eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

DETERMINACIÓN ESTRUCTURAL DE VIRUS CON ENVUELTA LIPÍDICA MEDIANTE CRIO-MICROSCOPÍA ELECTRÓNICA

ISAAC SANTOS PÉREZ

TESIS DOCTORAL

FEBRERO 2020

THESIS SUPERVISOR:

DR. NICOLA G.A. ABRESCIA

MDe
Master eta Doktorego Eskola
Escuela de Máster y Doctorado
Master and Doctoral School

INDICE GENERAL / GENERAL INDEX

ÍNDICE DE FIGURAS / FIGURES INDEX

ABREVIATURAS / ABREVIATIONS

(+)ssRNA: positive-sense single-stranded RNA (RNA genómico de cadena simple positiva)

aa: Amino acids (Aminoácidos)

Ab: Antibody

ABV: Acidianus bottle shaped virus

ATCC: American Type Culture Collection

BSL: Biosafety level (nivel de bioseguridad)

BTV: "Bluetongue virus" (Virus de la lengua azul)

BVDV: Bovine viral diarrea virus (Virus de la diarrea viral bovina)

CCD: charge-coupled device

Cp: cytopathic (citopático)

Cryo-EM: Cryogenic electron microscopy

Cryo-ET: cryo-electron tomography

CryoTEM: Transmission Electron Cryomicroscopy

CSFV: Classical swine fever virus (Virus de la peste porcina clásica)

CST: Central Section Theorem

CTF: Contrast Transfer Function

DNA: "Deoxyribonucleic acid" (Ácido desoxirribonucleico)

ELISA: Enzyme-Linked ImmunoSorbent Assay

EM: Enfermedad de las mucosas

Fab: antigen-binding fragment

FliP: Flavobacterium infecting, lipid-containing phage

FSC: Fourier Shell Correlation

GPS (protein): Global-Positioning-System proteins

HAdV-2: Human adenovirus C serotype 2

HCIV-1: *Haloarcula Californiae* Icosahedral Virus 1

HCV: Hepatitis C Virus (virus de la hepatitis C)

HHIV-2: *Haloarcula Hispánica* Icosahedral Virus 2

HR: High resolution

IAU: "Icosahedral Asymmetric Unit" (Unidad Asimétrica Icosaédrica)

ICTV: "International Comitee on Taxonomy of Viruses" (Comité Internacional de Taxonomía de los Virus)
IgG: Immunoglobulin G

ITR(s): "Inverted Terminal Repeats" (Secuencias invertidas terminales)

LUCA: "last universal common ancestor" **(**último antepasado común universal)

mAb: Monoclonal Antibody

MCP(s): "Major Capsid Protein(s)" (Proteínas Principales de Cápside)

MGM: Minimum Growth Medium

MOI: Multiplicity of infection

nano-LC–MS/MS: Nanoscale liquid chromatography coupled to tandem mass spectrometry

Ncp: Non cytopathic No citopático

NTE buffer: NaCl-Tris-EDTA buffer

OD: Optical Density

ORF: "Open Reading Frames" (Marcos abiertos de lectura)

PBCV-1: Paramecium bursaria chlorella virus 1

PBS buffer: Phosphate-buffered saline buffer

PCR: Polymerase chain reaction (Reacción en cadena de la polimerasa)

PFV1: Pyrobaculum filamentous virus 1

PI: Persistently infected (Persistentemente infectado)

RE: Retículo endoplasmático

RNA: "Ribonucleic acid" (Ácido ribonucleico)

SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis

SNR: Signal-to-noise ratio

SPA: “Single Particle Analysis” (Análisis de Partículas Individuales)

SSV1: Sulfolobus spindle-shape virus 1

T: “Triangulation number” (Número de triangulación)

TI: Transitoriamente infectado

VNCAGT: virus nucleocitoplasmáticos de ADN de gran tamaño

WBP: weighted back-projection

Capítulo 1:
INTRODUCCIÓN

1.1 Descripción general de los virus

Los virus pueden definirse como agentes biológicos parásitos que necesitan de una célula huésped para poder multiplicarse. Cuando el virus es independiente de su huésped forma una partícula denominada virión, que contiene todos los elementos necesarios para iniciar una nueva infección. Dichos elementos consisten en un genoma viral, RNA o DNA, contenido dentro de una cubierta proteica llamada cápside. En algunos virus existe además una membrana lipídica, denominada envoltura cuando se sitúa en la superficie.

Durante el estado de virión no existe ninguna actividad metabólica y el virus permanece inerte (salvo excepciones [1]), siendo siempre la propagación su única función. La etapa activa del virus comienza cuando el virión reconoce a uno o más receptores de la célula huésped. El papel de los receptores es proporcionar un punto de unión a la membrana celular que sirva como señal para el inicio de la infección [2]. En este punto el virión introduce el código genético en el interior de la célula, existiendo varios mecanismos para lograrlo. Algunos inyectan su carga genética en el citoplasma, otros se fusionan con la membrana celular y un tercer grupo introduce el virión íntegramente en la célula por endocitosis. Una vez dentro de la célula el virus pierde su entidad física y comienza a modificar el ciclo celular del huésped, creando así un entorno favorable para su replicación [3]. De este modo, la maquinaria celular es aprovechada para la síntesis de los componentes virales. Por un lado se crean nuevas copias del ácido nucleico mediante replicación, mientras que por otro se sintetizan las proteínas virales necesarias para la síntesis, maduración y empaquetamiento del virus. Algunos virus sin embargo, después de infectar la célula permanecen en un estado de latencia temporal, llamada fase lisogénica, durante la cual el código genético es incorporado al ADN celular, siendo capaces de generar copias a medida que la célula se duplica, hasta que en un momento determinado la secuencia del virus es liberada de nuevo. En cualquier caso, el virus comienza posteriormente la fase de ensamblaje. En esta, varias copias de una o varias proteínas víricas estructurales sintetizadas se organizan para formar la cápside, que sirve para albergar el genoma viral en su interior. Además, en aquellos virus con envuelta, una o varias glicoproteínas virales pueden encontrarse embebidas en una bicapa lipídica de origen celular. Tras el ensamblaje tiene lugar la salida del virus del interior celular, en la etapa conocida como liberación. La salida del virus puede ser debida a la lisis celular, es decir, por roturas graves de la membrana, causadas por medios físicos o químicos que provocan la muerte de la célula. Sin embargo, muchos virus pueden mantener la

integridad de la célula durante un tiempo prolongado mientras liberan gran cantidad de viriones, provocando así una infección persistente. En estos casos el virus realiza pequeños orificios mediante enzimas líticos, suficientes para su liberación o bien aprovecha los mecanismos de exocitosis propios de la célula, tales como la gemación (budding) o la exocitosis mediada por receptor.

La relación de los virus con los organismos celulares es por tanto muy estrecha, lo que ha influido en la evolución de los organismos. Por un lado, la coevolución de los mecanismos de defensa de la célula y los de infección del virus generan una carrera armamentística [4] que resulta en adaptaciones específicas más complejas en ambos casos. Por otro lado los virus interactúan con el código genético celular, provocando grandes cambios funcionales que por selección natural pueden beneficiar al organismo [5]. Además son el elemento principal en la transferencia horizontal de genes entre organismos, habiendo generado gran parte de la diversidad genética actual [6]. Por tanto su etimología latina con el significado de 'veneno' dista radicalmente de su significado en la actualidad. Los virus son, sin duda, una de las principales herramientas en el presente y el futuro de la investigación, desarrollo e innovación (I+D+I)[7]. Su versatilidad y sencillez hacen que el virus sea un instrumento clave en la nanotecnología médica, empleándose en terapia génica, administración de fármacos, monitorización, inmunoterapia o vacunación, entre otros. Se emplean también en la fabricación de nanomateriales híbridos [8], baterías eléctricas, producción de biodiesel [9] y en infinidad de diseños a pequeña escala que se valen de la arquitectura y propiedades físicas presentes en los virus.

Y es que la gran variabilidad de formas, tamaños y mecanismos hacen de los virus entidades capaces de adaptarse a todo tipo de ambientes, siendo los entes más abundantes de la tierra[10]. Si atendemos a su morfología (Figura 1.1), esta depende principalmente de la cápside y las posibles envueltas lipídicas externas, clasificándose en tan solo 3 grandes conjuntos. La gran mayoría de virus conocidos presenta una estructura ordenada, ya sea en forma de icosaedro o de hélice. Ambas conformaciones suelen estar basadas en el ensamblaje de una o varias proteínas principales de cápside (MCPs) que definen el armazón protector del código genético situado en su interior. El tercer grupo en cambio es mucho más heterogéneo, definiendo a los virus que no presentan una simetría clara o que son claramente pleomorfos. A estos últimos suelen pertenecer la mayoría de virus con una envoltura lipídica externa. [11].

