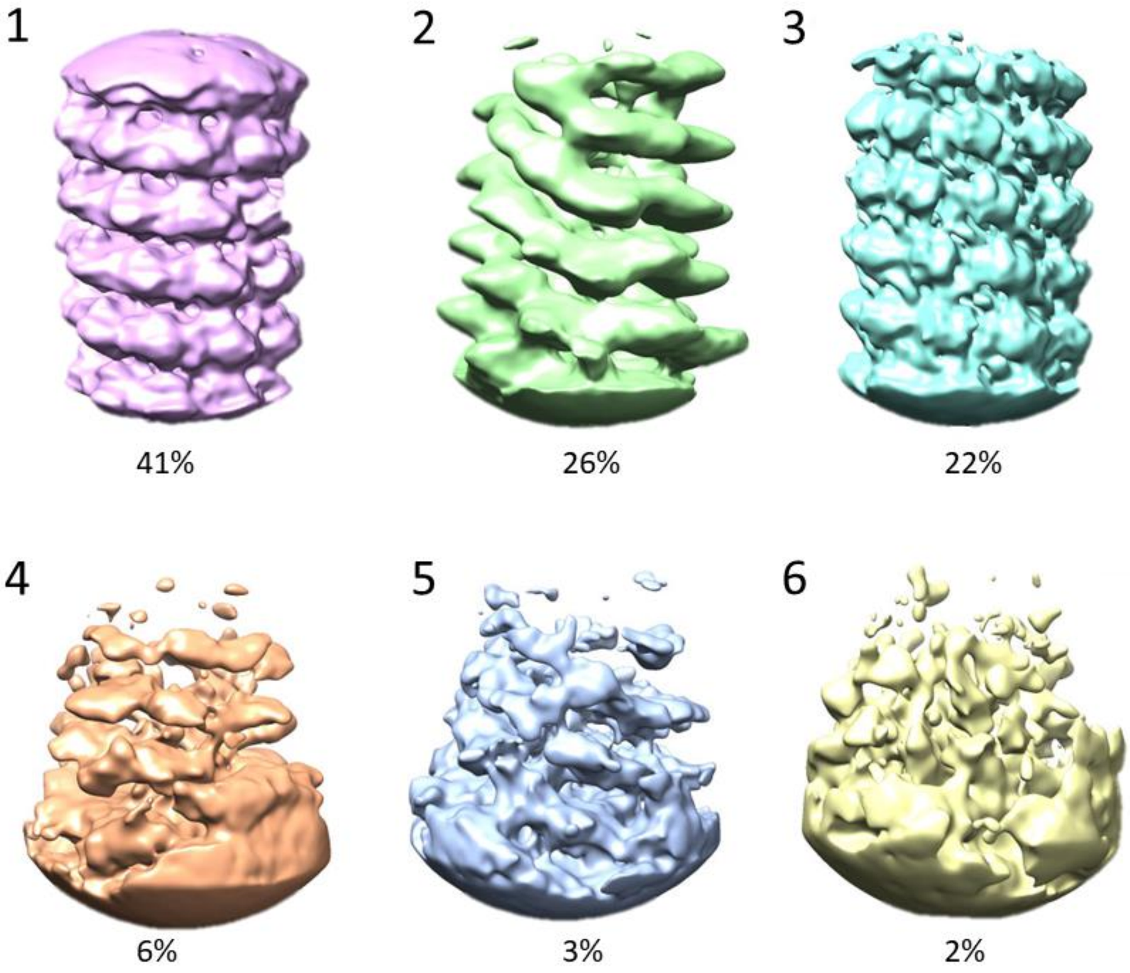


Fig 4.9 3D classification of TuMV VLPs. Only 63% of the particles, corresponding to classes **1** and **3**, displayed good characteristics and a well arranged helical form.

The absence of ssRNA in the VLPs produces labile multimers with distorted local regions along the filaments. The two classes with good features were selected for refinement imposing helical symmetry. Resolution of both maps was around 8 Å (Fig 4.10 A, B).

Low resolution of the filaments suggested that VLPs are less stable, structurally more heterogeneous, and for that reason their 3D averages showed limited structural details. Both classes, however, exhibit helical symmetry parameters identical to those of TuMV virions (Fig 4.10 A, B) meaning that the overall structure of the filament is maintained with local movements around the place where ssRNA should be placed.

At this level of resolution, it was not possible to build an accurate atomic model for the coat protein of the VLPs so, for the interpretation of their structures, we fitted the atomic coordinates modelled for TuMV virions as a rigid body. In the maps for both classes the path for the ssRNA resides in an empty passage (Fig 4.10, C), what confirms the absence of the genome in the VLPs and also that the fitting of the CP polymer is on the correct register with respect to the 3D maps.

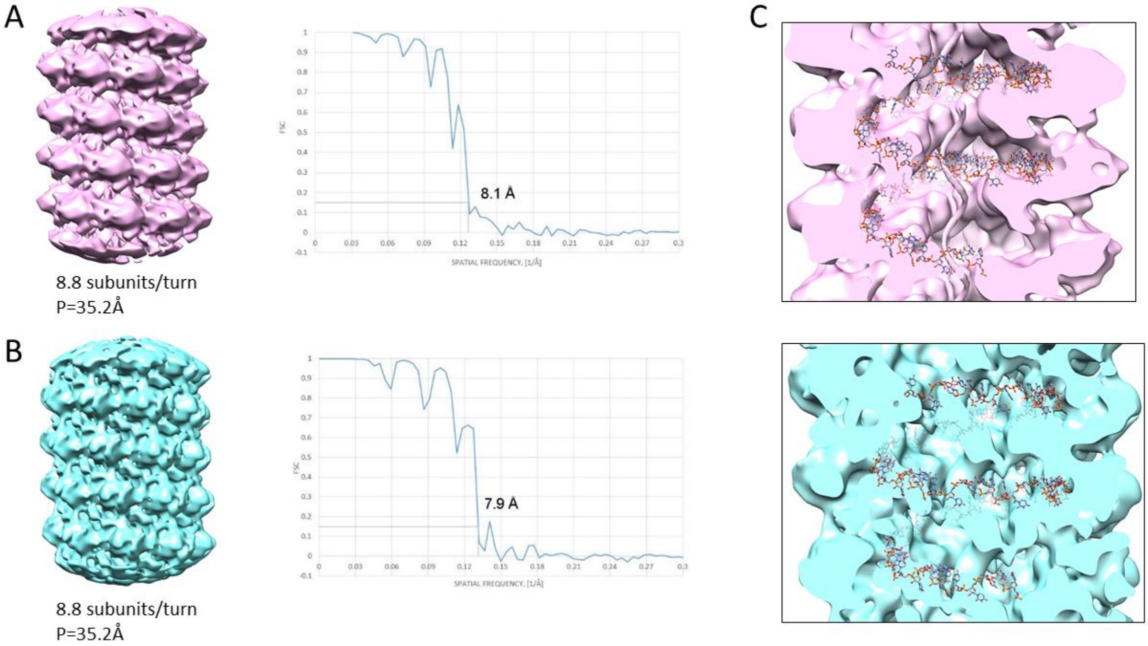


Fig 4.10 3D density maps for TuMV VLPs. A) 3D density map for class 1 at 8.1 Å with identical helical parameters as those of TuMV virions. **B)** 3D density map for class 3 at 7.9 Å with identical helical parameters as those of TuMV virions. **C)** 3D maps of class 1 and 3, clipped along helical symmetry axis showing how the ssRNA modeled for TuMV virions relayed in an empty passage inside the VLPs.

In class 1, helix H7, that delimits the ssRNA binding groove in the virions, seems to move towards the inner side of the filament. The fitting of the coordinates for the oligomer of CPs leaves the N-terminal arm outside the density: fully outside in class 1; or only in the last region that participates in axial interactions in class 3. Also, the densities for helices H1 and H5 are incomplete, and both secondary structure elements stick out at certain degree from the maps. Thus, the role of the N-terminal arm in polymerization and the position of helices H1 and H7 are perturbed in the absence of the ssRNA (Fig 4.11).

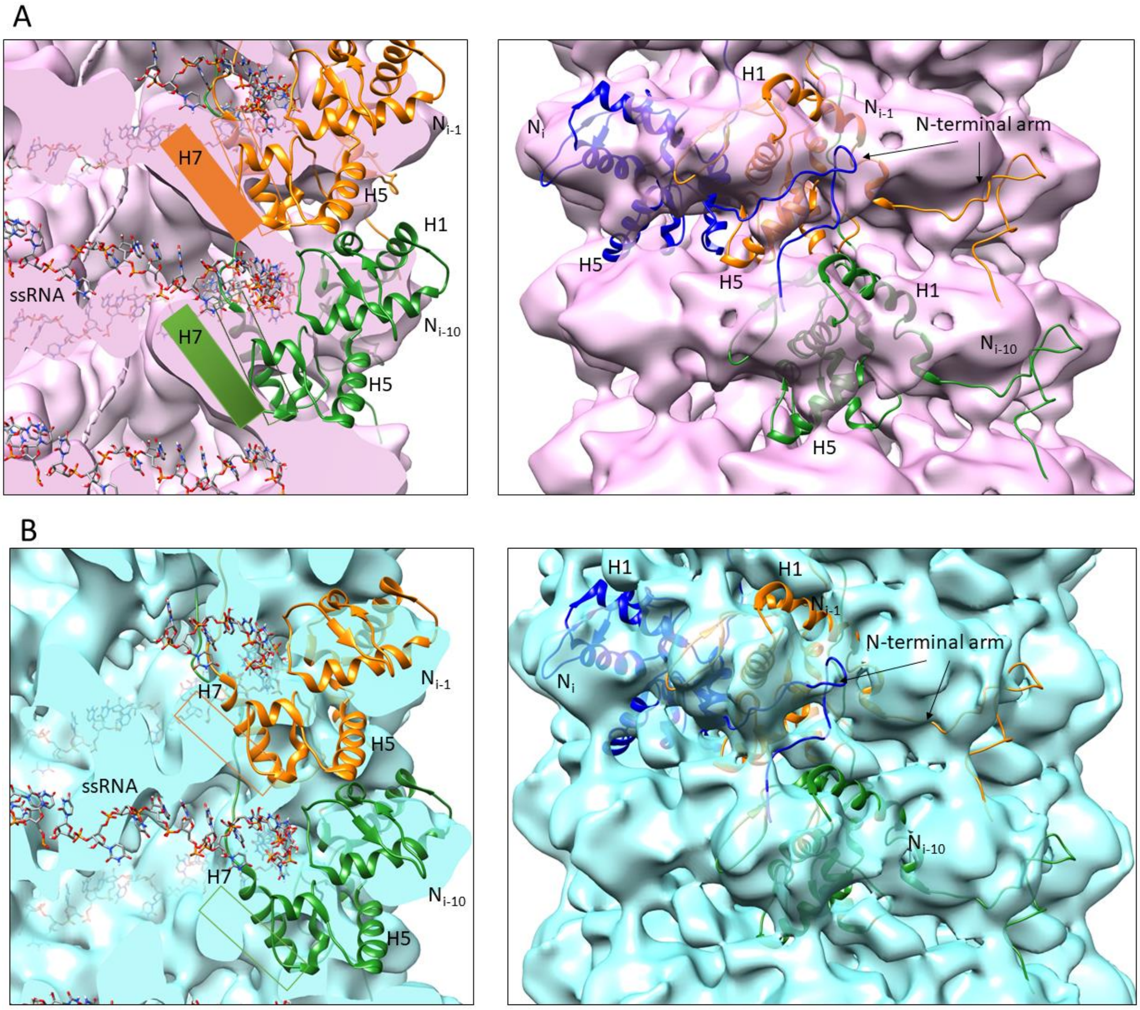


Fig 4.11 Analysis of changes on the CP in both classes in the absence of ssRNA. A) The density map for the Class 1 VLP is displayed with the CP atomic models of TuMV virions. Movements of the helices H7, H1 and H5 are pointed out, as well as the N-terminal arm completely outside of the density. **B)** The density map for the Class 3 VLP is displayed with the CP atomic models of TuMV virions. Movements of the helices H7 and H1 are displayed, also the N-terminal seems not to be on the track of TuMV virions CP.

To gain some insights into the influence of ssRNA over these structural elements we revisit the atomic model for TuMV virions. Firstly, it is important to observe how the CPs are arranged along the helix to understand the connections in its entirety. As mentioned before, the helix is composed of 8.8 subunits per turn. We listed the subunits (N) with the following criteria, from N_i to N_{i-10}, in order to explain the interaction between them (Fig 4.12).

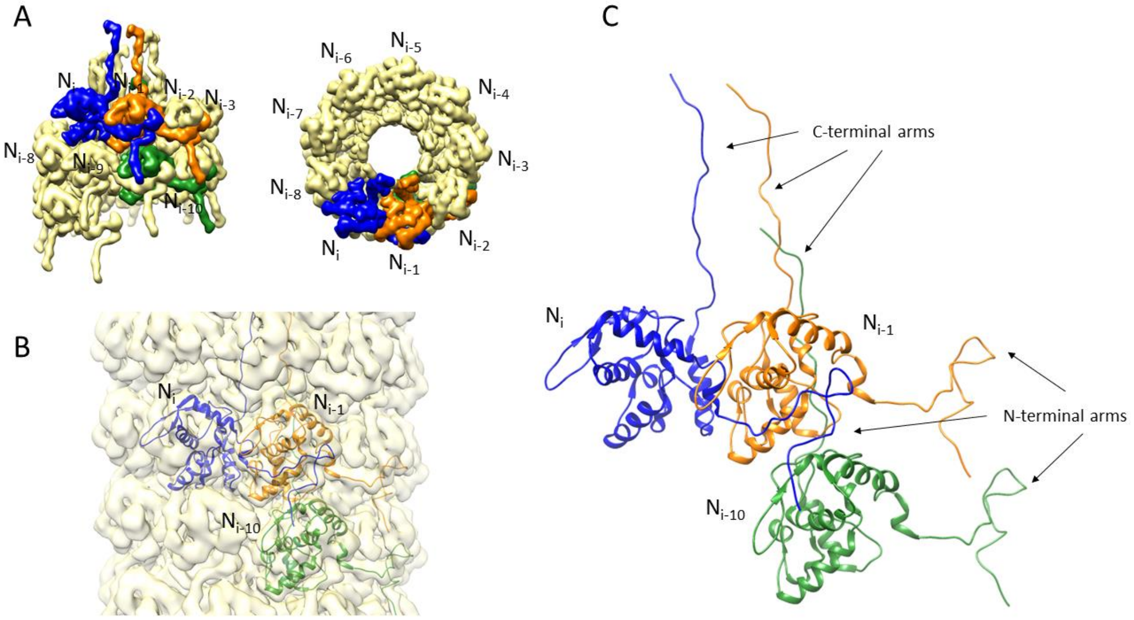
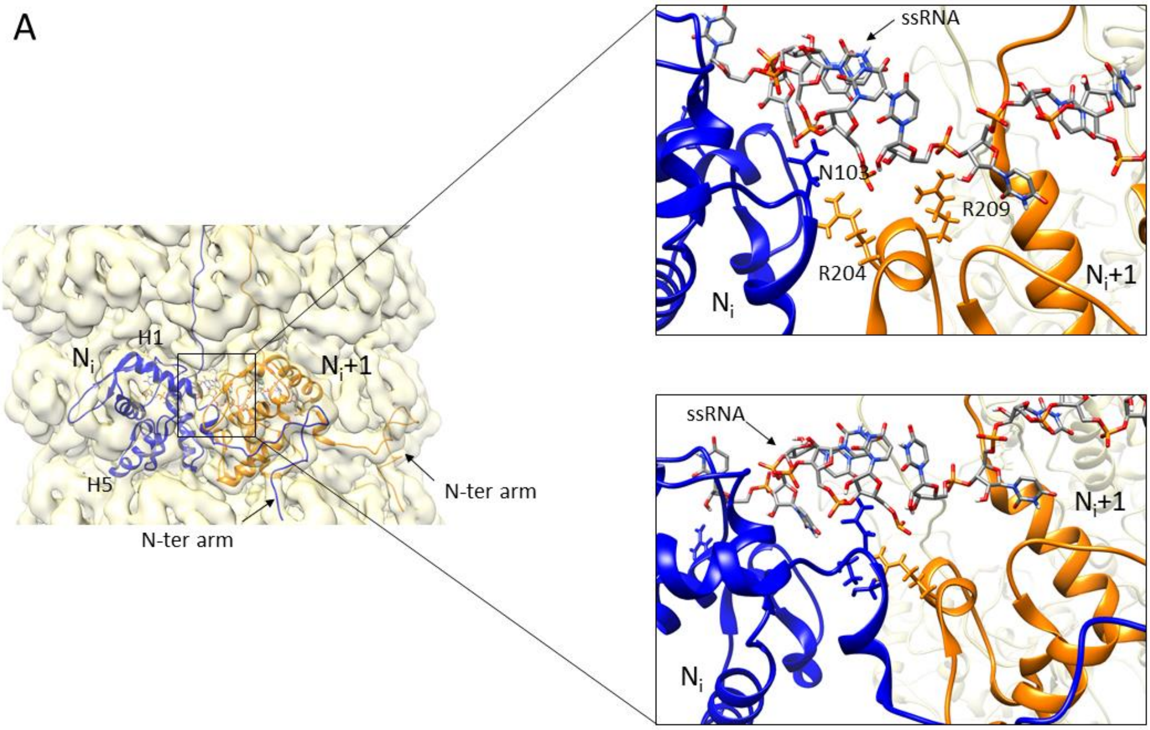


Fig 4.12 Arrangement of CP subunits along the helix. A) Density model of a polymer modelled with two turns of the helix of the filament. Near every subunit there is their assigned name that goes from N_i to $N_{i-10.}$ **B)** Density for the filament with 3 subunits highlighted in blue (N_i), orange (N_{i-1}) and green (N_{i-10}), that are the ones that illustrate the interactions between CPs. **C)** Atomic models of the subunits N_i, $N_{i-1,}$ $N_{i-10.}$

In the boundary between CP subunits there is a network of protein-RNA and protein-protein interactions that supports the proper orientation of the flexible N-terminal arm. Residue N103 from one subunit (N_i), and the pair R204 and R209 from the adjacent CP (N_{i-1}) interact with the phosphate backbone of the ssRNA (Fig 4.13, A). At the same time, these two regions are connected between them, in such a way that R204 interacts with the beginning of the N-terminal arm that contains the aforementioned N103, and S102 and T104 at the neighbouring subunit (Fig 4.13, A). These local interactions with the ssRNA and between CPs serve to anchor helix H1 and the N-terminal arm of one CP subunit (N_I) and helices H5 and H6 of the neighbour (N_{i-1}). Since helix H5 builds part of the groove for the interaction with the N-terminal arm, the contacts with the ssRNA modulate both the donor and the acceptor in the interaction via the N-terminal arm. The three residues that make direct contact with the ssRNA in this region are highly conserved in potyviruses (N103 90%, R204 80%, and R209 83%) and are also seen involved in the same interactions with the nucleic acid in WMV and PVY. In this same local region, helix H1 and the N-terminal arm (subunit N_i) interact with the N-terminal arm of the other subunit from the next helical turn (N_{i-9}) (Fig 4.13, B). Here, the hydrophobic interaction F115-Y80 and the salt bridge E97-R76 are key to set the 90º turn of the N-terminal arm towards the next turn of the helix. The F115-Y80 connection between TuMV CPs has equivalent pairs in WMV and PVY, where the hydrophobic pair is established between Tyr (Y) and Val (V) residues. However, the E97-R76 salt bridge has no counterparts in the other two potyviruses, probably due the high diversity of sequences at the N-terminal arm (Fig, 4.13, B).