En lo que concierne al tamaño, aún no se saben las dimensiones máximas que un virus puede llegar a alcanzar, ya que en los últimos años esta marca se ha visto superada en varias ocasiones. Actualmente la familia *Pandoraviridae* contiene los virus más grandes en cuanto a genoma (1839 genes y 2,5 Mb), mientras que el género *Pithovirus* alcanza las mayores dimensiones conocidas, superando las 1.5 micras de longitud , aunque a su vez contiene uno de los genomas menos compactos, con "solo" 600 Kb. [12] [13] . Opuestamente, el virus más pequeño constatado pertenece al género *Circovirus,* y unícamente cuenta con 2 genes indispensables (1,76 Kb) y 17 nm de diámetro.[14]

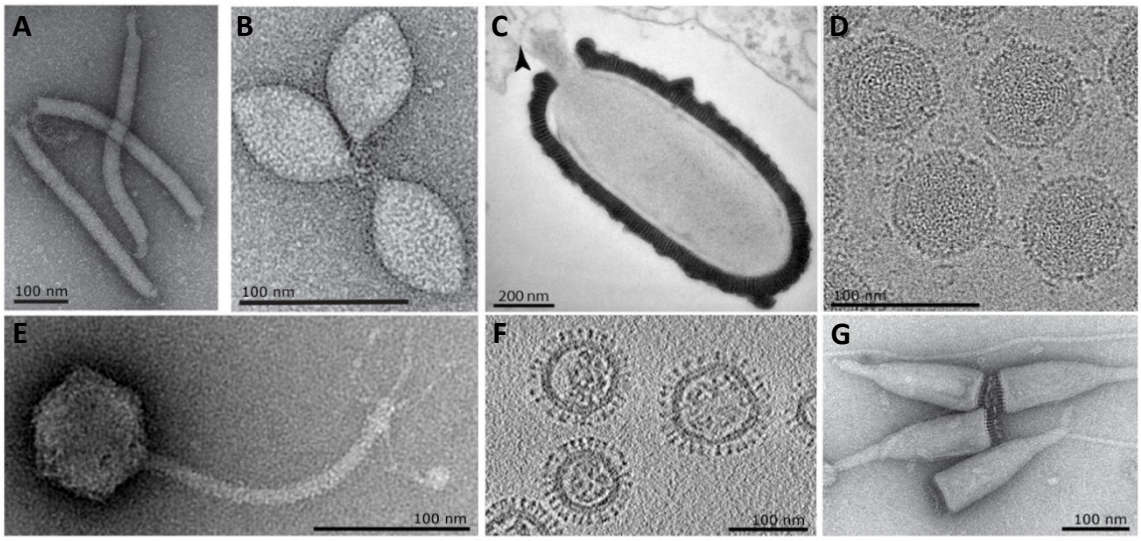


Figura 1.1. Imágenes tomadas mediante microscopía electrónica de partículas virales (viriones) con diferentes tamaños y morfologías. **(A)** Virus PFV1 (Fam. *Tristromaviridae*), con simetría helicoidal y envoltura lipídica (Prangishvili et al. 2017 [200]). **(B)** Virus SSV1 (Fam. *Fuselloviridae*) con morfología de limón y envoltura lipídica (Prangishvili et al. 2017 [200]). **(C)** Virus gigante *Pithovirus sibericum*, de morfología pleomórfica (Legendre et al. 2014 [13]) . **(D)** Virus HCIV-1, con simetría icosaédrica y membrana interna (Santos-Perez et al. 2019 [75]). **(E)** Bacteriófago DT571/2 (Fam. *Siphoviridae*), con cápside icosaédrica y larga cola helicoidal (Golomidova et al. 2016 [201]). (F) Virus de la influenza A (Fam. *Orthomyxoviridae*), con morfología y tamaño altamente variables (Harris et al. 2016 [202]). (G) Virus ABV (Fam. *Ampullaviridae*), en forma de botella (Prangishvili et al. 2017 [200]).

Se calcula que existen al menos 10^{31} particulas virales [15] en nuestro planeta, las cuales infectan a todo tipo de organismos celulares conocidos, existiendo incluso en nichos ecológicos donde las condiciones son extremas. Toda esta diversidad se ha englobado en la llamada virosfera, [16] que abarca a la totalidad de virus existentes y sus interrelaciones.

1.2 Clasificación taxonómica de los virus

Todos los seres vivos conocidos se han tratado de clasificar filogenéticamente, siendo comúnmente aceptado el origen de todas las formas existentes a partir de un "ultimo antepasado común universal" (LUCA) [17]. Sin embargo, los virus no han sido incluidos en esta clasificación al no considerarse estrictamente como organismos vivos [18].

Esto ha hecho necesaria la búsqueda de alternativas que den explicación a la relación de parentesco que también se presume entre los virus. Uno de los métodos de agrupación más utilizado ha sido la clasificación de Baltimore [19], basada en el tipo de genoma y método de replicación del virus. Este sistema es fácil de aplicar pero no es consistente con la relación filogenética, por lo que existe una segunda clasificación que pretende utilizar una nomenclatura linneana, basada en distintos niveles jerárquicos, asemejándose así a la aplicada en los seres vivos. Dicha nomenclatura se encuentra regulada por el Comité Internacional de Taxonomía de Virus (ICTV) [20]. Desafortunadamente, la diversificación tan temprana de los virus y la falta de datos plantea una gran dificultad a la hora de establecer relaciones de secuencia. Por este motivo, en los últimos años se ha propuesto una nueva clasificación basada en linajes estructurales [21], en cada uno de los cuales los virus conservan una homología estructural en sus proteínas principales de cápside (MCPs), incluso cuando la homología de secuencias es baja o inexistente, estableciendo así nuevas relaciones evolutivas entre virus aparentemente inconexos. No obstante, existen indicios de que esta homología estructural en las MCPs esté también respaldada por la similitud de secuencia conservada en los codones de dichos genes, reforzando así el origen común de varias familias de virus tanto filogenética como estructuralmente [22].

Hasta ahora son 4 los linajes establecidos basados en estructura, a saber: "Picorna-like" "PRD1-like" "HK97-like" y "BTV-like" (Figura 1.2). El linaje PRD1-like, por ejemplo, se caracteriza por presentar dos dominios barril-β verticales en su proteína principal de cápside (MCP). La mayor parte de virus con cápside icosaédrica están englobados en estos linajes. A pesar de ello, esta clasificación plantea desafíos tales como englobar a los virus helicoidales y pleomórficos, cuyas proteínas de cápside difieren notablemente en muchos casos a las formadoras de estructuras icosaédricas o están ausentes, respectivamente. Afortunadamente, recientes estudios han sido capaces de establecer homologías estructurales en estos virus [15], que extienden por tanto este sistema

taxonómico a la gran comunidad formante de la virosfera, estimada en más de 100 millones de especies de virus, de las cuales una ínfima parte han sido catalogadas[23]. En esta dirección, nuevos linajes estructurales han sido ya propuestos también para virus sin cápside icosaédrica , vaticinándose aún más en los próximos años[24].

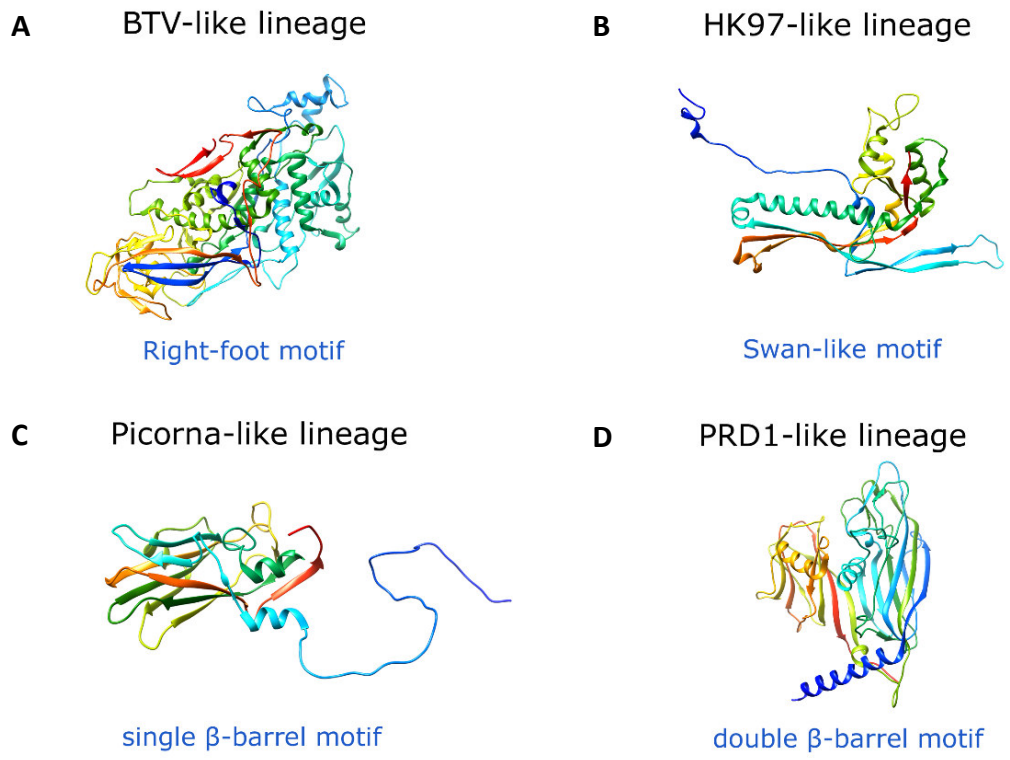


Figura 1.2. Proteínas principales de cápside (MCPs) de los 4 linajes establecidos basados en estructura. **(A)** MCP del virus L-A (Fam. *Totiviridae*), cuyo motivo proteico en forma de pie (right-foot motif) es característico del linaje BTV-like. (Código PDB del modelo atómico: 1M1C (Naitow et al. 2002 [25])). **(B)** MCP del bacteriófago HK97 (Fam. *Shiphoviridae*), cuyo motivo proteico en forma de cisne (Swan-like motif) es propio del linaje HK97-like. (Código PDB del modelo atómico: 1OHG (Helgstrand et al. 2003 [26])). **(C)** MCP del virus de la fiebre aftosa (foot-and-mouth disease virus) (Fam. *Picornaviridae*), cuyo motivo proteico es un barril-β único horizontal (single β-barrel motif), presente en el linaje Picorna-like. (Código PDB del modelo atómico: 1FOD (Logan et al. 1993 [27])) **(D)** MCP del bacteriófago PRD1 (Fam. *Tectiviridae*) cuyo motivo proteico es un doble barril-β vertical (double β-barrel motif), característico del linaje PRD1-like. (Código PDB del modelo atómico: 1W8X (Abrescia et al. 2004 [28])) **.** En todos los casos, los modelos atómicos han sido representados mediante el software gráfico Chimera [29].

En este punto subyace una de las preguntas evolutivas fundamentales. ¿Son los virus anteriores o posteriores a la existencia de la vida celular? Responder a esta pregunta es el perfecto paradigma de lo desconocido. En primer lugar habría que determinar si los virus, al igual que las células, poseen un antepasado común universal, es decir, si su origen es monofilético o es en cambio polifilético. Lo cierto es que no hay evidencias que apoyen

un antepasado único, ya que no existe un solo gen viral que sea reconocido como homólogo en todos los virus estudiados [30]. Por el contrario, los linajes estructurales apuntan a un origen monofilético dentro del mismo pero polifilético entre ellos [31]. Dichos linajes cuentan con virus que infectan a los 3 dominios celulares existentes, lo que apoya la hipótesis de un origen anterior al último antepasado común universal (LUCA) celular. No obstante existen distintas conjeturas, mejor o peor argumentadas, que intentan explicar el posible origen de los virus.

Por un lado se encuentran las hipótesis que respaldan un origen celular de los virus. Algunas de las siguientes son las más conocidas:

Hipótesis regresiva: Los virus se originaron a partir de organismos unicelulares parásitos de otras células mayores, perdiendo durante el curso de la evolución algunos de sus elementos celulares salvo aquellos imprescindibles para su capacidad replicativa [32].

Hipótesis del escape: Los virus derivan de fragmentos celulares de RNA o DNA, tales como plásmidos o transposones. Estos elementos son capaces de replicarse e infectar otras células, pudiendo adquirir formas más complejas incorporando genes de sus hospedadores[33].

Por otro lado hay hipótesis que defienden un origen de los virus previo a la aparición de las primeras formas celulares siendo la siguiente la principal:

Hipótesis de "virus primero": En este caso los virus surgieron a partir de moléculas orgánicas complejas, como las proteínas y ácidos nucleicos, contribuyendo posteriormente a la aparición de los primeros seres celulares [34] .

Finalmente hay hipótesis que apuestan por un origen y evolución paralelos de células y virus. Entre ellas destacan las siguientes:

Koonin et al. (2017) amplía el concepto de virus no solo a aquellos poseedores de cápside sino a cualquier elemento que actúe como parasito de un código genético[35]. De este modo habrían ocurrido numerosas transiciones recíprocas entre virus y elementos parásitos sin cápside a lo largo de la evolución.

Otros investigadores toman como elemento inicial a la membrana lipídica, la cual gracias a la formación de vesículas actuaría como generadora de las protocélulas y de algunos de

los linajes de virus más antiguos, facilitando la coevolución entre ambos gracias a la existencia de transferencia horizontal [36] [10].

También se ha relacionado a los virus como principales impulsores en la incorporación de algunos de los orgánulos principales en las células eucariotas, tales como el nucleo, la mitocondria y el cloroplasto [33].

A pesar de que todas las hipótesis aportan datos empíricos sobre el posible origen de los virus, esta es una cuestión aún por resolver. Por tanto, resulta necesario el estudio de nuevos modelos estructurales, así como comprender la relación específica de cada virus y hospedador, para poder entender mejor los procesos evolutivos que han dado lugar a la gran diversidad de virus actual.

1.3 Virus con membrana lipídica

Importancia de la membrana en los virus

La presencia de una bicapa lipídica no es una característica universal entre los virus. Sin embargo, como parásitos intracelulares obligados que son, su interacción con las bicapas lipídicas del huésped ocurre en varias fases del ciclo viral. Además, los virus poseedores de membrana, ya sea interna o externa, toman sus componentes lipídicos a partir del huésped, por lo que su composición es seleccionada a partir de lípidos presentes en la célula infectada [37]. Como ya se ha citado anteriormente, las vesículas extracelulares involucradas en el tráfico célula-célula son en ocasiones física y químicamente equivalentes a las envueltas víricas, utilizando mecanismos comunes en el tráfico vesicular tanto en la entrada como en la salida a través de la membrana plasmática celular [38]. Existen de hecho hipótesis que sitúan a la membrana lipídica como un elemento común entre virus y células antes incluso de que estos se diferenciaran en entidades distintas tal y como hoy se definen [39]. Debido a que los lípidos de membrana derivan del hospedador el origen de esta estructura es polifilético, dependiendo siempre de la composición lipídica de la célula, aunque la proporción de lípidos es en muchos casos controlada por el virus [40].

Las estructuras y funciones específicas de las membranas lipídicas virales difieren enormemente entre virus, reflejando la gran diversidad de estrategias desarrolladas por estos para adaptarse al hospedador de un manera los más eficiente posible [41].

Entre los virus poseedores de membrana se pueden establecer 2 grandes grupos y un tercero mucho menos numeroso (Figura 1.3). El primero englobaría a aquellos virus cuya membrana, también definida como envuelta, es la capa más externa del virión. En el segundo grupo estarían incluidos los virus cuya membrana se emplaza internamente respecto a la cápside, siendo esta última generalmente de naturaleza icosaédrica. En un tercer grupo más reducido estarían incluidos los virus que presentan ambas membranas, tanto interna como externa. [42, 43] .