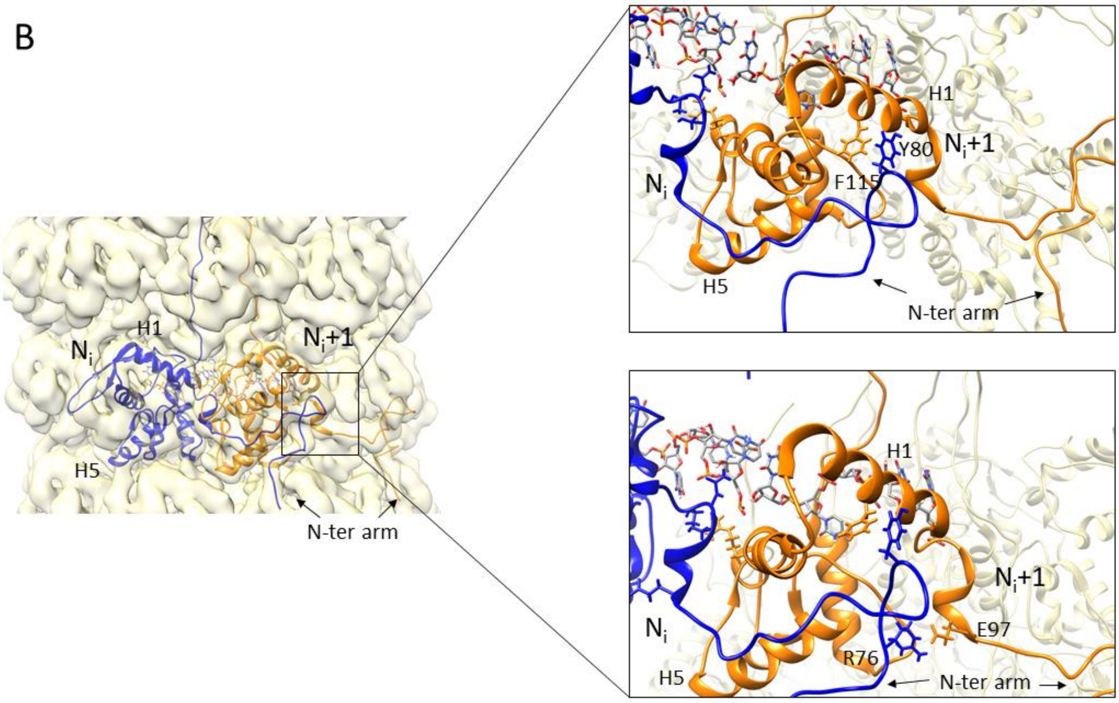


Fig 4.13 Network of ssRNA-CP and CP-CP interactions. A) Protein-protein and ssRNA-protein interactions between adjacent subunits. Implicated amino acids are indicated in the image. The image on the left indicates de orientation of the CPs in the virion and the place where those residues are interacting between them. **B)** Same scheme as above but displaying different residues in different locations of the protein. Implicated helices are also indicated.

3D density map resulting from this work was uploaded in the EMDB database (www.emdatabank.org) with the accession code EMD-10374.

4.2 CRYOET AND SUBTOMOGRAM AVERAGING OF TuMV VIRAL ENDS

In the last few years, the structural data available for flexible filamentous plant viruses has grown significantly. As the helical processing takes the central part of the filament for the reconstruction of the whole virion, there is no information about the terminal ends and how the filaments are capped. Here we studied the terminal ends of the potyvirus TuMV by cryoET. Structurally, little is known about the ends but genomically it’s known that the 5’ terminal end of the genome of potyviruses has a protein attached to it, called VPg, and the 3’ terminal end has a poly (A) track. The goal of this project was to get structural information about the organization of filamentous viruses at their filament ends, and also to explore the presence of the VPg protein attached to 5’ terminus end by cryoET followed by subtomogram averaging. In cryoEM image reconstruction, large set of particles displaying views from different angles are needed in order to achieve a good 3D reconstruction. In our case, this entailed a big issue since in a classical potyviral cryoEM micrograph we can see only a few viral ends (Fig 4.14) since filaments are so long that they do not fit in one micrograph. Besides, filamentous particles tend to lay horizontally on the grid which means having only one view of our particles. In contrast, in cryoET a series of projection images from the same object are recorded as the sample is tilted to various angles relative to the incident electron beam. The images are subsequently aligned and reconstructed to generate a 3D tomogram. It provides a 3D volume of a single unique specimen which later can be postprocessed via subtomogram averaging to obtain a 3D reconstruction of the macromolecule of interest. Structural information about the filament tips was recovered thanks to the multiple orientations with respect to the tilt axis that the tips presented in the tomograms overcoming the issue of the missing wedge that affects all tomographic data.

TuMV virions were used and the samples were purified exactly as described for the cryoEM analysis of TuMV filaments. The presence of VPg protein was confirmed by proteomics analysis. For cryoET grid preparation we added 4µl of gold beads to the viral sample, coated with bovine serum albumin (BSA) which are used to better alignment of the tilt series of the tomograms. Apparently, coating the gold fiducials with BSA prevents the fiducials adhering to the sample, also it can prevent clumping and promotes even distribution on the grid [124].

Tomographic data was collected in a Talos Artica electron microscope operated at 200kV using a Falcon III direct detector camera at the “Centro Nacional de Biotecnología” (CNB-CSIC). Usually, a 200kV microscope is not suitable for tomographic data acquisition because tomographic samples tend to be thick and at this voltage the electrons are accelerated at lower speeds and penetrate less inside the sample. Nevertheless, our sample is very thin since we only need a thin layer of ice that contains the filaments and a 200kV microscope is suitable for our work. Tilt series were acquired at a magnification of x57,000 with a corresponding pixel size of 1.83 Å using EPU software [125]. The electron dose per projection was 2.32 e^- /Å^2 with a total tilt dose of 95 e^- /Å^2. Tilt range was from -60° to 60° with 3° increments. Collection resulted in 86 tilt series with 41 tilts per tomogram, recorded in 3,440 frames.

Alignment and reconstruction of the tilt series was accomplished using the IMOD software package [126]. Tilt series alignment determines shifts between each image and refines the orientation of the tilt axis, refines the angles and compensates image deformations [93]□Gold fiducials were used for the automatic alignment of the tilt series but sometimes manual selection of the fiducials was needed. Tomogram reconstruction was made following the back-projection method. Tomograms were binned by a factor of two what resulted in a pixel size of 3.66 Å. Tomogram reconstruction was followed by subtomogram averaging. In subtomogram averaging individual repeating particles (subtomograms) are aligned in 3D and averaged to improve resolution and signal-to-noise ratio (SNR) [127]. Particle picking was made using 3dmod interface. We selected the end part of the filaments with a box of 120 pixels, including a small part of the filament but also the apparently “empty” space outside the filament. (Fig 4.14, B).

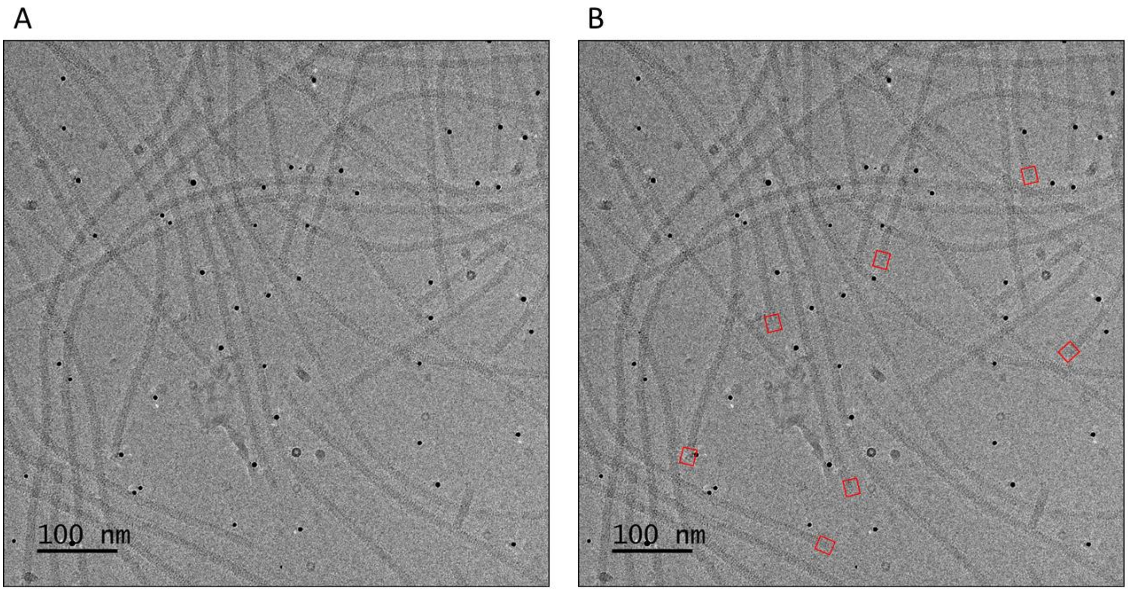


Fig 4.14 Central plane from a full reconstructed tomogram. A) The black dots are the gold beads used as fiducials markers for the alignment of the tomogram. **B)** Red boxes show how the particle picking was made, selecting the ends of the filaments.

For CTF refinement and extraction of the particles we followed Relion's subtomogram averaging protocol [115]. Next steps of 3D classification and refinement were done following the single particle methodology [128].

598 subtomograms were extracted from 74 tomograms. 3D refinement was performed using the density map for TuMV with a low pass filter of 60 Å as initial reference and a mask diameter of 400 Å. This mask radius is the diameter for a soft circular mask that is applied to the experimental images. This mask allowed us to see the reconstruction of our entire picked particle (Fig 4.15, A). This alignment showed a filament with different layers, where we could also slightly appreciate some blobs at the surface that could be attributed to the coat proteins. Also, the diameter of TuMV virions is 135 Å in agreement with our preliminary density map.

As our interest was to focus on the end of the filaments, we did a 3D classification applying a smaller mask (100 Å) in order to see features that allowed us to separate the two ends of the virions. We performed 3D classification with 3 classes. All the particles distributed mainly in two classes, the first one with 53% of the particles and the second one with the 42% (Fig 4.15, B). The third class had the remaining 5% and it didn't show any features. First class displayed and end similar to that seen in the prior 3D refinement with an abrupt end, but the second class displayed a longer end with the shape of a helical cap (Fig 4.15, B). We refined both classes with a mask radius of 150 Å. The refined map for both classes resulted in better aligned maps with more defined features. Both maps were postprocessed and resulted in two maps at a resolution of 31.37 Å (Fig 4.15, C).

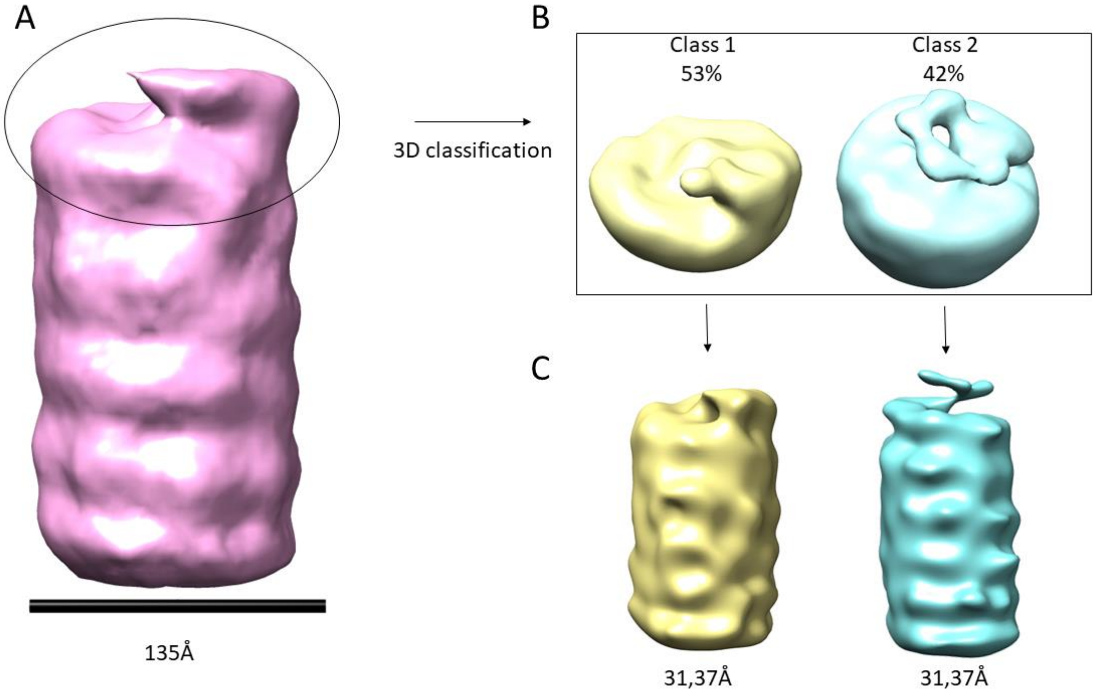


Fig 4.15 3D density maps for 3D refinement and 3D classification. A) 3D density map from the 3D refinement of TuMV subtomograms. **B)** Class 1 and 2 from 3D classification with the distribution of the particles. **C)** Final 3D density map reconstruction for both classes, displaying the different ends at their extremes and the achieved resolution.

Although the resolution achieved is very low, it allows to interpret our maps with the help of the atomic model for PVY [99]. We used PVY CP atomic model since it has the C-terminal arm described in its completeness and we assume that TuMV CP will have the similar architecture. We observe that there are two populations of particles, presumably corresponding to the two ends of the potyviral filaments. As said before, it is known that the VPg is covalently attached to the 5’ terminus of the genome but there is no structural data available for the virus tips. Previous work revealed the existence of an unusual structure at one of the ends of PVY, showing a protruding tip with the shape of a helix [99]. It was proposed that this helical end could correspond to the 5’ terminus of the genome and that it was composed by the two viral proteins VPg and HCPro [129]. Nowadays, we know the relative orientation and arrangement of CPs and the genome inside the virions. The C-terminal region is buried in the lumen and interwoven in an helical manner along the longitudinal axis, forming a compact cone-shaped structure that could be exposed in its vicinity to the RNA 3' terminus (Fig 4.16, A) [99]. Our helical cone structure in class 2 of TuMV virions reminded the cone-shaped structure of PVYs C-terminal, so this helical structure might correspond to the C-terminal arms of the coat proteins from the last turns of the helix that are exposed to the surface since the filament is reaching its final part. With this in hand we can relate this class to the 3' end of the viral genome, meaning that the other class represents the 5' end which implies the location of the VPg protein on that density map (Fig 4.16, B).

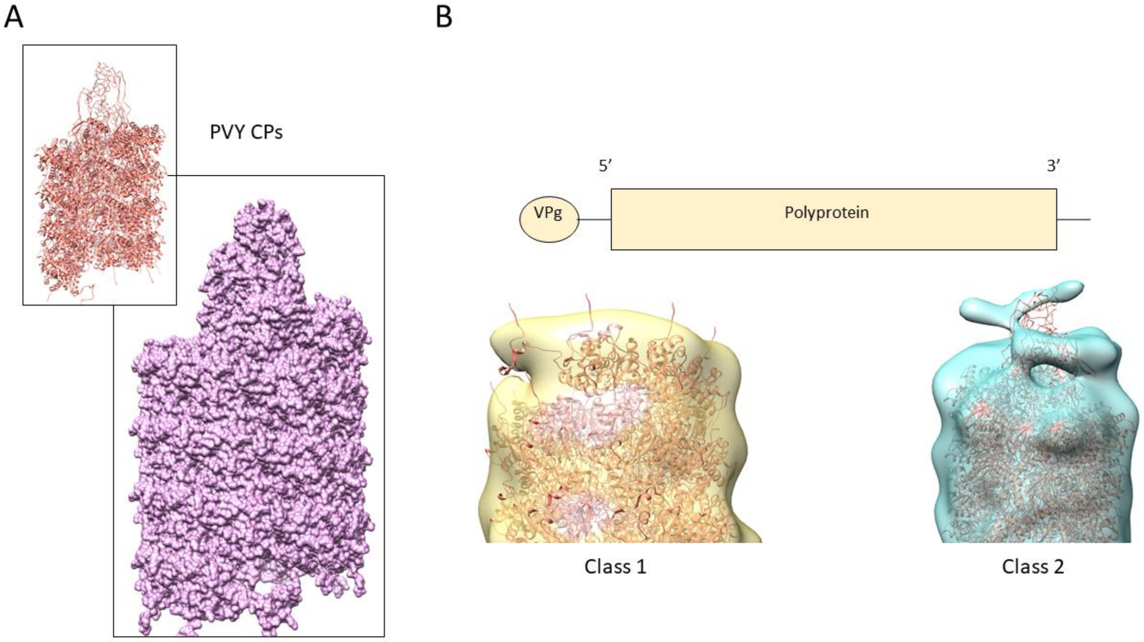


Fig 4.16 Assignment of 3D structures with the two ends of viral filaments. A) Polymer of CPs of PVY showing the architecture for the C-terminal ends of the proteins, ending in a cone helical shape. **B)** Scheme of the potyviral genome and assignation to both 5' and 3' ends thanks to the fitting of the atomic model for PVY onto the density maps of class 1 and class 2. Class 1 corresponds to 5' terminal end and class 2 to 3' terminal end.