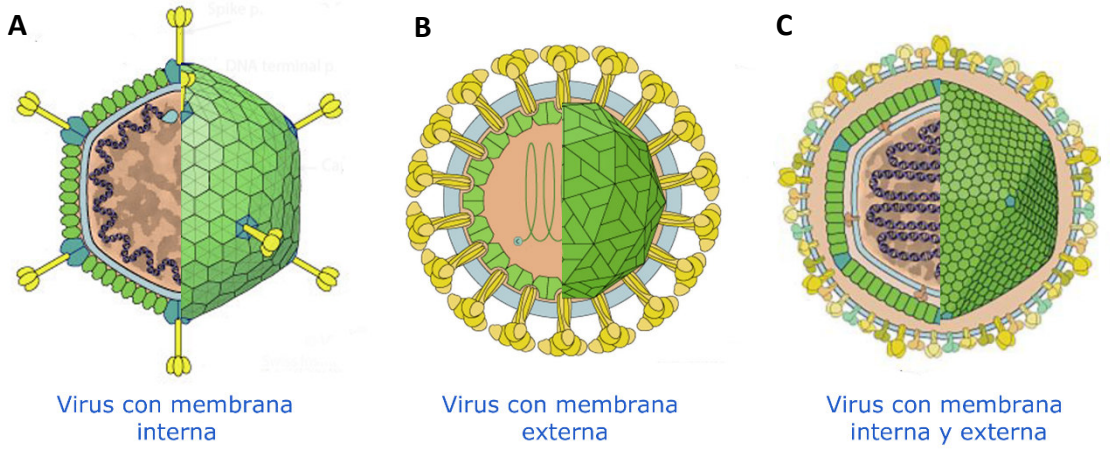


Figura 1.3. Representación de las 3 posibles localizaciones de la membrana en el contexto estructural del virión. Coloración de la membrana (azul claro) y de la cápside (verde). **(A)** Localización de la membrana internamente respecto a la cápside (Ej. Virus fam. *Tectiviridae*). **(B)** Localización externa de la membrana respecto a la cápside (Ej. Virus fam. *Togaviridae*). **(C)** Localización de la membrana a ambos lados de la cápside, interna y externamente. (Ej. Virus fam. *Asfarviridae*). Esquemas de los virus modificados a partir de imágenes de ViralZone (https://viralzone.expasy.org/) (Hulo et al. 2011 [44]).

1) virus con membrana interna

La primera identificación de virus con membrana interna fue en 1968, cuando el bacteriófago denominado PM2 fue analizado, descubriéndose la presencia de una bicapa lipídica en el virión, adquirida a partir de su hospedador [45, 46]. Desde entonces han sido numerosos los virus descritos con membrana interna, destacando el bacteriófago PRD1, el cual ha sido estudiado ampliamente [47]. En todos los casos la membrana está compuesta por lípidos de origen celular y proteínas de origen viral. En estos virus la estructura y el espacio ocupado por la membrana están altamente definidos (con simetría icosahédrica (Figura 1.3. A). Por tanto los lípidos y proteínas, se distribuyen de manera específica en base a sus interacciones y morfología para definir la curvatura necesaria de la membrana [48].

La membrana interna desempeña varias funciones importantes para el virus, siendo algunas de ellas probablemente originadas por un fenómeno de exaptación. Entre otras, actúa como andamiaje para el ensamblaje de la cápside externa. En algunos virus también se ha estudiado su papel en la fusión con la membrana externa del hospedador durante la etapa de entrada [49]. En ocasiones participa activamente formando una estructura tubular que sirve como conducto eyector del genoma viral [50]. Asimismo se le atribuye un papel protector frente al estrés mecánico [51].

2) virus con membrana externa.

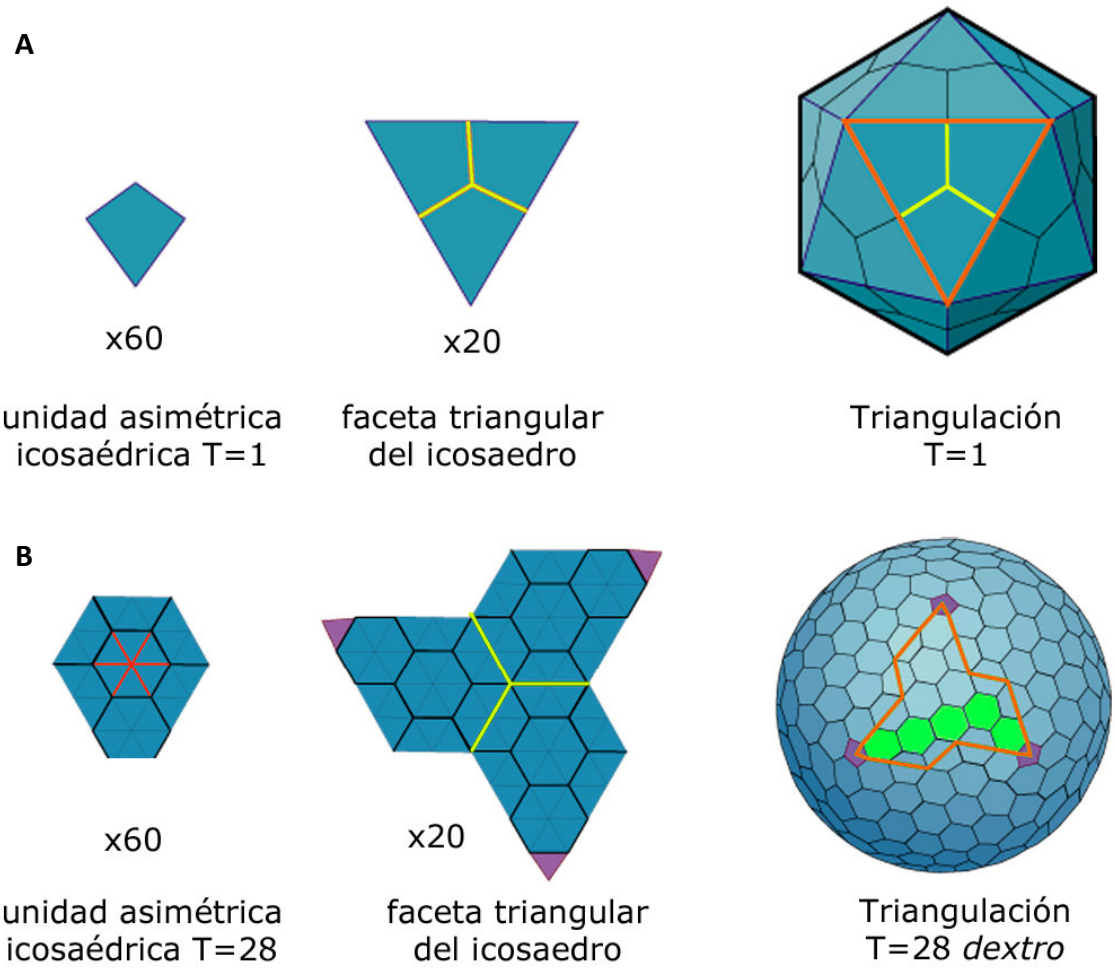
El primer virus aislado con envuelta externa fue el virus del mixoma, perteneciente a la familia *Poxviridae*, en 1898 [52]. Desde entonces numerosos virus con envuelta han sido descubiertos y caracterizados, incluyéndose entre estos a virus patógenos para el ser humano. La membrana externa de estos virus está compuesta por una o varias bicapas lipídicas tomadas del hospedador asociadas con proteínas de membrana, integrales o periféricas, de origen viral [53]. La morfología de la envuelta es muy variable y depende en gran medida del grado de ordenamiento de las proteínas presentes en la membrana, existiendo formas icosaédricas (Figura 1.3, B), elipsoidales o pleomórficas, entre otras.

Gracias a su disposición externa, la envuelta participa en los procesos de entrada al hospedador, reconociendo a los receptores celulares específicos. Algunos virus fusionan su bicapa lipídica con la membrana celular para llegar al citoplasma. En otras ocasiones las partículas virales son introducidas por endocitosis mediada por clatrina [54]. Las proteínas incluidas en la membrana a menudo están glicosiladas, denominándose glicoproteínas. Estas participan en las etapas iniciales de la infección, siendo responsables del reconocimiento del receptor del hospedador y permitiendo la ulterior fusión de membranas o endocitosis del virus.

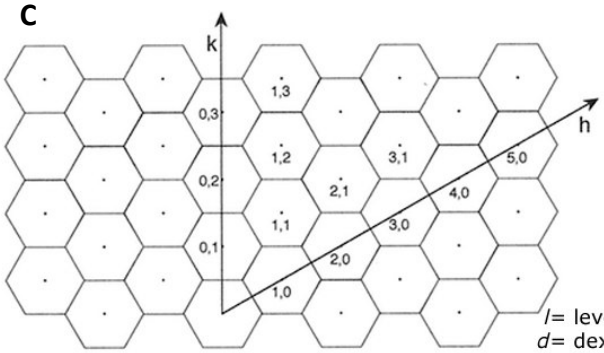
1.4- Virus simétricos y pleomórficos

Las diferencias entre virus simétricos y pleomórficos no residen únicamente en su morfología, pues su organización estructural y funcionalidad también se ven condicionadas en cada caso.

En primer lugar los virus simétricos mantienen una interrelación entre proteínas que les permite mantener una forma definida. Para ello es necesario que las uniones sean estables pero a la vez resistentes a los posibles cambios del medio ambiente extracelular [55] . Las principales proteínas del virus mantienen una relación proporcional y un número definido de unidades. Por ejemplo, en virus icosaédricos la cantidad de subunidades proteicas que forman la cápside es muy variable (desde 60 a 72.000 copias [56]), pero en cada especie el número es siempre el mismo. Esto es debido a que en los virus icosaédricos, los pentágonos formados en torno a los vértices y los hexágonos que completan la cápside, están formados por el mismo bloque proteico, representado por unidades triangulares equiláteras (Figura 1.4). El número de "triángulos" presente en cada una de las 60 unidades asimétricas del icosaedro define el número de Triangulación (T), dando idea de la organización y cantidad de proteínas necesarias para construir la cápside. En los virus pleomórficos, en cambio, el número de copias proteicas es variable, ya que no guardan una posición definida en el contexto general de la estructura del virus. Esta asunción, sin embargo, contiene salvedades en ambas morfologías. En el ensamblaje de estructuras simétricas se cumple el principio de cuasi-equivalencia, según el cual no todas las proteínas formadoras de la envuelta o cápside pueden ocupar la misma posición espacial, produciéndose pequeños cambios conformacionales que estabilizan la estructura del virus (Figura 1. 4. A, B) [57]. Tampoco los virus pleomórficos son totalmente asimétricos, ya que algunos contienen simetrías u asociaciones locales de proteínas [58].



C



h k	T	Ejemplos
1 0	1	bacteriófago ΦX174
1 1	3	virus TBSV del tomate
2 0	4	virus Sindbis
1 2	7*d*	polyomavirus
3 1	13*l*	reovirus
1 3	13*d*	virus IBDV
4 0	16	herpesvirus
5 0	25	adenovirus
4 2	28*d*	Alphasphaerolipovirus

Figura 1.4. Cálculo del número de triangulación (T) en las cápsides de virus icosaédricos. **(A)** Ejemplo de cápside icosaédrica con T = 1. La unidad icosaédrica asimétrica (IAU) representa el espacio mínimo ocupado por el conjunto de proteínas con posiciones cuasi-equivalentes (Una sola proteína en este caso). La cápside completa está formada a partir de 60 IAU idénticas, las cuales pueden ser posicionadas a partir de una sola IAU aplicando operadores de simetría icosaédricos. Cada faceta o cara del icosaedro está formada por 3 IAU, las cuales no necesariamente convergen en el espacio exacto delimitado por las aristas de la faceta. Las 3 IAU conforman entre sí un eje ternario de simetría en el centro de la faceta (líneas amarillas), mientras que a su vez cada IAU establece un eje de simetría binario con las IAU de las facetas adyacentes (líneas naranjas). A su vez, 5 IAU de 5 facetas distintas forman un eje quinario de simetría en torno al vértice pentagonal del icosaedro (pentágonos del vértice en morado). **(B)** Ejemplo de cápside icosaédrica con T = 28. En este caso la IAU está formada por 28 posiciones

cuasi-equivalentes, representadas por triángulos equiláteros (líneas rojas) que forman polígonos hexagonales además de pentágonos en los vértices del icosaedro (en color azul y morado respectivamente. **(C)** Sistema de coordenadas de una red plana hexagonal. El punto de origen de las coordenadas h y k representa el vértice pentagonal del icosaedro. Dichas coordenadas forman un ángulo de 60º que definen el número de hexágonos que deben atravesarse hasta llegar a un nuevo vértice pentagonal, definiendo el número de triangulación del virus a partir de la fórmula $T = h^2 + h \cdot k + k^2$. Por último, h y k también definen la orientación de la cápside icosaédrica. h es el número de unidades en línea recta hacia el próximo pentágono, k es el número de unidades desplazadas a cada lado (derecha (*dextro*) o izquierda (*levo*) para alcanzar el próximo pentágono. A la derecha se muestran los números de triangulación de algunas especies de virus. En el ejemplo (B) (al igual que en los virus icosaédricos de esta tesis), la orientación es dextro (coloreado en verde). Las imágenes han sido modificadas a partir del sitio web ViralZone (https://viralzone.expasy.org/) (Hulo et al. 2011 [44]). El esquema de coordenadas pertenece a Baker et al. 1999 [59].

En segundo lugar, los virus simétricos contienen también asimetrías necesarias para la funcionalidad del virus, como la existencia de un vértice único diferenciado implicado en la infección viral [60]. Además, la simplicidad de estructura que supone el uso de subunidades proteicas homólogas para el ensamblaje del virus a veces es insuficiente para definir correctamente la posición de cada una de ellas, por lo que resulta necesaria la asistencia de proteínas accesorias que faciliten el correcto posicionamiento de dichas subunidades (Figuras 4.10-12) [57]. Por el contrario, los virus pleomórficos tienen como desventaja una envuelta externa lípido-proteica que es mucho más inestable en el tiempo en el ambiente extracelular, pero que a su vez facilita la infección del hospedador mediante fusión [61].

En tercer lugar, algunos virus con simetría icosaédrica aprovechan la gran resistencia al estrés mecánico que confiere la cápside proteica para contener su genoma empaquetado a altas presiones, permitiendo un aprovechamiento máximo del espacio interno [55]. Sin embargo, el volumen interno capaz de albergar el genoma viral es invariable, limitando en gran medida la diversidad genética y adaptabilidad del virus. Los virus pleomórficos en cambio, tienen un rango de morfologías y tamaños variables [42]. Su genoma no está tan altamente empaquetado, lo que podría favorecer una mayor diversidad en la longitud del genoma viral, aumentando de este modo la variabilidad genética.

Por último, refiriéndonos al uso de la microscopía electrónica de transmisión, las técnicas estructurales utilizadas para estudiar las partículas virales son habitualmente distintas en cada caso. Los virus simétricos suelen abordarse mediante un análisis de partículas

individuales (SPA), mientras que para los virus pleomórficos, debido a su variabilidad conformacional, se utiliza la tomografía crioelectrónica.

1.5 Organización estructural de virus icosahédricos con dsDNA

Tal y como se ha mencionado anteriormente, los virus icosaédricos pueden ser clasificados a través de la homología estructural de la(s) proteína(s) de cápside en linajes virales. Dentro de estos linajes, la mayor parte de los virus no solo presentan estructuras homólogas de cápside, sino que también la tipología del material genético suele ser común. Atendiendo a esta premisa, son dos los linajes cuyo material genético suele estar codificado en forma de DNA bicatenario (dsDNA), estos son los linajes HK97-like y PRD1-like. A parte de estos, las familias *Papillomaviridae* y *Polyomaviridae* del linaje Picornavirus-like contienen también DNA bicatenario [24].

El linaje HK97-like incluye al 96% de los bacteriófagos conocidos, englobados en el orden *Caudovirales*. Su estructura está formada por una cápside proteica icosaédrica, que contiene al genoma empaquetado a alta presión, conectándose por uno de los vértices a una cola o tubo, también de base proteica [62]. Este grupo carece de membrana, exceptuando a la familia herpesviridae, también incluida en este linaje, que presenta una envuelta externa [63].

El linaje PRD1-like está compuesto por virus que infectan a los tres dominios celulares (Figura 1.5). A diferencia del grupo anterior no forman tubos proteicos, pero también poseen un vértice diferenciado por donde generalmente se empaqueta y eyecta el DNA, que siempre es bicatenario salvo en el caso del bacteriofago FLiP, donde el DNA es monocatenario [64]. La mayoría de familias de este grupo presentan una membrana interna bajo la cápside (exceptuando la familia *Adenoviridae*), algo único de este linaje, que ayuda en el ensamblaje de la cápside y en la protección del DNA interno [51]. Algunas especies como el propio PRD1 o Bam 35, tienen la peculiaridad de formar un tubo proteolipídico a partir de la membrana interna contenida en el virión. Dicha estructura se forma durante la etapa de infección, tras haber reconocido al receptor del hospedador[65]. El tubo participa en la translocación del DNA vírico desde el virus hacia el citoplasma celular, atravesando la pared y membrana celulares y ejerciendo una fuerza activa que ayuda en la eyección del DNA al interior celular [50]. En lo que respecta a la cápside, el tamaño y número de proteínas que la forman es muy variable, más si cabe tras haberse descubierto en los últimos años hasta diez familias de virus gigantes de eucariotas, para los que se ha

propuesto la creación de un cuarto orden nombrado como *Megavirales* (Familias *Ascoviridae* - *Asfarviridae* - *Iridoviridae* - *Marseilleviridae* - *Megaviridae* - *Mimiviridae* - *Pandoraviridae* - *Phycodnaviridae* - *Pithoviridae* - *Poxviridae*). Son conocidos también como virus nucleocitoplasmáticos de ADN de gran tamaño (VNCAGT). Estos virus se cree que puedan derivar del linaje de PRD1, ya que conservan el mismo plegamiento en su proteína principal de cápside (doble barril-β vertical) (Figura 1.5), pero solo algunos conservan la estructura icosaédrica, como la familia *Asfarviridae* que también presenta una membrana interna y otra externa a la cápside [66].