The first class had an arrow-head shape density structure at the end of the filament. This shape could be the density for an additional protein or for a last coat protein. We fitted the available atomic structure for VPg [57] and the polymer of PVY CPs [99]. VPg structure adopts a well-folded core composed of a 5-stranded β-sheet with two consecutive α-helices between β-strands 2 and 3. It has two unstructured regions at the N- and C-termini and a flexible loop between β4 and β5 strands (Fig 4.17, A) [57]. We fitted the central core of the protein into the end of the density map. The structure of VPg protein fits in good agreement into the shape of the density at the end of the filament (Fig 4.17, B). We focused only on the structured core of the protein since the flexible loop and the C- and N- unstructured regions would not be visible at such low resolution.

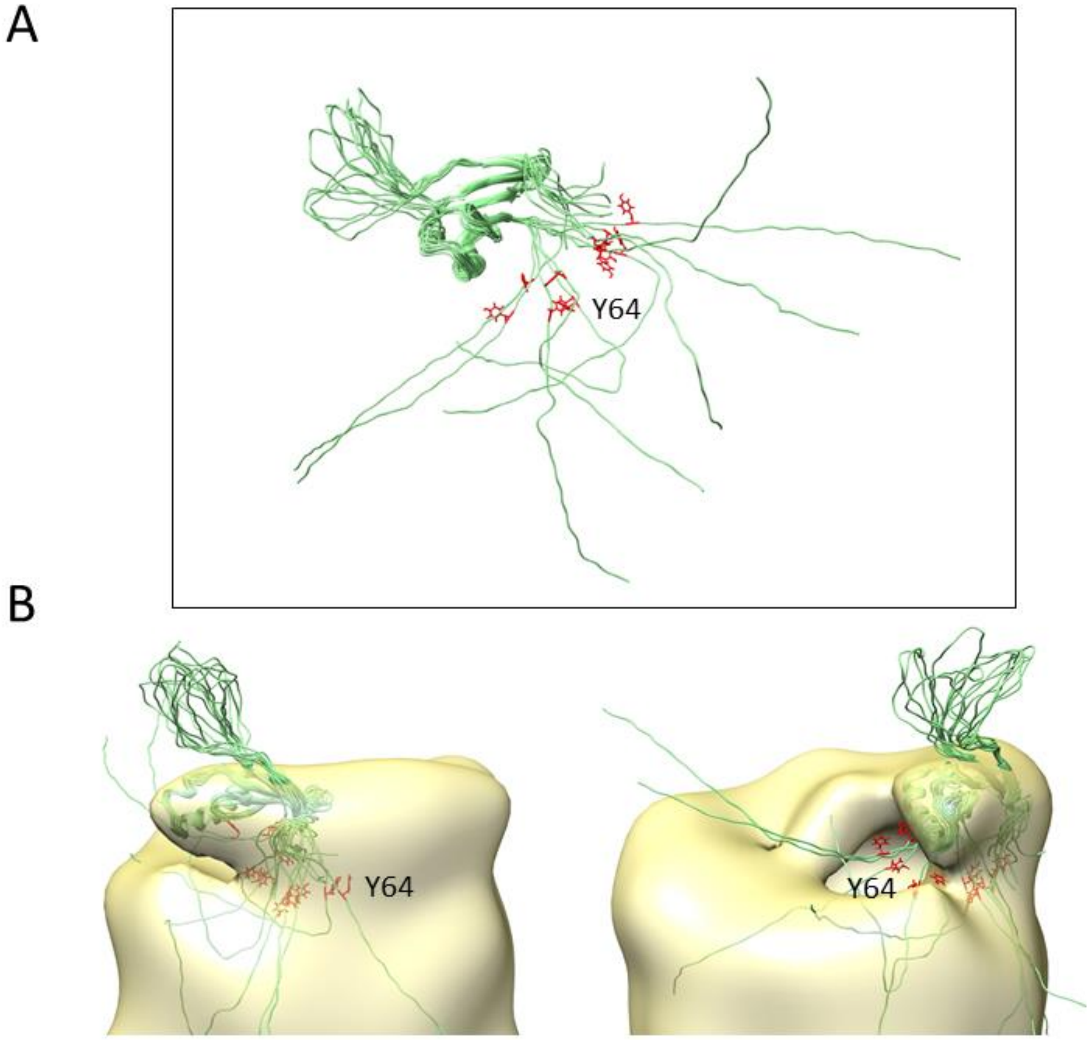


Fig 4.17 Fitting of VPg structure in TuMV subtomogram density map. A) Ensemble of VPg structure resolved by NMR [57]. Y64 is highlighted in red since is the residue that covalently attaches to the ssRNA. **B)** Fitting of VPg in the arrow-head shape density at the end of our tomographic density map. Again, Y64 is highlighted in red.

We also moved the polymer of PVY CPs in order to fit the last CP onto the final density to test the possibility of that density to be one of the coat proteins of the virion. The globular structure of the CP didn’t fit in the arrow-head shaped density, meaning that it was very unlikely that that density corresponded to a final CP (Fig 4.18, A). Finally, we fitted the CP polymer and the VPg within the 3D map for the end of the filament. In this fitted disposition the residue Y64, which is covalently attached to ssRNA, is placed close to the RNA which supports the location of the VPg (Fig 4.18, B).

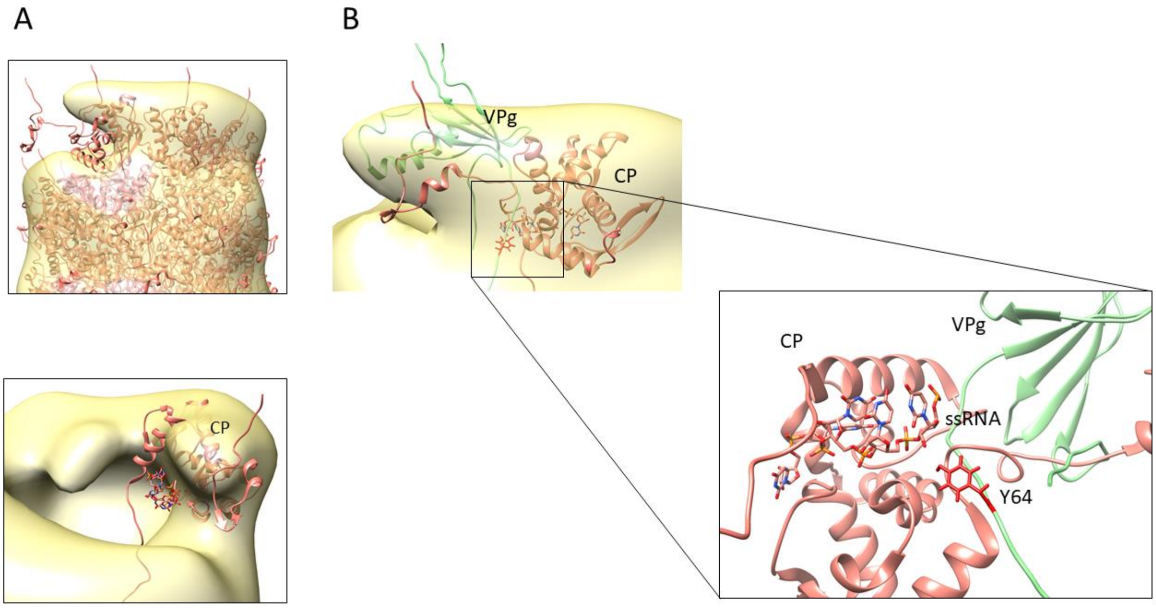


Fig 4.18 Fitting of CP onto TuMV subtomogram density map and analysis of VPg interaction with viral genome. A) Fitting of viral CP onto the final tip of the density. Since the CP is a globular protein it does not fit in the region. **B)** Here the interaction between viral genome and VPg is shown. We selected one of the conformations of the VPg and highlighted Y64 in red, seeing that this residue and the viral ssRNA are in close contact.

Although resolution is very limited, we have found some interesting features that may lead to a better understanding of the structure at the filament ends of potyviruses.

4.3 STRUCTURAL STUDIES OF PVY VLPs AS PLATFORM FOR CHARACTERIZATION OF COMPOUNDS THAT TARGET VIRAL CPs

Structural data about plant viral particles may be of big impact for the viral plague control in agriculture. Nowadays a wide array of phytosanitary disease control methods are being used but their efficiency is limited, and in some cases, they are not environmentally acceptable. For that reason, there is a great need of developing effective chemical treatments against plant virus. Here, we wanted to explore the ability of active molecules to modulate the viral activity of filamentous plant viruses. We know that these viruses contain a universally conserved region in their union with the ssRNA, which could be a potential target for the development of compounds for chemical control of plagues. The goal of this research was to characterize the interaction between potentially active compounds and the CP of flexible filamentous viruses. This characterization was studied using structural methods such as NMR and cryoEM. Revealing the nature of the interaction would be the first step for the creation of a specific active molecule with biotechnological application in agricultural crops. In collaboration with the NMR platform at CIC-bioGUNE and with our collaborators at CEBAS-CSIC in Murcia, we explored the interaction between selected chemical compounds and four different CPs from the families *Alphaflexiviridae, Potyviridae and Closteroviridae*.

CPs from PepMV and WMV, a potexvirus and a potyvirus, were used for an *in silico* screening of a chemical library (Maybridge Ro3 fragment library). The goal of the screening was to identify potentially active molecules that interact with the conserved binging site of the CP with the ssRNA of PepMV and WMV. The *in silico* screening resulted in a battery of molecules that scored high values in estimated binding affinity. We purified PepMV CP labelled with N15 for NMR experiments for further testing with the candidate compounds. PepMV's Heteronuclear single quantum coherence (HSQC) shows a disperse spectrum indicative of a properly folded protein with a tertiary structure (Fig 4.19). Three tryptophan side-chains are present as expected from the sequence. Arginine side-chain region shows three mayor signals. Tested compounds showed chemical shift perturbation (CSP) with three of the tested compounds, a clear indication of protein-ligand interaction (Fig 4.19, B); some CSPs were preserved in the three cases such as arginine's NE-HE or some backbone HN groups, but some other were ligand specific indicating that the binding mechanism could be different in each case (Fig 4.19, B). The selected compounds are designated by an internal code as: 3.G9, 25.D8, and 28.C6. More details about the compounds are not disclosed since they are not relevant in practical terms and there is a signed non-disclosure agreement with another institution.

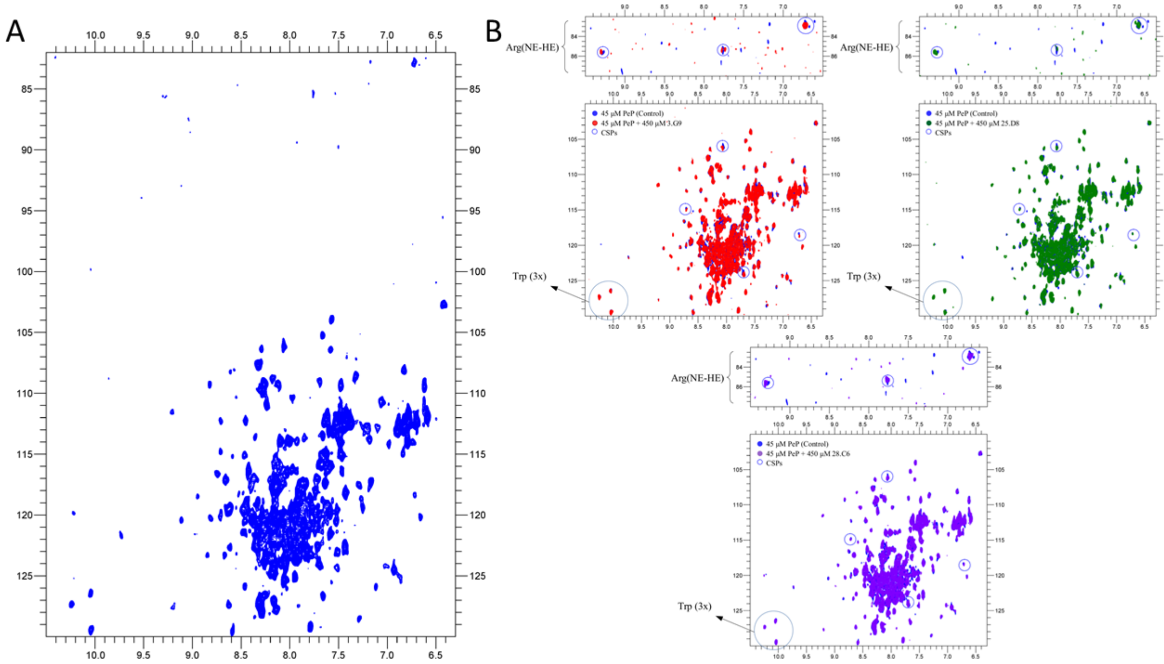


Fig 4.19 HSQC spectra of PepMV CP alone and with selected compounds. A) In blue, HSQC spectrum of PepMV CP for control. **B)** Three HQSC spectra are shown, all of them overlapped with the spectrum for the control. In red the spectrum for PepMV CP with compound 3.G9, in green with 25.D8 and in purple with 28.C6. All of them displayed chemical shift perturbation, showing that there is protein-ligand interaction at some specific residues.

According to the atomic structure for PepMV CP, the side chain of R124 should be a clear reporter of ligand-protein interaction in the RNA binding pocket. Unfortunately, R124 cannot be easily assigned, at least with this analysis, so we mutated the R124 for glutamine (R124Q) and did the same experiment. The HSQC experiments showed that the signal for one of the arginines of the mutated CP (colour black in the spectra) was not seen compared with the wild type CP (colour blue) (Fig 4.20). This allowed us to assign the residue R124 in the spectrum.

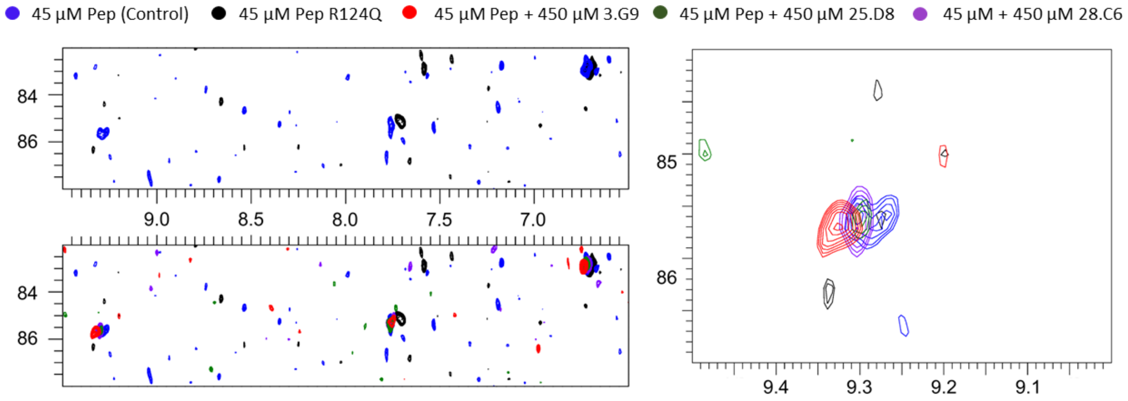


Fig 4.20 HSQC spectra of PepMV CP alone, mutated and with compounds. In the upper left spectrum we see the arginines for control PepMV CP (in blue) and the ones for the mutated sample, R124Q (in black). The image down left displays all the spectra overlapped and the image on the right shows the arginine that does not appear on one of the spectra, the mutated one.

We selected compounds 3.G9 and 25.D8 for testing their interaction with different potyviral samples. WMV virions were incubated with both compounds in order to test if they had some effect on the whole virions and not only at CP level. The ratio used for this experiment was 1:20 of CP:ligand to secure saturation, meaning 2 µl of virus sample + 0.26 µl of compound solution in 18 µl of citrate buffer. We incubated the samples 30 minutes at RT and did negative stain grids to check samples by TEM. In the micrographs, in the presence of the compound 25.D8 some local distortions on the filaments are observed, which we attribute to partial openings of the helix (insets in red boxes in Fig 4.21, A). We performed the same experiment but left the sample incubating with the ligand ON at 4ºC. In the control, virions appear as long filaments, well distributed and organized (Fig 4.21, B). With compounds 3.G9 and 25.D8 the effect was notorious, and the majority of the virions were broken and aggregated (Fig 4.21, B).