Por último, los denominados como virófagos, parásitos de los virus gigantes mencionados que requieren de su presencia para su propia replicación, también poseen una cápside icosaédrica y dsDNA como código genético [67]. Según varios estudios evolutivos, el hecho de que virófagos y virus gigantes estén relacionados tanto biológica como estructuralmente podría deberse a un origen común [68]. Tal origen estaría representado por los polintons o polintovirus, los cuales son transposones de DNA integrados en el núcleo de células eucariotas que en ocasiones contienen todos los genes necesarios para formar un virus icosaédrico. Estos elementos tendrían su origen en bacteriófagos del tipo PRD1-like que se integraron en células protoeucariotas, para más tarde desligarse del genoma nuclear en diferentes células eucariotas, formando la variedad de virus actual [69].

1.5.1 *Haloarcula hispanica* icosahedral virus 2 (HHIV-2) y *Haloarcula californiae* icosahedral virus 1 (HCIV-1) de la familia *Sphaerolipoviridae*

Los virus pertenecientes a la familia *Sphaerolipoviridae* presentan una cápside icosaédrica y una membrana interna que protegen a su genoma, formado por dsDNA. Dentro de esta familia existen tres géneros distintos basados en la similitud genética y estructural de sus proteínas principales de cápside y la ATPasa viral, al ser estas proteínas las más conservadas [70]. Los géneros *Alphasphaerolipovirus y Betasphaerolipovirus* coinciden en su nicho ecológico, ya que los virus encontrados pertenecientes a estos grupos infectan a células del dominio Archaea, concretamente al filo Euryarcheota**,** que viven en ambientes halófilos. El género *Gammasphaerolipovirus* en cambio, tiene como hospedador a bacterias termófilas del género *Thermus* [71]. Existen además provirus integrados en el genoma de bacterias y arqueas de ambientes similares, cuyos genes son homólogos a los tres géneros citados, por lo que también han sido clasificados dentro de estos [72].

La característica principal que comparten todos ellos es la utilización de dos proteínas principales de cápside distintas, las cuales están organizadas a lo largo de toda la superficie externa del virión formando una cápside proteica. Ambas proteínas comparten el mismo dominio estructural , que se basa en un plegamiento en forma de barril-β vertical, conservando por tanto una semejanza estructural que rememora al linaje PRD1-like, solo que en este caso la proteína principal no contiene dos dominios barril-β verticales, sino que cada proteína presenta un único motivo (Figura 1.5). Una de las dos proteínas formadoras de esta cápside presenta un dominio barril-β adicional en forma de torre, siendo por tanto la proteína mayor de cápside, en contraposición de la proteína menor de cápside.

En el caso de HHIV-2, estudios estructurales previos afirman la presencia de dos MCPs distintas, presentando cada una un motivo barril-β vertical [73]. Ambas proteínas forman en proporción numérica distinta, la cápside del virus. Respecto a HCIV-1, a pesar de la ausencia de información acerca de su organización estructural, los análisis filogenéticos respaldan su pertenencia al género *Alphasphaerolipovirus*, en el cual se incluye también a HHIV-2. El genoma de HCIV-1 presenta una organización muy similar a otro virus de este grupo, compartiendo una alta similitud de secuencia especialmente en el caso de los genes codificantes de las MCPs y de la proteína ATPasa [70].

Las interacciones entre las dos MCPs formantes de la cápside han sido descritas en otros virus de la familia, concretamente en el fago P23-77 [74], perteneciente al género *Gammasphaerolipovirus*. Sin embargo, actualmente no existe una hipótesis que justifique el patrón de organización característico de las proteínas formadoras de la cápside, lo cual resulta esencial para comprender el sistema de ensamblaje de estos virus.

En este trabajo [75], por tanto, se ha tratado de profundizar en el conocimiento estructural acerca del ensamblaje de los virus HHIV-2 y HCIV-1.

En relación a la simetría de cápside, curiosamente todas las estructuras virales resueltas hasta ahora en esta familia coinciden en el número de triangulación y orientación, siendo esta T = 28 dextro [70]. Por último, otro posible miembro de este grupo podría ser el virus Salisaeta icosahedral phage 1 (SSIP-1), presente en bacterias halofílicas y también con dos proteínas principales de cápside, aunque en este caso el número de proteínas que forman la cápside es mucho mayor [76].

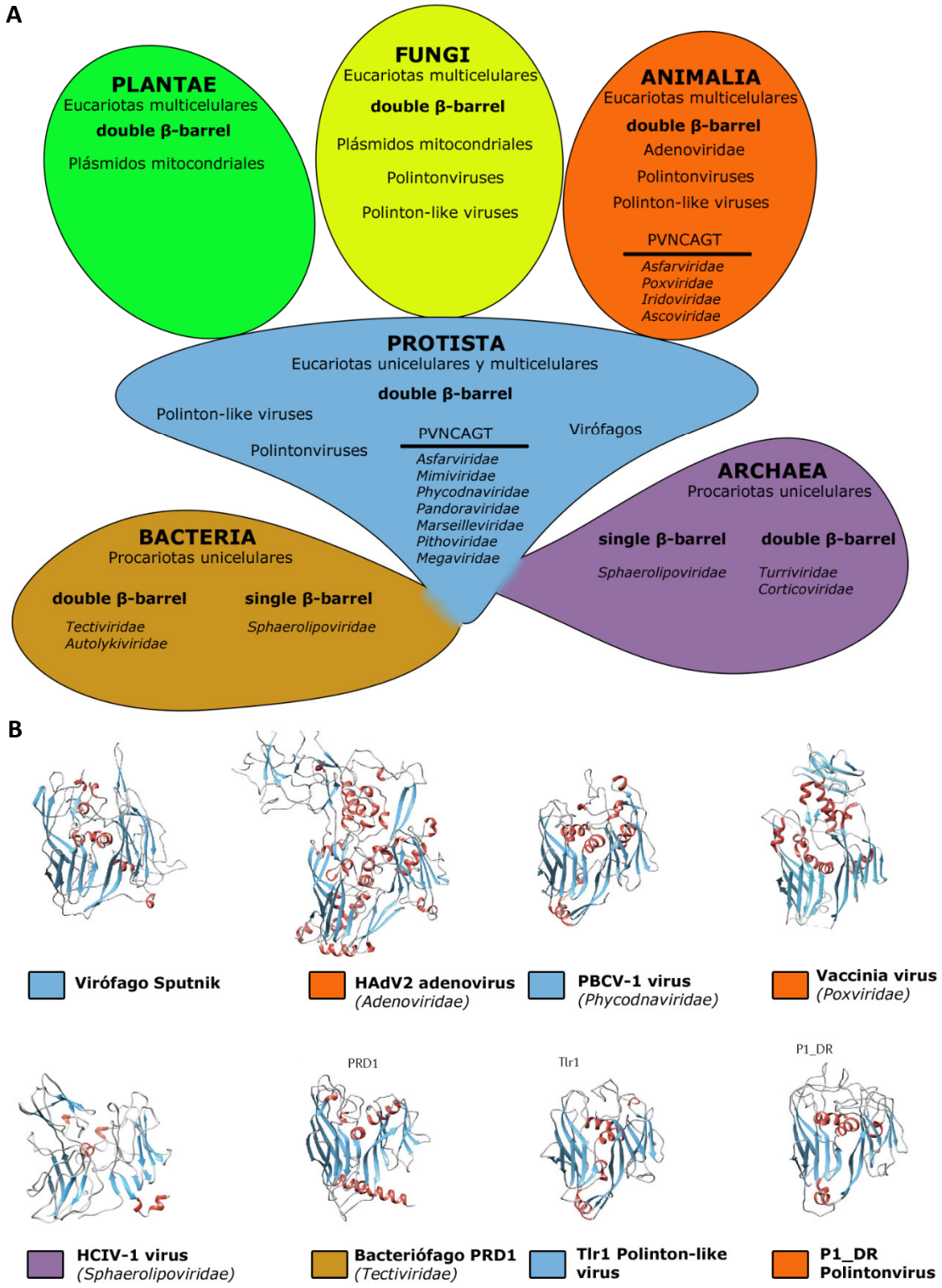


Figura 1.5. Distribución del linaje estructural PRD1-like en base a los organismos infectados. **(A)** Los virus vinculados a este linaje están presentes en seres vivos pertenecientes a todos los dominios y reinos filogenéticos (ilustrados en 7 colores diferentes). El motivo proteico en forma de doble barril-β vertical (double β-barrel motif) ha sido encontrado en MCPs resueltas a nivel atómico en virus del reino animal (Fam. Adenoviridae y virus nucleocitoplasmáticos de ADN de gran tamaño (VNCAGT)). Estos

virus gigantes también están presentes en organismos eucariotas del reino Protista, siendo a su vez parasitados por virófagos con el mismo motivo estructural en sus proteínas de cápside. También hay virus de este tipo en los dominios Archaea (Fam. *Turriviridae*, Fam. *Corticoviridae*) y Bacteria (Fam. *Tectiviridae*, Fam.*Autolykiviridae*). Además, en algunos organismos eucariotas existen plásmidos y elementos genéticos transponibles (polintoviruses y polinton-like viruses) cuya secuencia ha sido identificada con este mismo motivo proteico. Finalmente, los virus con un motivo proteico en forma de un único barril-β vertical y dos proteínas principales de cápside también son incluidos en este linaje, habiéndose encontrado solo en organismos procariotas (Fam. *Sphaerolipoviridae*). **(B)** Modelos estructurales de las proteínas principales de cápside (MCPs) de distintas especies virales englobadas en el linaje PRD1-like. El recuadro coloreado indica al reino o dominio celular al que infectan. La estructura secundaria en forma de α-hélice (color rojo) y láminas beta (color azul claro) evidencia la gran similitud estructural existente en las proteínas de las distintas especies. Códigos PDB: virófago Sputnik:3J26 (Zhang et al. 2012 [77]); adenovirus HAdV2:1P2Z (Rux et al. 2003 [78]); bacteriófago PRD1:1W8X (Abrescia et al. 2004 [79]); PBCV-1 virus:1M4X (Nandhagopal et al. 2002 [75]); Vaccinia virus:3SAQ (Hyun et al. 2011 [80]); HCIV-1 virus (Santos-Perez et al. 2019 [75]); Los modelos estructurales del polinton-like virus Tlr1 y el polintovirus P1_DR no han sido resueltos estructuralmente, tomándose como referencia la predicción de estructura a partir de Krupovic et Koonin 2015 [81])

Respecto a los virus *Haloarcula hispanica* icosahedral virus 2 (HHIV-2) y *Haloarcula californiae* icosahedral virus 1 (HCIV-1), los dos son integrantes del género *Alphasphaerolipovirus,* compartiendo por tanto muchas de sus características biológicas y estructurales. En la mayoría de los casos, los genes estructurales del género conservan su homología entre especies, nombrándose por convención de mayor a menor longitud de secuencia (VP1-VP7, VP9, VP10, VP12 y VP13) (Figura 1.6).

En este género la similitud de secuencia entre las proteínas principales de cápside es altísimo (rango de similitud de aminoácidos del 86 % – 99 % en la proteína menor de cápside VP7, 90 % – 97 % en la proteína mayor de cápside VP4 (Figura 1.6). Asimismo, otras proteínas estructurales, como las dos proteínas principales asociadas a la membrana, VP10 y VP12, conservan una similitud *intragenera* de su secuencia de aminoácidos del 57 % - 83% y 92 % -97 %, respectivamente [71].

En el caso de HHIV-2, su aislamiento se produjo en 2009 a partir de células arqueas de la especie Haloarcula hispánica, provenientes de la salina de Margherita di Savoia, Italia [82]. Su genoma es linear, con una longitud de 30.578 pb y un diámetro de extremo a extremo ≈ 800 Å [73]. Los complejos de unión al receptor celular se encuentran situados en los vértices de los pentones de la cápside y en el caso de HHIV-2 no son homólogos al resto de alphasphaerolipovirus estudiados, ya que los genes que lo constituyen forman parte de un cluster de cinco genes diferente en el resto de virus [83].

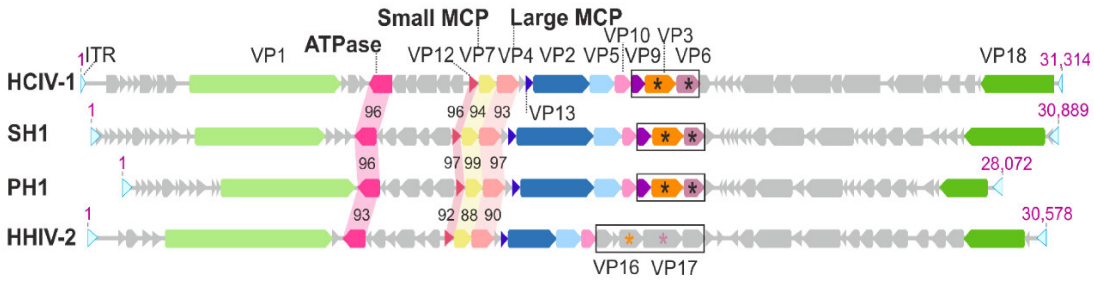


Figura 1.6. Genomas de los virus pertenecientes al género *Alphasphaerolipovirus* (Especies HCIV-1, SH1, PH1 y HHIV-2). (Figura de Demina et al. 2017 [71]). Flechas grises: marcos abiertos de lectura (Open Reading Frames) de cada gen viral. Flechas coloreadas: La misma coloración en distintos genomas indica la homología entre proteínas estructurales. Las flechas en azul claro de los extremos indican las secuencias invertidas terminales (Inverted Terminal Repeats). Las proteínas estructurales, salvo VP1 y VP18, están agrupadas en la misma región del genoma, manteniéndose el mismo orden de lectura en las 4 especies virales, excepto en el caso de las proteínas VP16 y VP17 de HHIV-2 (marcadas con asteriscos), cuya secuencia de aminoácidos no es homóloga a ninguna proteína del resto de especies. El recuadro negro comprende las proteínas estructurales que forman el pentón y el complejo de unión al receptor celular. Los números sobre los genes que codifican para la ATPasa, VP12, VP17 y VP4 reflejan la similitud de la secuencia de aminoácidos de cada proteína respecto a sus homólogas de HCIV-1).

HCIV-1 fue aislado en 2010 a partir de células de la especie *Haloarcula Californiae,* originarias de la salina de Samut Sakhon, Tailandia [84]. Su tamaño es semejante al de HHIV-2, presentando también un genoma linear de 31.314 pb de longitud. Sus genes asociados a la formación del complejo de unión al receptor son homólogos a otros virus del genero *Alphasphaerolipovirus,* como SH1 y PH1 (Figura 1.6) [71]. Respecto al mecanismo de infección de HCIV-1, imágenes tomadas mediante microscopía electrónica de transmisión (TEM), muestran la formación de un tubo a partir del virus que podría asemejarse al sistema utilizado en otros virus del linaje estructural, como es el caso del propio bacteriófago PRD1 [70].

1.5.2 Importancia científica y social de los virus icosahédricos con dsDNA

El estudio de virus icosaédricos con dsDNA tiene una gran relevancia por diversas razones. En primer lugar existen varias especies de virus, entre ellas los adenovirus, que causan distintas enfermedades en el ser humano, para muchas de las cuales todavía no se han desarrollado respuestas terapéuticas satisfactorias [85]. No obstante, los mismos adenovirus así como los herpesvirus, también tienen aplicaciones beneficiosas para

nuestra especie, empleándose como vectores virales en tratamientos mediante terapia génica [86].

Una segunda aplicación de estos virus está relacionada con la emergente resistencia bacteriana a antibióticos. Muchos bacteriófagos, entre ellos PRD1, infectan específicamente a bacterias con plásmidos conjugativos concretos. Dichos plásmidos pueden contener genes de resistencia a antibióticos y por tanto la eliminación por lisis de estas bacterias reduciría su población y la transferencia horizontal de los genes de resistencia a otras bacterias [87].