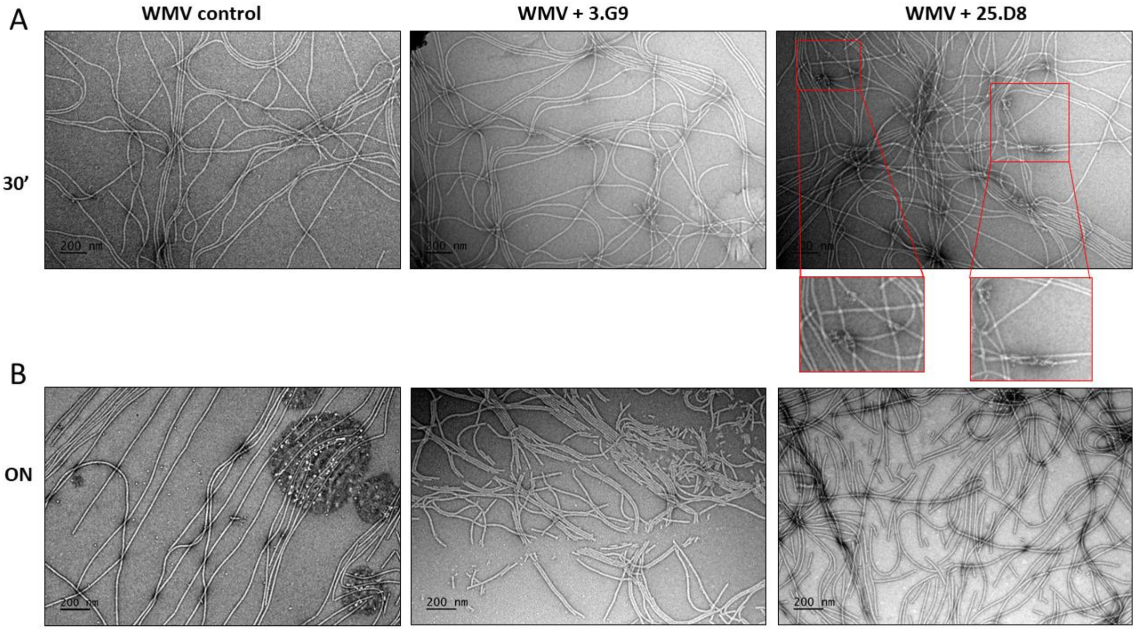
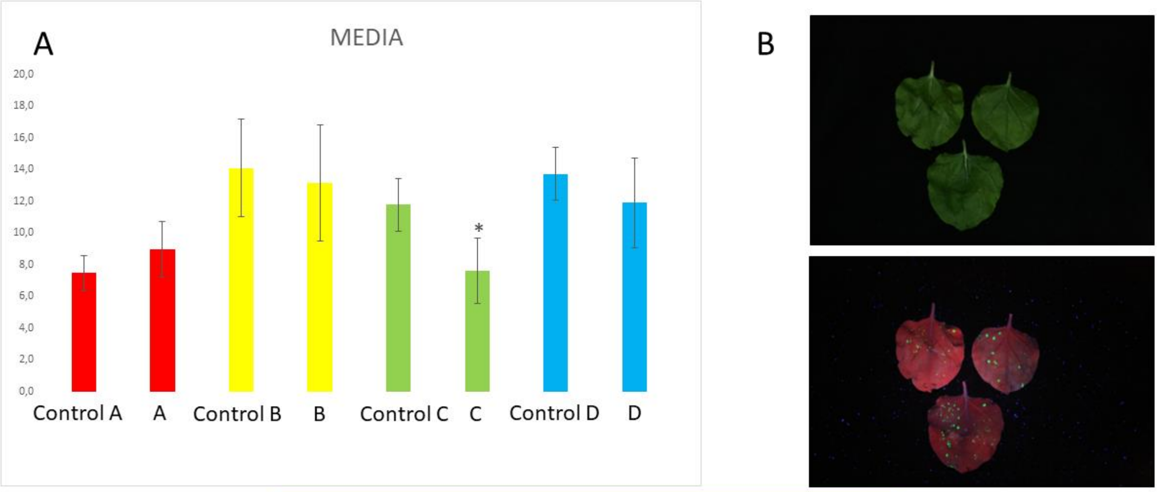


Fig 4.21 Negative staining micrographs of WMV virions with selected compounds. A) Micrographs showing, from left to right, WMV virions without ligand, WMV with 3.G9 ligand and WMV with 25.D8 ligand, after 30 minutes incubation. **B)** Micrographs showing the same samples as in **A** but after an incubation ON.

At the same time, our collaborators at CEBAS-CSIC did a preliminary test in carborundum-dusted *N.benthamiana* leaves. This treatment removes the wax layer of the leaves and makes wounds on them. They added PepMV virions to the surface of the leaves and let it penetrate inside. The genome of these virions has been modified to include the green fluorescent protein (GFP) and when the plant infection is effective, then GFP is expressed in the infected cells. The infection zones are fluorescent under ultraviolet light thanks to the GFP (Fig 4.22, B). They did 4 experiments with 10 leaves for each experiment.

Leaves were differentiated in two halves, one side was infected with a compound free sample of PepMV virions, and the other one was treated with one of the three compounds (Fig 4.22, B). Compounds were incubated for 30 mins at RT. They tested four compounds; 3G.9 (compund A), negative control (compound B), 25.D8 (compound C) and 28.C6 (compound D) (Fig 4.22). The negative control used in these experiments is a compound library that showed negative results for interaction with CPs in the chemical shift experiments.

Although it was a very preliminary experiment, we could see some differences calculated following an ANOVA statistics analysis. ANOVA is an analysis of variance in a population. It decomposes the variance of spots into two components: a between-group component and a within-group component. The F-ratio is a ratio of the between-group estimate to the within-group estimate. If the P-value of the F-test is less than 0.05 then there is statistically significant different between the mean spots from one level treatment to another at the 95.0% confidence level. Compound C proved to be the only one that showed statistically significant differences, meaning that the compound might interfere with the viral activity (Fig 4.22, A, C).



C

	P-values	F-ratio
A → 3.G9	0.4022	0.71
B → -CONTROL	0.7032	0.15
C → 25.D8	0.0113*	6.85*
D → 28.C6	0.5720	0.32

Fig 4.22 Preliminary experiments of the three selected compounds in plant leaves. A) This chart displays the four experiments differentiated by colours. Colour red represent experiment with compound A (3.G9), in yellow the one with the negative control, in green the experiment with compound B (25.D8) and in blue the last one with compound D (28.C6). The asterisk indicates the only experiment that gave statistical differences. **B)** Photographs of leaves with no UV light and under UV light. The left part of the leaves was sprayed with virus control and the other half with the virus plus the corresponding compound.

We selected compound 25.D8 as a potential candidate for interaction with the conserved region of flexible filamentous plant viruses. To further strengthen this possibility, we tested this compound for its interaction with members from other flexible filamentous plant virus families. We included the CPs from “Citrus tristeza virus” (CTV), a closterovirus, and from WMV, a potyvirus, for our next NMR experiments. Besides, we mutated R172 for an A (R172A) in WMV CPs to test if this change at the conserved RNA binding site had any influence in the binding. “Saturation transfer difference” (STD) experiments shows only ligand signals when there is binding between the protein and the small molecule. We tested five different samples; PepMV CP, PepMV CP + 25.D8, CTV + 25.D8, WMV + 25.D8 and WMV R172A + 25.D8 (Fig 4.23).

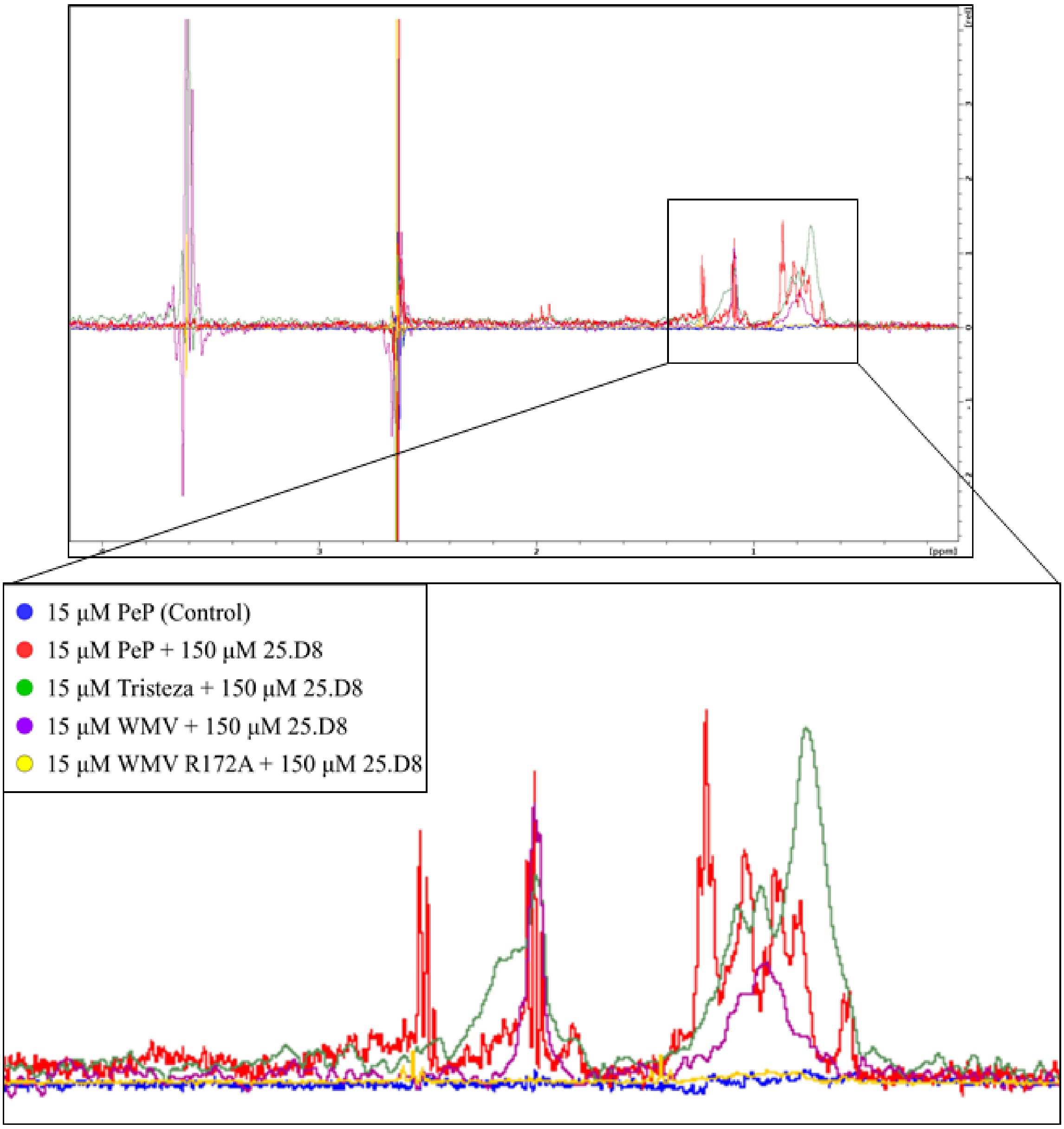


Fig 4.23 Spectra for STD experiments with the coat proteins of different viruses. The upper image shows the whole STD spectra for all tested viral coat proteins, each CP spectrum displayed in a different colour. The imaged above shows the area of interest where the signal is different if the ligand binds to the CP. The colour code is situated on the upper left of the above image.

PepMV CP sample, in blue, served as a control of signal. The same sample but with 25.D8 show signals of energy transfer between the CP and the ligand (red spectrum). The same result was seen for CTV in green and WMV in purple. For mutated WMV CP R172A we couldn't detect any signal although the sample included the compound. This means that there is no protein-ligand interaction, confirming that the residue R172 in the binding pocket of the protein is required for the interaction with the compound.

We decided to move forward and carry out structural studies for the characterization of the interactions between CPs and selected compounds. The aim was to develop a platform to explore this binding at atomic resolution and face structure-guided selection of new candidates. For this purpose, we need a viral scaffold without the ssRNA that blocks the binding site. We selected PVY VLPs since their structure had been recently revealed at 4Å resolution [99] and their production was easy and accessible. We couldn't use TuMV VLPs because, due to their high flexibility, the resolution achieved in their reconstruction was in the range of 7-8 Å. At this resolution level it is not possible to build an atomic model for the CP-compound interaction.

85 µl of PVY VLP sample was incubated with 15 µl of the compound 25.D8 in PBS buffer and 4 µl of the mixture sample was placed on Quantifoil 2/2 holey carbon grids coated with a thin layer of carbon. Movie frames images were collected in a Titan Krios FEI electron microscope operated at 300kV using a K3 direct detector camera (GATAN) at "The Netherlands Centre for Electron Nanoscopy" (NeCEN) (Fig 4.24). Images were taken in super resolution mode at a nominal magnification of x105,000 with a corresponding pixel size of 0.417 Å/pixel (super resolution pixel size). Micrographs were taken with time exposures of 2.1s in electron counting mode resulting in movies with 40 frames each and a total dose of 40 e^{-}/ $Å^2$. Motion between frames was corrected using MotionCor2 [130] using dose weighting with frames 1 to 40. 2,502 micrographs were obtained. Images were motion corrected and their contrast transfer function was estimated using

CTFFIND3 (Mindell & Grigorieff, 2003). Filaments were manually selected resulting in 45,489 overlapped segments for PVY VLPs with a step size between them of 2 asymmetrical units (85.3 Å) and a size of 400 pixels. Particles were binned by a factor of two, what resulted in particles with a pixel size of 0.834 Å. Image processing of the cryoEM data was done using Relion3 [130]. 2D classification of PVY VLPs revealed a stacked-ring structure (Fig 4.24) in contrast with the helical architecture of TuMV VLPs, as it was previously described [99]. We selected the groups that exhibited good characteristics and discarded the bad ones. That left us with 28,847 segments.

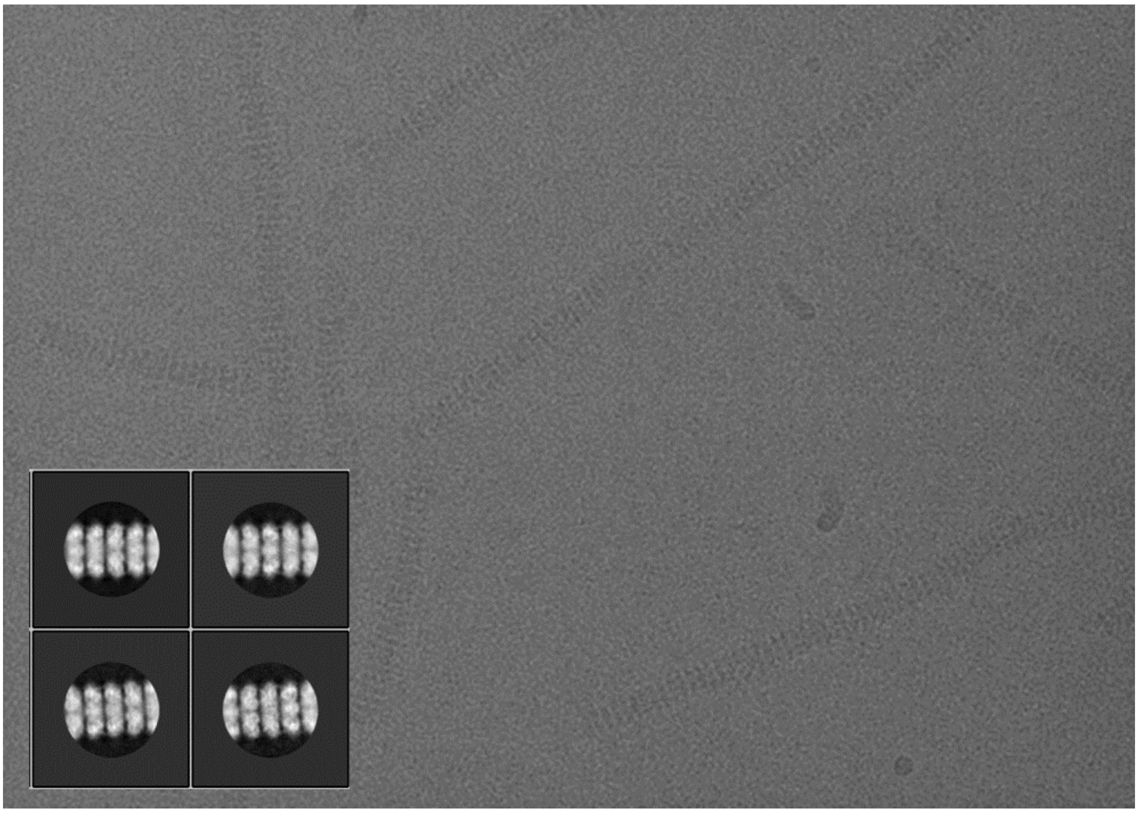


Fig 4.24 Electron micrograph of PVY VLPs sample and 2D classification average. The electron micrograph shows long and well defined filaments, where we could also discern the stacked rings previous to the image processing. On the bottom left four classes out of the 40 classes from 2D classification are shown, where the stacked ring architecture is clear.

After selecting the averaged classes that presented good features we did a 3D refinement, using the map published for PVY VLPs [99] as initial reference (accession code EMBD-0298) low pass filtered at 10 Å. We applied C8 and helical symmetry as each disk contains 8 CP subunits and we also applied helical symmetry, with parameters of twist 13.24˚and rise 42.65 Å. Refinement yielded in a map at 8.34 Å of resolution which was used for creating a mask spanning three disks or layers of the filament. We continued previous refinement from the last iteration but applying the new mask and this work gave us a map at 8.13 Å. The filament presented piled disks with 8 subunits on each of them (Fig 4.25). The fitting of the atomic structure for the PVY VLPs was in register, meaning that our density map was correct (Fig 4.25).

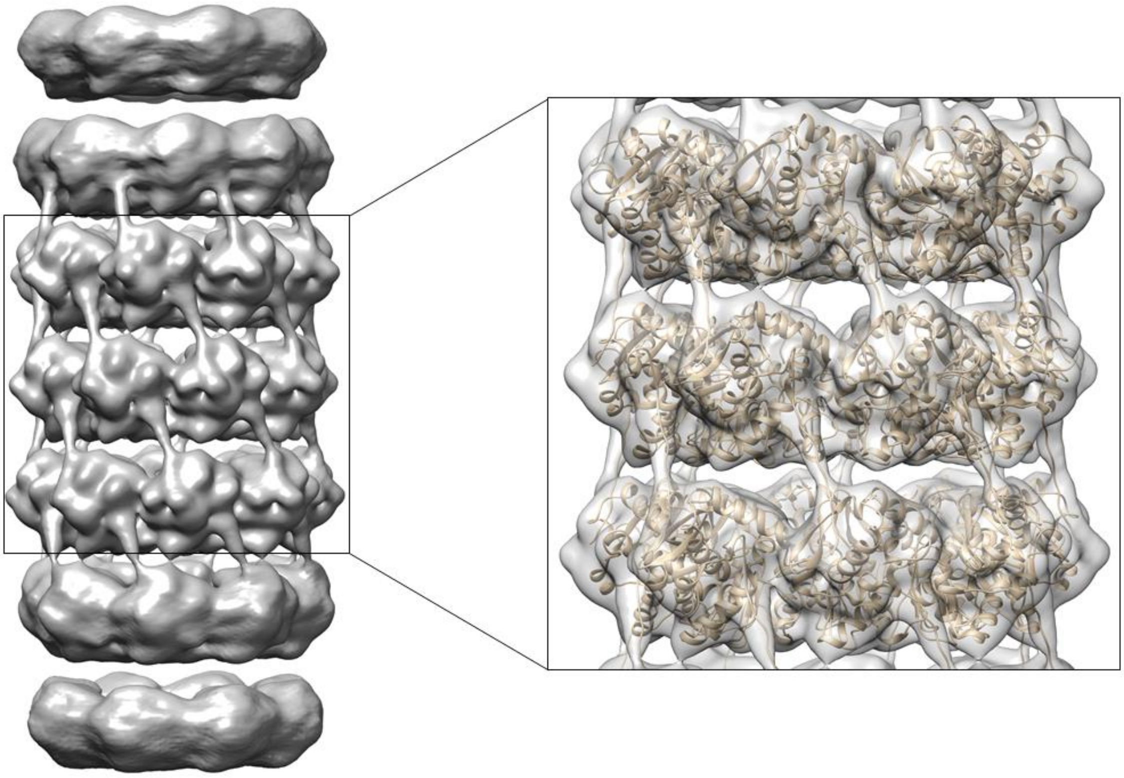


Fig 4.25 Density map for PVY VLPs. The density map on the left shows the whole virion structure with a four layers mask applied at the refinement step, that's why only the four disks in the centre of the map are well defined. The image in the left shows a close up to the map with a transparent surface exhibiting the atomic model for PVY VLPs inside, with all its features in register with our density map.

Our goal wasn't to reconstruct the filament structure but to decipher the nature of the interaction between the VLPs and the compound 25.D8. Assuming that our sample is heterogeneous and that not all the VLPs had included the compound in their binding pockets, we performed a 3D classification in order to separate different populations of particles, those having the compound attached, those who didn't and those that may have additional structural heterogeneity.

We used the previous refined structure as reference and used the same three layers mask as well as the same symmetry parameters. The classification was run without aligning the particles. It yielded 5 classes. Two classes presented most of the particles, class 1 with 37% and class 2 with 41%, meaning that they represented the main populations in the sample. At first sight, both density maps seemed almost identical, but after an exhaustive search of structural differences we saw that in the binding groove of one of the classes there was what seemed to be an extra density. For that reason, we refined both classes and postprocessed them, what resulted in two maps at 6.3 Å and 5.2 Å (Fig 4.26).

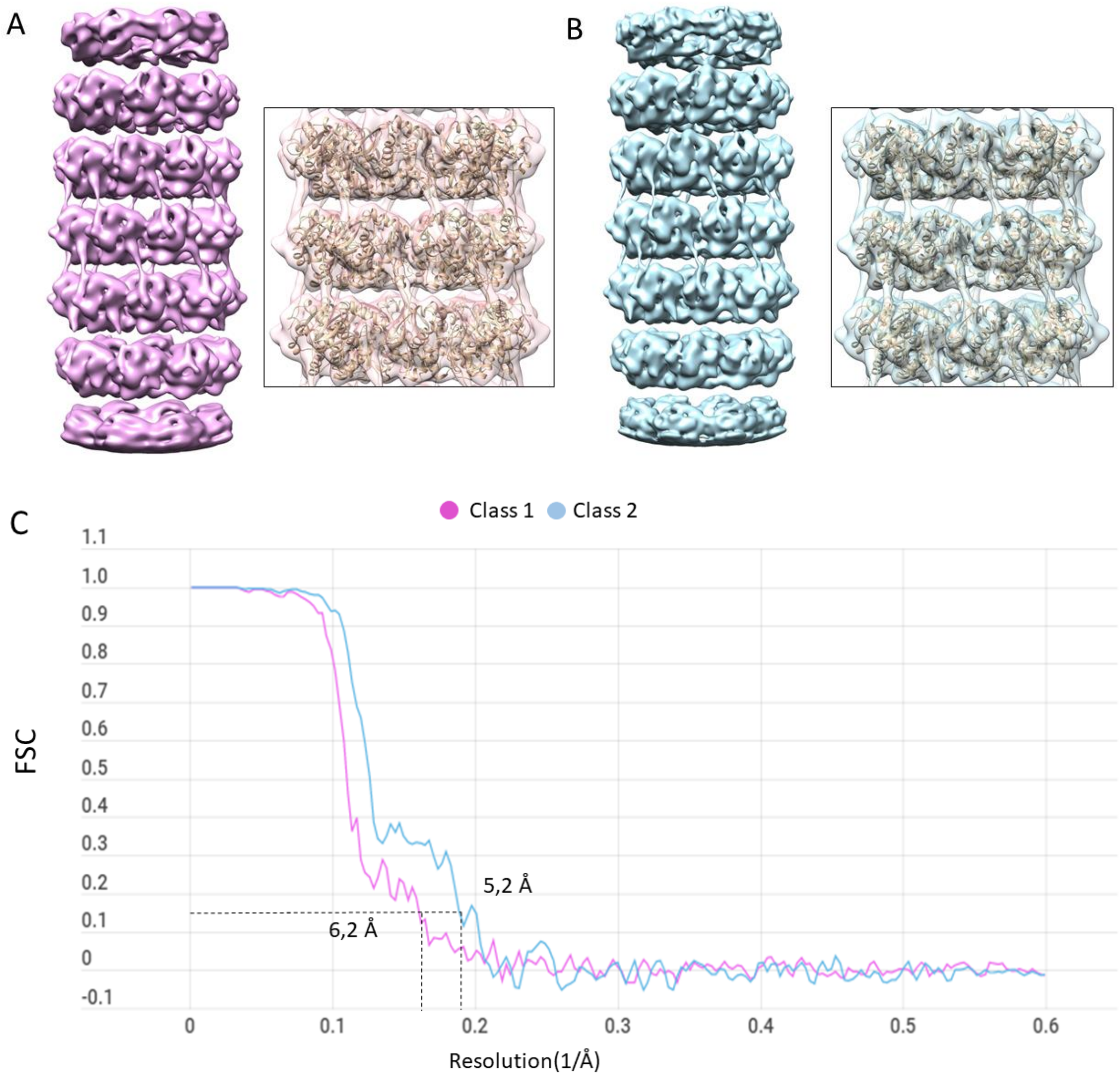


Fig 4.26 Density maps and FSC curves for class 1 and class 2. A and B) Density maps for class 1 in pink **(A)** and class 2 in blue **(B)** with the images on the right side of each map, showing the

transparent density maps with the atomic model for PVY VLPs fitted inside. **C)** FSC curve for both maps, displaying the resolution following the gold-standard method with cutoffs at 0.143.

When fitting the atomic model, we also displayed the three residues conserved in the binding pocket of the CP of PVY in both density maps. Clipping the maps allowed us to see the interior of the filament and compare the two of them. In class 1 the density for C-terminal arm extends to the center of the filament, and the binding groove, where the ssRNA is located in virions, remains empty since VLPs lack the genetic material (Fig 4.27, A). On the contrary, in class 2 we found that the binding groove was not completely empty and that there was a density spanning from the C-terminal arm to the binding pocket were the three conserved residues are located (Fig 4.27, B). This density had a long and thin shape that could agree with the chemical structure for the compound 25.D8 (Fig 4.27, C). The resolution achieved in these reconstructions wasn't enough for an atomic modelling or to distinguish secondary structure elements on the density maps, meaning that we couldn't decipher the nature of the interaction with the compound. Although we tried to improve the resolution of the maps running 3D refinements and classifications with different parameters, using different masks, and also extracting the particles with different step sizes, we didn't achieve any better results that the ones exposed here.

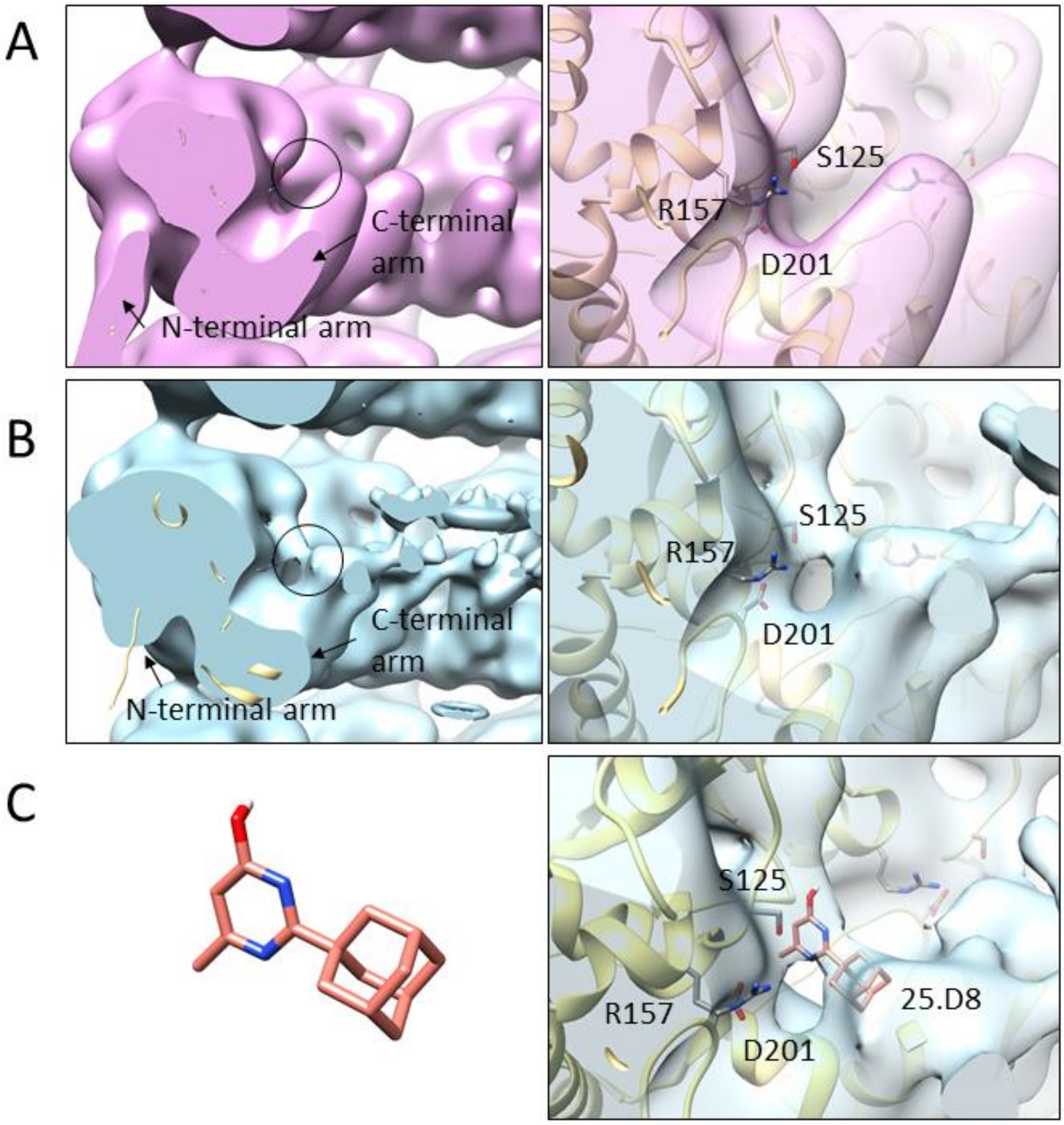


Fig 4.27 Difference between the binding pocket of Class 1 and Class 2 and fitting of the chemical structure of 25.D8. The transversally clipped density maps show how the C-terminal arm goes into the center of the filament and the binding groove where the ssRNA is placed in the virions. Class 1 **(A)** has the binding groove empty but Class 2 **(B)** has an extra density that could be related with the compound 25.D8 **(C)** that goes from the C-terminal arm to the serine 125 in the binding pocket of the CP.

The fact that each layer of the VLPs is poorly connected with the ones above and below make them highly flexible what is a huge impediment for cryoEM reconstruction. Multibody refinement is an option available in Relion that allows the exploration of movements between different components of the macromolecule of interest. In our case, we wanted to explore if the different layers that form the filament move relative to each other, changing the helical parameters, what will lead to worst reconstructions since the helical symmetry will not be equal all along the filament. Using the first reconstructed map of the PVY VLPS, we created five different masks to explore the movements between them. The first mask (Mask 1) was masking the top part of the filament. Three masks (Mask 2, Mask 3 and Mask 4) were masking the three middle layers of our reconstructed filament and the fifth mask (Mask 5) was used in the bottom part of the filament (Fig 4.28).

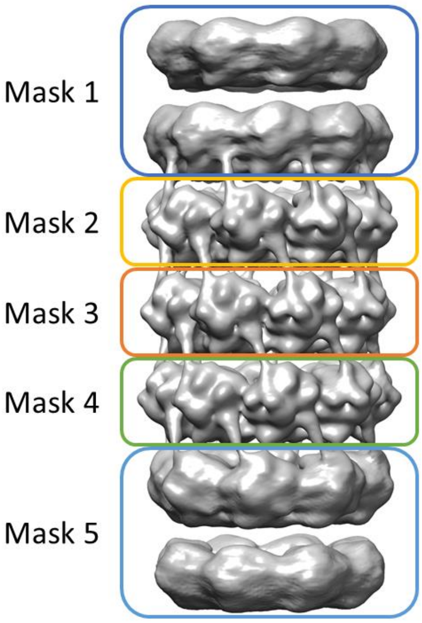


Fig 4.28 Schematic representation of the five masks used for the multibody refinement in the PVY VLP filament density map. The image above shows the density map reconstructed for the PVY VLP with the three middle layers well defined. Each mask used for the multibody refinement is represented as a colored box around the density that will be masked. Mask 1 is masking the upper density, represented in a deep blue contoured box. Mask 2 is masking one defined layer, in a yellow box. Mask 3, orange box masking the middle well defined layer. Mask 4, in green, masking the bottom defined layer and Mask 5, in light blue, is masking the bottom part of the filament.

The Multibody refinement yielded in two principal motions that represent the principal movements occurring in the filaments. Those principal motions explain the particle variability in the sample. The first motion was a notorious side to side movement in the longitudinal axis of the filament (Fig 4.29 A, B, C). What this movement is telling us is that, although straight parts of the filament were selected in the manual picking, the filaments are bended in a different plane that cannot be seen in the micrographs. The other movement displayed is a twist between the different layers of the filament (Fig 4.29 D, E, F). This movement may be breaking the helical symmetry as the twist and rise parameters can be different for each particle, since the layers could be moved between each other in different particles that are contributing to the reconstruction.

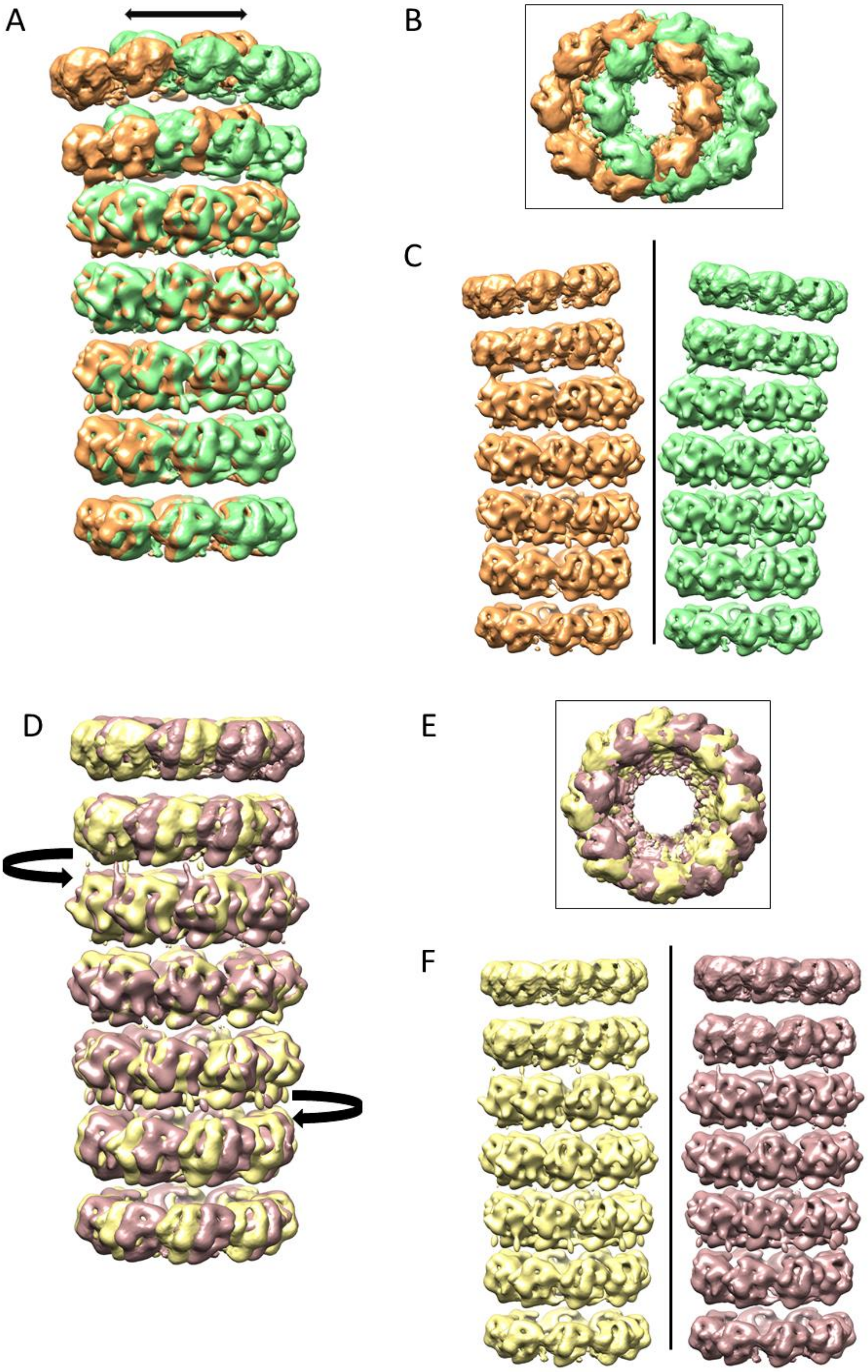


Fig 4.29 Density maps extracted from the Multibody refinement, corresponding to the different movements present in the filament. A) This image shows both most extreme density maps of the side to side movement of the filament, one in light green and the

other in light orange. The black arrow at the top shows the movement that the filament does. **B)** Top view of A where it can be seen the movement from another perspective. **C)** Same both maps of A and B but separated allowing the visualization of the movement of the top part of the filament. **D)** This image shows both extreme density maps corresponding to the twist movement between the different layers of the filament. One density map is displayed in light yellow and the other one in rose gold. Both black arrows are showing the twist movement that happens between layers. **E)** Top view of the twist movement. Here the twist movement is clear as there are monomers occupying the empty space that should be between monomers. **F)** Both density maps separated.

These movements corroborate the high flexibility of the structure and could explain the difficulty to reach a high-resolution reconstruction.

Chapter 5:

DISCUSSION

5.1 STRUCTURE OF FILAMENTOUS PLANT VIRUSES

All filamentous plant viruses have a helical arrangement with a very strong helical symmetry. This high symmetry is of great advantage for structure determination by X-ray crystallography or cryoEM. Filamentous plant viruses can be rigid rod-shaped filaments or flexible filaments. Flexibility entails a great issue for structural determination and the structures for flexible filamentous plant viruses couldn't be resolved until few years ago. The first structure ever resolved for a virus was the one for TMV virions using fiber diffraction in the late 70's [131–134]. TMV virions are filamentous rod-shaped plant viruses with a highly symmetric helical architecture. This feature makes them the perfect samples for X-ray crystallography experiments. X-ray crystallography was the main method for structural determination, but flexible samples couldn't be resolved this way, and it wasn't until the development of cryoEM and helical reconstruction algorithm methods that the first structure for a flexible filamentous virus was resolved. The most important step of helix reconstruction is the determination of the helical parameters. These parameters describe how the subunits/elements that form the helix are related between each other and how the helix is placed and organized in space.

In our work we present the structure for TuMV virions which have helical parameters of 8.8 subunits per turn of the helix and a pitch of 32.5 Å. These results agree with the parameters for other members of the *Potyviridae* family as the ones for WMV [98] and PVY [99] which share the same number of subunits per turn. The pitch is slightly different for PVY virions which is 34.7 Å [99]. Soybean mosaic virus (SMV), another potyvirus was studied at low resolution and its helical parameters were set as 8.8 subunits per turn and a pitch of 35 Å [135]. This almost identical parameters reveal that the overall helical structure of potyviruses is conserved along the whole family. But not only potyviruses have been demonstrated to display that architecture, also members of the genus potexvirus share very similar architecture. Potexvirus are also flexible filamentous plant viruses and are the second genus of plant viruses economically more relevant. PVX is the archetypical member for its genus and also has a helical structure with helical parameters of 8.7 subunits per turn and 34.6 Å [101], the same parameters as for PepMV

[31]. The structure for another potexvirus, BaMV revealed a pitch of 35 Å and 8.8 subunits per turn [100] and very similar parameters for Alternanthera mosaic virus (AltMV) with 8.7 subunits per turn and 35.7 Å [136].

We have also presented the atomic model for the CP of TuMV virions. This structure shows a helical core formed by 7 α-helices and two long unstructured terminal arms which agrees with the structures of other two potyviruses studied at atomic resolution, WMV and PVY [98,99]. The helical core shares structural homology also with potexviruses as PepMV [31] meaning that both genus of flexible plant viruses may share a common ancestor.

In TuMV virions, as for WMV and PVY, the N-terminal arm participates in hydrophobic side-by-side interactions with a groove in the neighbouring subunit. Moreover, after a 90° turn, it reaches another subunit in the next turn of the helix and interacts with it via electrostatic interactions. This axial interaction after the 90° turn of the N-terminal arm seems to be a signature of potyviruses. Potexviruses lack the 90° turn and only present the side-by-side interactions as it has been seen for PepMV, PapMV and PVX virions [31,32,101]. In PapMV virions, the deletion of 26 [137] or 13 [138] residues from the N-terminal abolishes the CP-CP interaction and makes the protein assembly incompetent which strongly suggests that the hydrophobic interactions participated by residues at the N-terminal region are the underlying mechanism of polymerization of flexible filamentous viruses [139]. The same was seen for PVY and TuMV VLPs. For PVY, the deletion of 49, 50 and 69 N-terminal residues prevented filament assembly *in vitro* and *in vivo*. Besides, the 50 residues N-terminal deletion mutants weren't able to systemically infect plants [99]. The deletion of 31 N-terminal residues in TuMV VLPs had the same result, proposing that interactions taking place between the N-terminal and the core domain are essential for viral assembly [140].

The C-terminal arm of TuMV couldn't be fully traced within the cryoEM map, and the atomic model lacks the last 16 residues. This arm, in all potyviruses and potexviruses, is located inside the filament and it interacts with subunits above and below in the axial plane. In PVY virions the C-terminal could be completely traced and it was seen that it is fully shielded inside the virion, thus the C-terminal region can only be exposed to the solvent at the RNA 3' end or in damaged virions [99]. Potyviral terminal arms are longer than the potexvirus ones and it results in an increased CP-CP interaction network, involving 12 neighbouring subunits in potyviruses compared to 8 in potexviruses [32]. Altogether, C- and N-terminal arms of potyviruses and potexviruses seem to be the key of flexible filamentous virion assembly. Their flexible links between CP subunits allow relative movements between CPs, explaining the overall flexibility of the virions.

But not only deciphering the interaction between CPs is important to understand the nature of flexible virus assembly, but also the interaction with the ssRNA plays a crucial role. In TuMV the ssRNA sits in the binding pocket of each CP, which spans for five nucleotides of the viral genome. The binding site is placed in a groove between the helix H7 of the helical core of the protein and the C-terminal arm. Four of the five nucleotides are facing the inner side of the filament but one of them is deep inside the binding site of the CP subunit and surrounded by three residues, S146, R178 and D222. These residues have been found in the same place in PepMV, WMV, PVY and PVX [31,98,99,101]. The binding site seems to be conserved in the CPs from four families of flexible filamentous plant viruses although they do not share significant sequence homology [98].

With all this information in hand, it is clear that all of the flexible filamentous plant viruses are structurally related.

5.2 STRUCTURE OF TuMV VLPS

The study of VLPs has been widely used for different nanotechnological applications, including vaccine carriers and drug delivery systems [68]. The first plant VLPs were constructed from TMV and were obtained after structural gene expression in bacterial cells [141]. Afterwards more VLPs of filamentous viruses have been formed *in vitro* and *in planta* but no structural data was available until few years ago. In this work we presented the structure for TuMV VLPs at low resolution. We reconstructed two maps, both with helical arrangement and defined CPs. CryoEM data at 13 Å resolution for AltMV VLPs showed filaments following a helical architecture [136]. Although we couldn't build an atomic model for TuMV VLPs we can see some changes in the CP structure. It seems that in the absence of viral RNA, the helix that delimits the binding groove is moved towards the inner side of the filament, presumably occupying the gap left by the genome. This CP movement without the viral RNA was also proposed for AltMV VLPs, although there is no atomic model to support this idea [136]. To date, the only atomic model available for a flexible filamentous VLP is the one for PVY VLPs. Interestingly they found that their filaments do not present a helical architecture as TuMV or AltMV virions, but they display an architecture of stacked octameric rings [99]. The overall structure of the CP is maintained at the core of the protein, but its flexible nature allows for some inner movements when there is no binding partner. In PVY VLPs the RNA binding loop is moved towards the inner site of the filament, as we have observed for our TuMV VLPs. The difference in the architecture between the helical TuMV and AltMV VLPs and for the stacked PVY VLPs is really surprising. While TuMV VLPs have been produced *in planta*, AltMV and PVY VLPs were produced *in vitro* and in *E.coli,* respectively. PVX and PapMV CPs do not assemble *in vitro* without the presence of RNA, so it seems that virion-RNA interactions are crucial for the assembly of both potexviruses [101,142]. In PVX virions Asn51 has been proposed as the residue that may be crucial for the assembly. It is placed in the N-terminal arm of the protein and interacts with both the ssRNA and the adjacent CP. This dual role may be responsible of the impossibility of the CPs to assemble if the RNA is not present. AltMV VLPs have been formed without the presence of RNA but as there are no atomic models for this virion we cannot unravel why members from the same genus do not behave equally under similar conditions. Nevertheless, potyviruses as TuMV and PVY have shown to be able to assemble without a guiding RNA [99]. It could be interesting to compare differences between both genera in terms of their assembly without a guiding RNA, but for that task more structural data is needed.

As opposed to icosahedral viruses, in helical viruses the genetic material is bound to copies of the viral nucleoprotein or CP along the entire genomic length, and each nucleoprotein subunit interacts with the genome. Thus, the absence of the nucleic acid in VLPs is expected to modify the entire structure. The helical arrangement of TuMV VLPs allows the comparison of their structure with TuMV virions, and had shown that the interaction with the ssRNA in between subunits govern the network of contacts between

CPs mediated by N-terminal arms that play as molecular staples, and that these interactions are lost in the absence of the nucleic acid.

In the last years, plant VLPs have been widely studied for their applications in biomedicine and biotechnology, so having structural knowledge about them will help to improve and refine these types of technologies. As VLPs contain CPs that are repeated hundreds of times along the filament it has been seen that they can induce strong immune responses in mammalian organisms, so they seem to be good candidates for vaccine production strategies. Some plant virus VLPs have been already tested for safety and efficacy in human clinical trials [71].

5.3 TOMOGRAPHY OF TuMV VIRAL ENDS

Flexible filamentous viruses are known to have a rod-shaped helical structure formed by the same CP along the whole virion and protecting the ssRNA inside. Viral RNAs have a methyl-7-guanosine cap (m^7G cap) on their genomic 5’ end that allows them to bind the eukaryotic translation machinery, but some plant viruses, as members of the families *Secoviridae, Potyviridae, and Luteoviridae*, instead, present a viral encoded protein, the VPg [143] covalently attached. The structure for the potyviral VPg has been resolved recently [57] but there is no structural data available of the VPg attached to the filament since potyviral ends still remain structurally uncharacterized.

In our work we present a preliminary low-resolution study of both TuMV filament ends using cryoET. We have seen that there are two populations of particles and both display very different characteristics. One of the populations present a blunt end, and the other one has a protruding helical-like structure that is smaller in diameter than the virion itself. This protruding tips have also been seen for other potyviruses and also for other flexible filamentous viruses from different families [129,144]. AFM studies of Beet Yellows Virus (BYV), potato virus A (PVA) and PVY have demonstrated that these viruses have a similar tip at one of the ends of the filaments [129,144]. On one hand, for BYV the tip seemed to be subdivided in three segments and the authors proposed that it might be formed by the three viral-encoded proteins, Hsp70h, p64 and p20 [144]. On the other hand, PVY and PVA tips were proposed to be made up by the two potyviral proteins VPg and HC-Pro [129]. These works were published before the first atomic structure for a flexible filamentous virus was available, so they were merely based on the knowledge about the potyviral genome and assumptions of how the polarity of the genome was inside the filaments. Nowadays, we have several atomic models for potyviral and potexviral CP structures. In all the available structures the C-terminal arms run straight through the inner part of the filament, protected from the solvent. They are completely hidden until they reach the 3' end of the genome, where they could be exposed to the outside, constructing a helical-like structure. This structure resembles those structures seen by AFM for other filamentous viruses and also the protruding tip seen by cryoET in one of the TuMV ends. We propose that the protruding helical tip at the end of the virions corresponds to the C-terminal arms that protrude from the inside of the filaments.

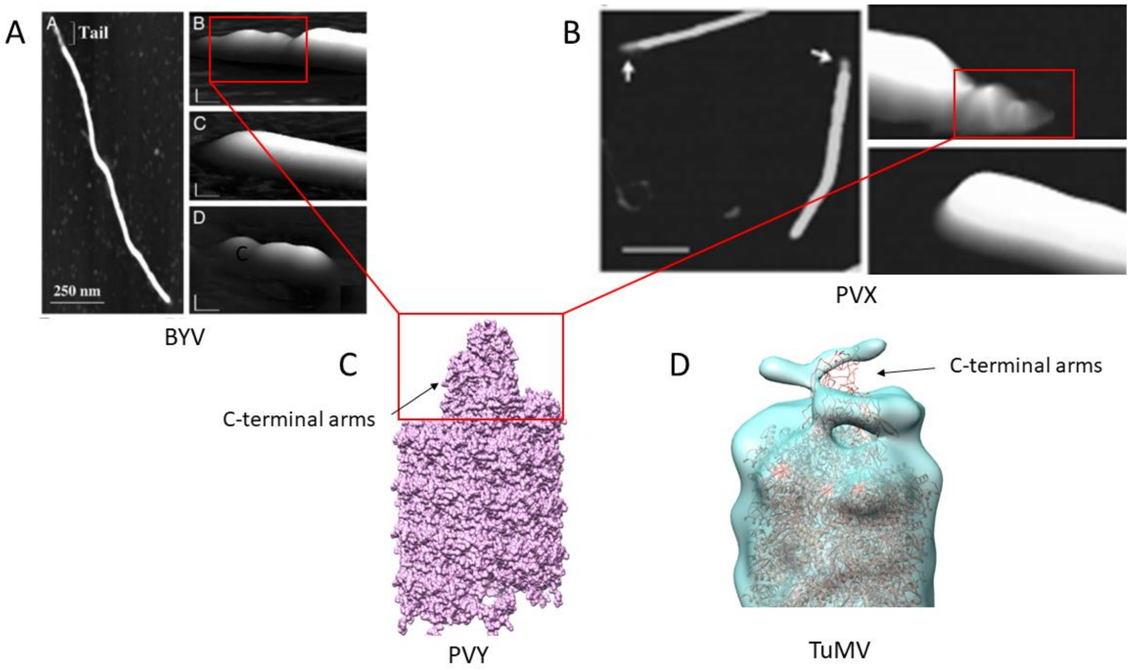


Fig 5.1 Comparison between the AFM studies of BYV and PVX with the cryoEM studies of PVY virions and our own tomography results. The above images show **A)** AFM studies of BYV virion ends. Filaments show a narrow tail with three segments at one of the ends. **B)** AFM studies of PVX in which a similar tail is seen in one of the tips. In this case, the tail had a helical structure instead of a structure of three segments. The other extreme of the virions was blunt. **C)** Atomic model for a multimer of PVY CPs where the C-terminal arms form a protruding helical-like structure at one end of the virions that corresponded with the genomic 3' end. **D)** Our own

tomographic density map for one of the TuMV viral ends with the atomic model for PVY CPs fitted inside. The density map displays a helical-like structure at the end which is narrower than the filament. The fitting seems to be in register, and the C-terminals could be forming the helical-structure at the end of the virions.

Therefore, the blunt end of the other population of particles may correspond to the 5’-terminal end of the genome. In our tomographic studies we have seen that this blunt end fits well with a last turn of the helix, but there is an extra density at the tip that does not display the globular shape of the CP. PVY VPg structure has been solved by NMR [57] and it has an arrow-head shape structure with two unstructured arms. It is known that the VPg of plant viruses is covalently attached to the 5’ terminal uridine residue of genomic RNA via tyrosine or serine residues [143]. The long shape of the potyviral VPg body fits perfectly with the density at the end of our map and also, in that position, the residue Y64, responsible for the interaction between the protein and the viral genome in PVY virions, is placed at a distance that could make the interaction possible. Our resolution is limited, and we need to increase the number of subtomograms to lead to better reconstructions.

5.4 EVALUATION OF PVY VLPs AS PLATFORM FOR STRUCTURAL STUDIES IN COMPOUND SCREENING

According to Food and Agriculture Organization of the United Nations (FAO) more than the 60% of the world´s population depends on agriculture to make a living. Besides, the rapidly growing population on the planet will be accompanied with a need of increasing the number of agricultural crops in order to produce enough food. In this regard, plant viral diseases suppose a great threat for plant farming. Plant viruses are spread all around the globe and are responsible for great economic losses in agriculture, therefore there is a great need of finding sustainable disease management approaches [34]. Nowadays there are several strategies for controlling plant virus infection as: basic hygienic methods; use of pesticides against vectors transmitting the virus; use of resistant plants selected by traditional methods or transgenic plants; use of certified plants that guarantee the absence of the pathogen; methods for detecting the virus; elimination of infected plants; and growing plants in a controlled *in vitro* environment. Some of these methods are not well accepted socially and environmentally. The most successful method for disease control is the combination of *in vitro* plant breeding, where plant tissue is grown in a sterile environment, and the production of plants certified as pathogen-free. Nevertheless, the effective control does not exist and there still a clear need to develop effective chemical treatments for the *in vitro* production of virus free plants.

In our work we have found a chemical compound that binds the CP of various flexible plant viruses from different families. We were able to see, by NMR studies, that the binding site involves residues at the RNA binding pocket of the CPs where the three conserved amino acids are located. We tried to unravel the nature of the interaction using cryoEM. The reconstructed structures allowed us to see some extra density that may correspond to the compound, but the medium resolution of the reconstruction made impossible to describe the interaction. In the work of Andreja and colleagues [99] they reconstructed the PVY VLPs at 4 Å resolution. Although our images are of good quality and the power spectra clearly shows information at high resolution (Fig 5.2), the difference between their results and ours could be due to two causes. In their work they had a dataset of 148,876 particles in contrast with our 28,847 particle dataset. Besides, in our work we have a chemical compound attached to the VLPs that may be affecting the stability of the sample and increasing its heterogeneity.

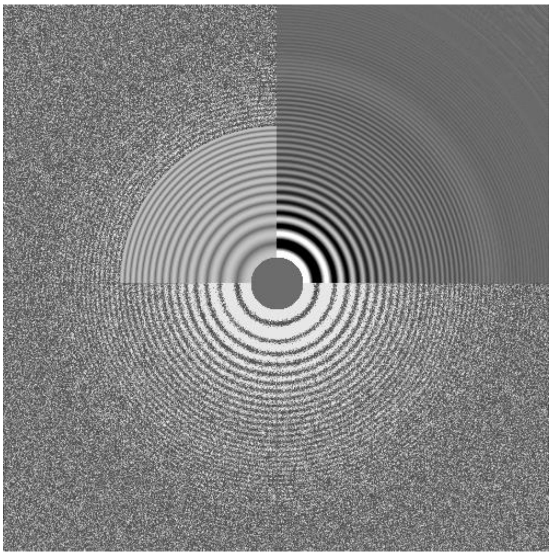


Fig 5.2 Power spectra of one micrograph selected for particle picking of the PVY VLP sample. The power spectra of the image is a representative one of the quality of the images of this dataset. The information available in these micrographs goes up to 2.6 Å resolution.

Moreover, a deeper research on the flexibility of the filaments unraveled two principal motions occurring in the filament structure. There is side to side movement in the longitudinal axis of the filament and although straight parts of the filaments were selected in the manual picking, filaments aren't straight in a different plane not visible from the 2D micrographs. The other movement is a twist motion between the different layers of the filament. These motions can be breaking the helical symmetry of the filament and the movements could be the cause of the limited resolution achieved in the reconstruction. To improve the resolution we will need to increase the initial dataset. As an alternative method in our lab we have also grown crystals of WMV CP with the compound to be studied by X-ray crystallography. Any of these two approaches may be the tool to unravel the nature of the interaction between compound candidates and the viral CPs.

CHAPTER 6:
CONCLUSIONS

TuMV and VLPs

1. In this work the structure of TuMV, a member of the family *Potyviridae* has been resolved at 5 Å resolution. TuMV virions show a left-handed helix with 8.8 CP per turn, in full agreement with the published data for other potyviruses.
2. An atomic model for the CP and for the ssRNA of TuMV has been generated thanks to the density map obtained and taking as a guide the atomic model published for WMV virions. TuMV structure includes the universally conserved pocket in flexible filamentous plant viruses formed by amino acids Ser, Arg and Asp.
3. The structure for the VLPs of TuMV has also been studied, this time at a resolution between 7 and 8 Å. The absence of ssRNA produces labile multimers with distorted local regions along the filaments, but VLPs maintain their helical symmetry with parameters identical to those of TuMV.
4. Absence of the ssRNA provokes a reorganization of the interaction network between subunits, altering the interaction between neighbor CPs and the structure of the CP itself.

CryoET and subtomogram averaging of TuMV viral ends

5. With this work we have seen that potyviruses have different structural organizations at both their ends. We have found that the 5’ terminal end has an arrow-head shaped structure that we attribute to the VPg protein.
6. The 3’ terminal end presents a helix-like conical density that we propose it corresponds to the C-terminal regions of the viral CPs that protrude from the inner side of the filaments at the last turn of the helix.

Structural studies of a chemical interaction between a selected compound and PVY VLPs

7. The NMR studies demonstrated that the compound 25.D8 interacts with the conserved binding site of the CP with the ssRNA.
8. *In planta* preliminary studies have shown that the compound may be interfering the infection of potyviruses in host leaves.
9. Cryo-EM studies have shown that there might be an interaction taking place at the ssRNA binding site of the CP. Nevertheless, the resolution achieved wasn’t enough to claim this statement.
10. Multibody refinement analysis revealed two principal motions on the structure of the VLPs that may be altering the helical symmetry of the filament and consequently limiting the resolution of the cryoEM study.

BIBLIOGRAPHY

1. Abrescia NGA, Bamford DH, Grimes JM, Stuart DI. Structure unifies the viral universe. *Annu Rev Biochem*. Published online 2012. doi:10.1146/annurev-biochem-060910-095130

2. Lustig A, Levine AJ. One hundred years of virology. *J Virol*. Published online 1992. doi:10.1128/jvi.66.8.4629-4631.1992

3. Oldstone MBA. History of virology. In: *Encyclopedia of Microbiology*. Elsevier; 2019:608-612. doi:10.1016/B978-0-12-801238-3.00078-7

4. Barrass S V., Butcher SJ. Advances in high-throughput methods for the identification of virus receptors. *Med Microbiol Immunol*. Published online 2020. doi:10.1007/s00430-019-00653-2

5. Poranen MM, Daugelavičius R, Bamford DH. Common principles in viral entry. *Annu Rev Microbiol*. Published online 2002. doi:10.1146/annurev.micro.56.012302.160643

6. Xu J, Xiang Y. Membrane Penetration by Bacterial Viruses. *J Virol*. Published online 2017. doi:10.1128/jvi.00162-17

7. Cann AJ. Replication of Viruses. In: *Encyclopedia of Virology*. ; 2008. doi:10.1016/B978-012374410-4.00486-6

8. Speck SH, Ganem D. Viral latency and its regulation: Lessons from the γ-Herpesviruses. *Cell Host Microbe*. Published online 2010. doi:10.1016/j.chom.2010.06.014

9. Krupovic M, Dolja V V., Koonin E V. Origin of viruses: primordial replicators recruiting capsids from hosts. *Nat Rev Microbiol*. 2019;17(7):449-458. doi:10.1038/s41579-019-0205-6

10. Koonin E V. Viruses and mobile elements as drivers of evolutionary transitions. *Philos Trans R Soc B Biol Sci*. Published online 2016. doi:10.1098/rstb.2015.0442

11. Buchmann JP, Holmes EC. Cell Walls and the Convergent Evolution of the Viral Envelope. *Microbiol Mol Biol Rev*. 2015;79(4):403-418. doi:10.1128/mmbr.00017-15

12. Louten J. Virus Structure and Classification. In: *Essential Human Virology*. ; 2016. doi:10.1016/b978-0-12-800947-5.00002-8

13. Goldsmith CS, Miller SE. Modern uses of electron microscopy for detection of viruses. *Clin Microbiol Rev*. Published online 2009. doi:10.1128/CMR.00027-09

14. Lin YR, Lin CS. Genome-wide characterization of vibrio phage ϕ{symbol}pp2 with unique arrangements of the mob-like genes. *BMC Genomics*. Published online 2012. doi:10.1186/1471-2164-13-224

15. Häring M, Rachel R, Peng X, Garrett RA, Prangishvili D. Viral Diversity in Hot Springs of Pozzuoli, Italy, and Characterization of a Unique Archaeal Virus, Acidianus Bottle-Shaped Virus, from a New Family, the Ampullaviridae. *J Virol*. Published online 2005. doi:10.1128/jvi.79.15.9904-9911.2005

16. Fauquet CM. Taxonomy, Classification and Nomenclature of Viruses. In: *Encyclopedia of Virology*. ; 2008:9-23. doi:10.1016/B978-012374410-4.00509-4

17. Baltimore D. *Expression of Animal Virus Genomes*. Vol 35.; 1971. Accessed March 24, 2021. http://mmbr.asm.org/

18. Simmonds P, Aiewsakun P. Virus classification – where do you draw the line? *Arch Virol*. Published online 2018. doi:10.1007/s00705-018-3938-z

19. Walker PJ, Siddell SG, Lefkowitz EJ, et al. Changes to virus taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2019). *Arch Virol*. Published online 2019. doi:10.1007/s00705-019-04306-w

20. Nasir A, Caetano-Anollés G. Identification of capsid/coat related protein folds and their utility for virus classification. *Front Microbiol*. Published online 2017. doi:10.3389/fmicb.2017.00380

21. Singh S, Awasthi LP, Jangre A. Transmission of plant viruses in fields through various vectors. In: *Applied Plant Virology*. Elsevier; 2020:313-334. doi:10.1016/b978-0-12-818654-1.00024-4

22. Lefeuvre P, Martin DP, Elena SF, Shepherd DN, Roumagnac P, Varsani A.

Evolution and ecology of plant viruses. *Nat Rev Microbiol*. 2019;17(10):632-644. doi:10.1038/s41579-019-0232-3

23. Hogenhout SA, Ammar ED, Whitfield AE, Redinbaugh MG. Insect vector interactions with persistently transmitted viruses. *Annu Rev Phytopathol*. 2008;46(1):327-359. doi:10.1146/annurev.phyto.022508.092135

24. Pirone TP, Perry KL. Aphids: Non-persistent transmission. *Adv Bot Res*. 2002;36:1-19. doi:10.1016/s0065-2296(02)36056-7

25. Insect vectors ~ ViralZone. Accessed May 7, 2021. https://viralzone.expasy.org/3738

26. Ng JCK, Perry KL. Transmission of plant viruses by aphid vectors. *Mol Plant Pathol*. 2004;5(5):505-511. doi:10.1111/J.1364-3703.2004.00240.X

27. Scholthof KBG, Adkins S, Czosnek H, et al. Top 10 plant viruses in molecular plant pathology. *Mol Plant Pathol*. 2011;12(9):938-954. doi:10.1111/j.1364-3703.2011.00752.x

28. Plant viruses ~ ViralZone. Accessed April 19, 2021. https://viralzone.expasy.org/231

29. Namba K, Pattanayek R, Stubbs G. Visualization of protein-nucleic acid interactions in a virus. Refined structure of intact tobacco mosaic virus at 2.9 Å resolution by X-ray fiber diffraction. *J Mol Biol*. 1989;208(2):307-325. doi:10.1016/0022-2836(89)90391-4

30. Namba K, Stubbs G. Structure of tobacco mosaic virus at 3.6 Å resolution: Implications for assembly. *Science (80-)*. 1986;231(4744):1401-1406. doi:10.1126/science.3952490

31. Agirrezabala X, Méndez-López E, Lasso G, Sánchez-Pina MA, Aranda M, Valle M. The near-atomic cryoEM structure of a flexible filamentous plant virus shows homology of its coat protein with nucleoproteins of animal viruses. *Elife*. Published online 2015. doi:10.7554/eLife.11795

32. Dimaio F, Chen C-C, Yu X, et al. The molecular basis for flexibility in the flexible

filamentous plant viruses. *Nat Publ Gr*. Published online 2015. doi:10.1038/nsmb.3054

33. Mumford RA, Macarthur & R, Boonham & N. The role and challenges of new diagnostic technology in plant biosecurity. Published online 2015. doi:10.1007/s12571-015-0533-y

34. Jones RAC, Naidu RA. Global Dimensions of Plant Virus Diseases: Current Status and Future Perspectives. *Annu Rev Virol*. 2019;6(1):387-409. doi:10.1146/annurev-virology-092818-015606

35. Patil BL. Plant Viral Diseases: Economic Implications. In: *Encyclopedia of Virology*. Elsevier; 2021:81-97. doi:10.1016/b978-0-12-809633-8.21307-1

36. Rivas-Valencia P, Loeza-Kuk E, Domínguez-Monge S, Lomas-Barrié CT. Infección crónica del virus de la tristeza de los cítricos en árboles de Citrus sinensis / C. aurantium en un régimen térmico restrictivo en Yucatán. *Rev Chapingo, Ser Hortic*. 2017;23(3):187-202. doi:10.5154/r.rchsh.2016.11.028

37. García JA, Glasa M, Cambra M, Candresse T. Plum pox virus and sharka: A model potyvirus and a major disease. *Mol Plant Pathol*. 2014;15(3):226-241. doi:10.1111/mpp.12083

38. Varveri C, Maliogka VI, Kapari-Isaia T. Principles for supplying virus-tested material. In: *Advances in Virus Research*. Vol 91. Academic Press Inc.; 2015:1-32. doi:10.1016/bs.aivir.2014.10.004

39. Yancheva S, Kondakova V. Plant Tissue Culture Technology: Present and Future Development. In: ; 2018:39-63. doi:10.1007/978-3-319-54600-1_16

40. Gibbs AJ, Hajizadeh M, Ohshima K, Jones RAC. The Potyviruses: An Evolutionary Synthesis Is Emerging. *Viruses*. 2020;12(2):132. doi:10.3390/v12020132

41. Wylie SJ, Adams M, Chalam C, et al. ICTV virus taxonomy profile: Potyviridae. *J Gen Virol*. Published online 2017. doi:10.1099/jgv.0.000740

42. Gadhave KR, Gautam S, Rasmussen DA, Srinivasan R. Aphid transmission of potyvirus: The largest plant-infecting RNA virus genus. *Viruses*. Published online

2020. doi:10.3390/v12070773

43. Ivanov KI, Eskelin K, Lõhmus A, Mäkinen K. Molecular and cellular mechanisms underlying potyvirus infection. *J Gen Virol*. 2014;95(PART 7):1415-1429. doi:10.1099/vir.0.064220-0

44. Revers F, García JA. Molecular biology of potyviruses. In: *Advances in Virus Research*. ; 2015. doi:10.1016/bs.aivir.2014.11.006

45. Ng Yi N G Sha N HO, Pasin F, Oa I, et al. Truncation of a P1 leader proteinase facilitates potyvirus replication in a non-permissive host. doi:10.1111/mpp.12640

46. Pasin F, Simón-Mateo C, García JA. The Hypervariable Amino-Terminus of P1 Protease Modulates Potyviral Replication and Host Defense Responses. *PLoS Pathog*. 2014;10(3). doi:10.1371/journal.ppat.1003985

47. Valli AA, Gallo A, Rodamilans B, López-Moya JJ, García JA. The HCPro from the Potyviridae family: an enviable multitasking Helper Component that every virus would like to have. *Mol Plant Pathol*. Published online 2018. doi:10.1111/mpp.12553

48. Valli A, Gallo A, Calvo M, Perez J d. J, Garcia JA. A Novel Role of the Potyviral Helper Component Proteinase Contributes To Enhance the Yield of Viral Particles. *J Virol*. Published online 2014. doi:10.1128/jvi.01010-14

49. Luan H, Shine MB, Cui X, et al. The Potyviral P3 Protein Targets Eukaryotic Elongation Factor 1A to Promote the Unfolded Protein Response and Viral Pathogenesis 1 [OPEN]. doi:10.1104/pp.16.00505

50. Wen RH, Hajimorad MR. Mutational analysis of the putative pipo of soybean mosaic virus suggests disruption of PIPO protein impedes movement. *Virology*. 2010;400(1):1-7. doi:10.1016/j.virol.2010.01.022

51. Cui H, Wang A. Plum Pox Virus 6K1 Protein Is Required for Viral Replication and Targets the Viral Replication Complex at the Early Stage of Infection . *J Virol*. 2016;90(10):5119-5131. doi:10.1128/jvi.00024-16

52. Deng P, Wu Z, Wang A. The multifunctional protein CI of potyviruses plays interlinked and distinct roles in viral genome replication and intercellular movement. Published online 2011. doi:10.1186/s12985-015-0369-2

53. Wei T, Zhang C, Hong J, et al. Formation of complexes at plasmodesmata for potyvirus intercellular movement is mediated by the viral protein P3N-PIPO. *PLoS Pathog*. 2010;6(6). doi:10.1371/journal.ppat.1000962

54. Movahed N, Patarroyo C, Sun J, Vali H, Laliberté JF, Zheng H. Cylindrical inclusion protein of turnip mosaic virus serves as a docking point for the intercellular movement of viral replication vesicles. *Plant Physiol*. 2017;175(4):1732-1744. doi:10.1104/pp.17.01484

55. Saha S, Mäkinen K. Insights into the functions of eIF4E-binding motif of VPG in potato virus A infection. *Viruses*. 2020;12(2). doi:10.3390/v12020197

56. Lé S, Viel C, Beauchemin C, Daigneault N, Fortin MG, Laliberté J-FO. Interaction of VPg-Pro of Turnip mosaic virus with the translation initiation factor 4E and the poly(A)-binding protein in planta. doi:10.1099/vir.0.19706-0

57. Coutinho De Oliveira L, Volpon L, Rahardjo AK, et al. Structural studies of the eIF4E-VPg complex reveal a direct competition for capped RNA: Implications for translation. 2019;116(48):24056-24065. doi:10.1073/pnas.1904752116

58. Mäkinen K, Hafrén A. Intracellular coordination of potyviral RNA functions in infection. *Front Plant Sci*. 2014;5(MAR):110. doi:10.3389/fpls.2014.00110

59. Tavert-Roudet G, Anne A, Barra A, Chovin A, Demaille C, Michon T. The potyvirus particle recruits the plant translation initiation factor eif4e by means of the vpg covalently linked to the viral RNA. *Mol Plant-Microbe Interact*. 2017;30(9):754-762. doi:10.1094/MPMI-04-17-0091-R

60. Cheng X, Wang A. The Potyvirus Silencing Suppressor Protein VPg Mediates Degradation of SGS3 via Ubiquitination and Autophagy Pathways. *J Virol*. 2017;91(1). doi:10.1128/jvi.01478-16

61. Zhao J, Zhang X, Hong Y, Liu Y. Chloroplast in plant-virus interaction. *Front*

Microbiol. 2016;7(OCT). doi:10.3389/fmicb.2016.01565

62. Bhattacharyya D, Chakraborty S. Chloroplast: the Trojan horse in plant–virus interaction. *Mol Plant Pathol*. 2018;19(2):504-518. doi:10.1111/mpp.12533

63. Patarroyo C, Laliberté JF, Zheng H. Hijack it, change it: How do plant viruses utilize the host secretory pathway for efficient viral replication and spread? *Front Plant Sci*. 2013;3(JAN). doi:10.3389/fpls.2012.00308

64. Potato Virus Y (PVY). Accessed April 20, 2021. https://cropscience.bayer.co.uk/threats/diseases/potato-diseases/potato-virus-y-pvy/

65. Lacomme C, Jacquot E. General characteristics of Potato virus Y (PVY) and its impact on potato production: An overview. In: *Potato Virus Y: Biodiversity, Pathogenicity, Epidemiology and Management*. Springer International Publishing; 2017:1-19. doi:10.1007/978-3-319-58860-5_1

66. Guerret MGL, Nyalugwe EP, Maina S, Barbetti MJ, Van Leur JAG, Jones RAC. Biological and Molecular Properties of a Turnip mosaic virus (TuMV) Strain that Breaks TuMV Resistances in Brassica napus. Published online 2017. doi:10.1094/PDIS-08-16-1129-RE

67. Roldão A, Mellado MCM, Castilho LR, Carrondo MJT, Alves PM. Virus-like particles in vaccine development. *Expert Rev Vaccines*. Published online 2010. doi:10.1586/erv.10.115

68. Steele JFC, Peyret H, Saunders K, et al. Synthetic plant virology for nanobiotechnology and nanomedicine. *Wiley Interdiscip Rev Nanomedicine Nanobiotechnology*. 2017;9(4):1447. doi:10.1002/wnan.1447

69. Kushnir N, Streatfield SJ, Yusibov V. Virus-like particles as a highly efficient vaccine platform: Diversity of targets and production systems and advances in clinical development. *Vaccine*. 2012;31(1):58-83. doi:10.1016/j.vaccine.2012.10.083

70. Narayanan KB, Han SS. Recombinant helical plant virus-based nanoparticles for

vaccination and immunotherapy. *Virus Genes*. 2018;54(5):623-637. doi:10.1007/s11262-018-1583-y

71. Balke I, Zeltins A. Recent advances in the use of plant virus-like particles as vaccines. *Viruses*. 2020;12(3). doi:10.3390/v12030270

72. Yang C-D, Liao J-T, Lai C-Y, et al. Induction of protective immunity in swine by recombinant bamboo mosaic virus expressing foot-and-mouth disease virus epitopes. Published online 2007. doi:10.1186/1472-6750-7-62

73. Denis J, Majeau N, Acosta-Ramirez E, et al. Immunogenicity of papaya mosaic virus-like particles fused to a hepatitis C virus epitope: Evidence for the critical function of multimerization. *Virology*. 2007;363(1):59-68. doi:10.1016/j.virol.2007.01.011

74. Hefferon KL. Repurposing plant virus nanoparticles. *Vaccines*. 2018;6(1). doi:10.3390/vaccines6010011

75. Lebel M ève, Chartrand K, Leclerc D, Lamarre A. Plant viruses as nanoparticle-based vaccines and adjuvants. *Vaccines*. 2015;3(3):620-637. doi:10.3390/vaccines3030620

76. Balke I, Zeltins A. Use of plant viruses and virus-like particles for the creation of novel vaccines. *Adv Drug Deliv Rev*. 2019;145:119-129. doi:10.1016/j.addr.2018.08.007

77. The Transmission Electron Microscope | CCBER. Accessed March 17, 2021. https://www.ccber.ucsb.edu/ucsb-natural-history-collections-botanical-plant-anatomy/transmission-electron-microscope

78. Getting started in cryo-EM | Coursera. Accessed April 13, 2021. https://www.coursera.org/learn/cryo-em#syllabus

79. Cheng Y. Single particle cryo-EM-how did it get here and where will it go. doi:10.1126/science.aat4346

80. Taylor KA, Glaeser RM. Electron diffraction of frozen, hydrated protein crystals. *Science (80-)*. 1974;186(4168):1036-1037. doi:10.1126/science.186.4168.1036

81. Dubochet J, Chang JJ, Freeman R, Lepault J, McDowall AW. Frozen aqueous suspensions. *Ultramicroscopy*. 1982;10(1-2):55-61. doi:10.1016/0304-3991(82)90187-5

82. Dubochet J, McDowall AW. VITRIFICATION OF PURE WATER FOR ELECTRON MICROSCOPY. *J Microsc*. 1981;124(3):3-4. doi:10.1111/j.1365-2818.1981.tb02483.x

83. Kühlbrandt W. The resolution revolution. *Science (80-)*. 2014;343(6178):1443-1444. doi:10.1126/science.1251652

84. Yip KM, Fischer N, Paknia E, Chari A, Stark H. Atomic-resolution protein structure determination by cryo-EM. *Nature*. 2020;587. doi:10.1038/s41586-020-2833-4

85. Doerr A. Structural analysis of macromolecular assemblies. *Nat Methods*. 2008;5(1):23. doi:10.1038/nmeth1160

86. Penczek PA. Fundamentals of Three-Dimensional reconstruction from projections. In: *Methods in Enzymology*. Vol 482. Academic Press Inc.; 2010:1-33. doi:10.1016/S0076-6879(10)82001-4

87. DeRosier DJ, Moore PB. Reconstruction of three-dimensional images from electron micrographs of structures with helical symmetry. *J Mol Biol*. 1970;52(2):355-369. doi:10.1016/0022-2836(70)90036-7

88. Nogales E, Scheres SHW. Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. *Mol Cell*. 2015;58(4):677-689. doi:10.1016/j.molcel.2015.02.019

89. He S, Scheres SHW. Helical reconstruction in RELION. *J Struct Biol*. Published online 2017. doi:10.1016/j.jsb.2017.02.003

90. Koning RI, Koster AJ, Sharp TH. Advances in cryo-electron tomography for biology and medicine. *Ann Anat*. 2018;217:82-96. doi:10.1016/j.aanat.2018.02.004

91. Kudryashev M. Resolution in Electron Tomography. In: ; 2018:261-282. doi:10.1007/978-3-319-68997-5_10

92. Zhang P. Advances in cryo-electron tomography and subtomogram averaging and classification. *Curr Opin Struct Biol*. 2019;58:249-258. doi:10.1016/j.sbi.2019.05.021

93. Wan W, Briggs JAG. Cryo-Electron Tomography and Subtomogram Averaging. In: *Methods in Enzymology*. ; 2016. doi:10.1016/bs.mie.2016.04.014

94. Clare DK, Pechnikova E V., Skurat E V., et al. Novel inter-subunit contacts in barley stripe mosaic virus revealed by cryo-electron microscopy. *Structure*. 2015;23(10):1815-1826. doi:10.1016/j.str.2015.06.028

95. Wang H, Culver JN, Stubbs G. Structure of ribgrass mosaic virus at 2.9 Å resolution: Evolution and taxonomy of tobamoviruses. *J Mol Biol*. 1997;269(5):769-779. doi:10.1006/jmbi.1997.1048

96. Tewary SK, Oda T, Kendall A, et al. Structure of hibiscus latent Singapore virus by fiber diffraction: A nonconserved His122 contributes to coat protein stability. *J Mol Biol*. 2011;406(3):516-526. doi:10.1016/j.jmb.2010.12.032

97. Wang H, Stubbs G. Structure determination of cucumber green mottle mosaic virus by x-ray fiber diffraction. Significance for the evolution of tobamoviruses. *J Mol Biol*. 1994;239(3):371-384. doi:10.1006/jmbi.1994.1379

98. Zamora M, Méndez-López E, Agirrezabala X, et al. Potyvirus virion structure shows conserved protein fold and RNA binding site in ssRNA viruses. *Sci Adv*. 2017;3(9):1-8. doi:10.1126/sciadv.aao2182

99. Kežar A, Kavčič L, Polák M, et al. Structural basis for the multitasking nature of the potato virus Y coat protein. *Sci Adv*. 2019;5(7):1-14. doi:10.1126/sciadv.aaw3808

100. DiMaio F, Chen CC, Yu X, et al. The molecular basis for flexibility in the flexible filamentous plant viruses. *Nat Struct Mol Biol*. Published online 2015. doi:10.1038/nsmb.3054

101. Grinzato A, Kandiah E, Lico C, Betti C, Baschieri S, Zanotti G. Atomic structure of potato virus X, the prototype of the Alphaflexiviridae family. *Nat Chem Biol*.

2020;16(5):564-569. doi:10.1038/s41589-020-0502-4

102. Yang S, Wang T, Bohon J, et al. Crystal structure of the coat protein of the flexible filamentous papaya mosaic virus. *J Mol Biol*. 2012;422(2):263-273. doi:10.1016/j.jmb.2012.05.032

103. Sainsbury F, Thuenemann EC, Lomonossoff GP. PEAQ: Versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol J*. 2009;7(7):682-693. doi:10.1111/j.1467-7652.2009.00434.x

104. Fromm SA, Bharat TAM, Jakobi AJ, Hagen WJH, Sachse C. Seeing tobacco mosaic virus through direct electron detectors. *J Struct Biol*. 2015;189(2):87-97. doi:10.1016/j.jsb.2014.12.002

105. Zivanov J, Nakane T, Scheres SHW. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ*. 2019;6(1):5-17. doi:10.1107/S205225251801463X

106. Rohou A, Grigorieff N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol*. 2015;192(2):216-221. doi:10.1016/j.jsb.2015.08.008

107. Rosenthal PB, Henderson R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol*. 2003;333(4):721-745. doi:10.1016/j.jmb.2003.07.013

108. Carroni M, D’imprima E, Kimanius D, Serna M. Hands on Methods for High Resolution Cryo-Electron Microscopy Structures of Heterogeneous Macromolecular Complexes. *Front Mol Biosci*. 2019;6:33. doi:10.3389/fmolb.2019.00033

109. Nakane T, Kimanius D, Lindahl E, Scheres SHW. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *Elife*. 2018;7. doi:10.7554/eLife.36861

110. Kucukelbir A, Sigworth FJ, Tagare HD. The Local Resolution of Cryo-EM Density

Maps HHS Public Access. *Nat Methods*. 2014;11(1):63-65. doi:10.1038/nmeth.2727

111. Turoňová B, Schur FKM, Wan W, Briggs JAG. Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4 Å. *J Struct Biol*. 2017;199(3):187-195. doi:10.1016/j.jsb.2017.07.007

112. Li X, Mooney P, Zheng S, et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods*. 2013;10(6):584-590. doi:10.1038/nmeth.2472

113. Mastronarde DN, Held SR. Automated Tilt Series Alignment and Tomographic Reconstruction in IMOD. doi:10.1016/j.jsb.2016.07.011

114. Radermacher M. Weighted back-projection methods. In: *Electron Tomography: Methods for Three-Dimensional Visualization of Structures in the Cell*. Vol 9780387690087. Springer New York; 2006:245-273. doi:10.1007/978-0-387-69008-7_9

115. Bharat TAM, Scheres SHW. Resolving macromolecular structures from electron cryo-tomography data using sub-tomogram averaging in RELION. doi:10.1038/nprot.2016.124

116. Mindell JA, Grigorieff N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol*. Published online 2003. doi:10.1016/S1047-8477(03)00069-8

117. Pintilie GD, Zhang J, Goddard TD, Chiu W, Gossard DC. Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *J Struct Biol*. 2010;170(3):427-438. doi:10.1016/j.jsb.2010.03.007

118. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25(13):1605-1612. doi:10.1002/jcc.20084

119. Emsley P, Lohkamp B, Scott WG, Cowtan K. Biological Crystallography Features and development of Coot. doi:10.1107/S0907444910007493

120. Adams PD, Afonine P V, Bunkó G, et al. Biological Crystallography PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Res Pap Acta Cryst*. 2010;66:213-221. doi:10.1107/S0907444909052925

121. Chen VB, Arendall WB, Headd JJ, et al. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr Sect D Biol Crystallogr*. 2010;66(1):12-21. doi:10.1107/S0907444909042073

122. Walsh I, Minervini G, Corazza A, Esposito G, Tosatto SCE, Fogolari F. Bluues server: Electrostatic properties of wild-type and mutated protein structures. *Bioinformatics*. 2012;28(16):2189-2190. doi:10.1093/bioinformatics/bts343

123. Fogolari F, Corazza A, Yarra V, Jalaru A, Viglino P, Esposito G. *Bluues: A Program for the Analysis of the Electrostatic Properties of Proteins Based on Generalized Born Radii*. doi:10.1186/1471-2105-13-S4-S18

124. Iancu C V., Tivol WF, Schooler JB, et al. Electron cryotomography sample preparation using the Vitrobot. *Nat Protoc*. Published online 2007. doi:10.1038/nprot.2006.432

125. *EPU 2 Software Sample Screening and Data Acquisition Software for Single Particle Analysis Workflow (Cryo-EM SPA)*.; 2020.

126. Kremer JR, Mastronarde DN, McIntosh JR. Computer visualization of three-dimensional image data using IMOD. *J Struct Biol*. Published online 1996. doi:10.1006/jsbi.1996.0013

127. Förster F, Hegerl R. Structure Determination In Situ by Averaging of Tomograms. *Methods Cell Biol*. Published online 2007. doi:10.1016/S0091-679X(06)79029-X

128. Fernandez-Leiro R, Scheres SHW. A pipeline approach to single-particle processing in RELION. In: *Acta Crystallographica Section D: Structural Biology*. ; 2017. doi:10.1107/S2059798316019276

129. Torrance L, Andreev IA, Gabrenaite-Verhovskaya R, Cowan G, Mäkinen K,

Taliansky ME. An unusual structure at one end of potato potyvirus particles. *J Mol Biol*. 2006;357(1):1-8. doi:10.1016/j.jmb.2005.12.021

130. Zivanov J, Nakane T, Forsberg BO, et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife*. Published online 2018. doi:10.7554/eLife.42166

131. Franklin RE, Holmes KC. Tobacco mosaic virus: application of the method of isomorphous replacement to the determination of the helical parameters and radial density distribution. *Acta Crystallogr*. 1958;11(3):213-220. doi:10.1107/s0365110x58000529

132. Holmes KC, Stubbs GJ, Mandelkow E, Gallwitz U. Structure of tobacco mosaic virus at 6.7 Å resolution. *Nature*. 1975;254(5497):192-196. doi:10.1038/254192a0

133. Stubbs G, Warren S, Holmes K. Structure of RNA and RNA binding site in tobacco mosaic virus from 4-Å map calculated from X-ray fibre diagrams. *Nature*. 1977;267(5608):216-221. doi:10.1038/267216a0

134. Bloomer AC, Champness JN, Bricogne G, Staden R, Klug A. Protein disk of tobacco mosaic virus at 2.8 resolution showing the interactions within and between subunits. *Nature*. 1978;276(5686):362-368. doi:10.1038/276362a0

135. Kendall A, McDonald M, Bian W, et al. Structure of Flexible Filamentous Plant Viruses. *J Virol*. Published online 2008. doi:10.1128/jvi.00895-08

136. Donchenko EK, Pechnikova E V., Mishyna MY, et al. Structure and properties of virions and virus-like particles derived from the coat protein of Alternanthera mosaic virus. *PLoS One*. 2017;12(8):1-11. doi:10.1371/journal.pone.0183824

137. Tremblay MH, Majeau N, Gagné MEL, et al. Effect of mutations K97A and E128A on RNA binding and self assembly of papaya mosaic potexvirus coat protein. *FEBS J*. 2006;273(1):14-25. doi:10.1111/j.1742-4658.2005.05033.x

138. Laliberté Gagné ME, Lecours K, Gagné S, Leclerc D. The F13 residue is critical for interaction among the coat protein subunits of papaya mosaic virus. *FEBS J*.

2008;275(7):1474-1484. doi:10.1111/j.1742-4658.2008.06306.x

139. Yang S, Wang T, Bohon J, et al. Crystal structure of the coat protein of the flexible filamentous papaya mosaic virus. *J Mol Biol*. Published online 2012. doi:10.1016/j.jmb.2012.05.032

140. Yuste-Calvo C, Ibort P, Sánchez F, Ponz F. Turnip Mosaic Virus Coat Protein Deletion Mutants Allow Defining Dispensable Protein Domains for 'in Planta' eVLP Formation. *Viruses*. 2020;12(6):661. doi:10.3390/v12060661

141. Zeltins A. Construction and Characterization of Virus-Like Particles: A Review. *Mol Biotechnol*. 2013;53(1):92-107. doi:10.1007/s12033-012-9598-4

142. Erickson JW, Bancroft JB. The self-assembly of papaya mosaic virus. *Virology*. 1978;90(1):36-46. doi:10.1016/0042-6822(78)90330-6

143. Jiang J, Laliberté JF. The genome-linked protein VPg of plant viruses - A protein with many partners. *Curr Opin Virol*. 2011;1(5):347-354. doi:10.1016/j.coviro.2011.09.010

144. Peremyslov V V, Andreev IA, Prokhnevsky AI, Duncan GH, Taliansky ME, Dolja V V. *Complex Molecular Architecture of Beet Yellows Virus Particles*.; 2004. Accessed March 3, 2021. www.pnas.orgcgidoi10.1073pnas.0400303101

PUBLICATIONS

Cuesta, R., Yuste-Calvo, C., Gil-Cartón, D. *et al.* Structure of Turnip mosaic virus and its viral-like particles. *Sci Rep* **9,** 15396 (2019). https://doi.org/10.1038/s41598-019-51823-4