En tercer lugar, la alta capacidad adaptativa a ambientes extremos de los virus de bacteria y arquea aportan nuevas utilidades aplicables en los sectores tecnológico e industrial. Aquellos adaptados a ambientes hipersalinos, como los virus HHIV-2 y HCIV-1 o el bacteriófago SSIP-1, toleran en algunos casos concentraciones de sal mayores a 3M, siendo la fuente de nuevos enzimas usados en aplicaciones biotecnológicas[41]. Otros como el bacteriófago P23-77 aportan información sobre el ensamblaje y estabilización entre lípidos y proteínas a altas temperaturas [88].

Aparte de las ya mencionadas, la oportunidades de innovación que ofrecen los virus icosaédricos con dsDNA se extienden a nuevos campos de investigación relacionados con la nanotecnología, biomateriales, biomedicina, producción de energía o control de calidad de alimentos [7].

Por consiguiente, el desarrollo de todas estas aplicaciones va ligado a un mejor conocimiento de la estructura viral. Las diferentes técnicas de biología estructural, entre ellas la microscopía electrónica de transmisión, ofrecen mapas a alta resolución que desvelan la disposición exacta de las moléculas virales, infiriéndose información indispensable para su uso en gran variedad de disciplinas, además de su estudio evolutivo.

1.6 Organización estructural en virus pleomórficos de la familia *Flaviviridae*

La familia *Flaviviridae* comprende una gran variedad de virus, algunos de los cuales causan graves enfermedades en el ser humano. A pesar de su gran diversidad, muchas de sus características son compartidas por todos los virus de la familia.

La gran mayoría tienen como hospedador principal a mamíferos y aves, siendo capaces en muchos casos de propagarse mediante vectores artrópodos. Estructuralmente presentan una envuelta lipídica externa que contiene de una a 3 glicoproteínas distintas embebidas, las cuales reconocen a los receptores celulares (Figura 1.7). Bajo esta capa lipídica la proteína estructural 'core' recubre y condensa el RNA genómico de cadena simple positiva ((+)ssRNA) [89]. La longitud del genoma está comprendida entre 9.0 y 13 kb en un solo fragmento con un único marco abierto de lectura (Figura 1.10). La replicación se produce en el citoplasma, tras lo cual la nucleocápside empaqueta al genoma. A continuación la envuelta externa se genera a partir de membranas del retículo endoplasmático, englobando a la nucleocápside. Finalmente, el virión maduro abandona la célula por exocitosis (Figura 1.8). En total son cuatro los géneros asignados a esta familia (*Flavivirus, Hepacivirus*, *Pegivirus* y *Pestivirus*), además de un dudoso quinto género que cumple sólo algunas de las características señaladas, los *Jingmenvirus* [90].

Si atendemos al conocimiento estructural de cada género, mientras que en el género *Flavivirus* se ha resuelto la estructura de hasta seis especies virales a alta resolución (entre ellos el virus del Zika, Dengue y virus del Nilo Occidental [91] [92] [93]), todas ellas de morfología icosaédrica, el resto de grupos carecen de estructuras ordenadas que permitan una fácil reconstrucción mediante análisis de partículas individuales (SPA).

Algunos virus pertenecientes a este segundo grupo aún por caracterizar tienen un gran impacto social, como es el caso del virus de la hepatitis C (HCV), el virus de la diarrea viral bovina (BVDV) o del virus de la peste porcina clásica (CSFV). La ausencia de información de alta resolución en estos virus pleomórficos impide caracterizar varios puntos importantes de su ciclo viral, tales como el modo de ensamblaje, organización de las glicoproteínas de membrana en la envuelta lipídica y su unión a receptor y anticuerpos. Es por tanto necesario el uso de nuevas técnicas estructurales, entre ellas la tomografía crioelectrónica, que ofrezcan un análisis estructural de aquellos virus sin una morfología definida que sin embargo presentan interacciones características entre proteínas a pesar de su variabilidad morfológica.

Similitudes de BVDV y Virus de la Hepatitis C

El virus de la hepatitis C (HCV) es, en ciertos aspectos, semejante al virus BVDV. Ambos pertenecen a la familia flaviviridae, aunque a géneros distintos (Género *Hepacivirus* y *Pestivirus*, respectivamente). No obstante la organización estructural y el mecanismo de infección son muy similares en ambos, poseyendo las mismas proteínas estructurales con la salvedad de la glicoproteína Erns, presente únicamente en *Pestivirus*. Dada la dificultad de crecer y purificar muestras de HCV in vitro, el BVDV se erige como un buen sistema modelo para estudiar su ciclo viral, especialmente en la etapas de entrada y fusión con las membranas celulares[94 95]. Además, las glicoproteínas E1 y E2 de HCV han resultado tener una mayor homología de secuencia con las equivalentes de BVDV que con las glicoproteínas de los *Flavivirus* [96], compartiendo mecanismos de fusión que han sido observados gracias a mapas obtenidos por cristalografía y microscopía electrónica[97]. Al contrario que los *Flavivirus*, las glicoproteínas de HCV y BVDV no pertenecen a ninguno de los tres grupos estructurales de proteínas de fusión de membrana (Clases I, II y III) presentes en el resto de virus con envuelta [98]. La similitud entre los dos virus no es solo estructural, ya que sus proteínas no estructurales cumplen funciones semejantes en la replicación y asociación con las membranas citoplasmáticas de la célula[99]. Algunas de estas funciones se han examinado en virus quiméricos de BVDV en los que se han introducido uno o varios genes de HCV, resultando constructos con una viabilidad similar a las cepas del virus original [100].

Género Flavivirus

Los virus de este género exhiben en la superficie de la envuelta una cápside icosaédrica constituida por la glicoproteína E (Figura 1.7). Esta pertenece a las proteínas de fusión de membrana tipo II y su organización sufre una transición conformacional desde una agrupación dimérica nativa a pH neutro a una forma trimérica fusogénica en condiciones ácidas [101]. En consecuencia, la conformación del virus también cambia a una simetría con entornos cuasi-equivalentes [102, 103]. En el caso de BVDV, en cambio, la mayor acidez del medio no parece inducir ningún cambio conformacional, aunque no existen estructuras a alta resolución que lo verifiquen[104]. Además de la organización estructural de sus glicoproteínas, se ha demostrado también cierto orden en los lípidos de membrana, como en el caso del virus del Dengue. Las diferentes especies lipídicas presentes en la monocapa externa de la envuelta se distribuyen heterogéneamente y en estrecha relación con las glicopoteínas del virus, las cuales están en contacto con hasta el 68% de las

cabezas polares de esta capa [105]. Por tanto la envuelta de estos virus es más estable y la difusión de lípidos es reducida.

La composición lipídica de BVDV también diverge de la distribución de la bicapa celular original, enriqueciendo la cantidad de esfingomielina y colesterol, pero no guarda relación con la composición de familias lipídicas encontradas en *Flavivirus* ni en HCV.

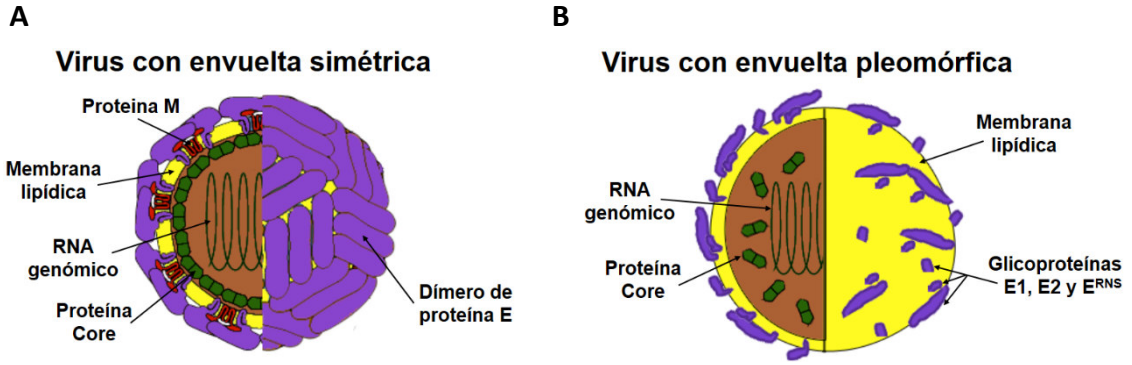


Figura 1.7. (A) Esquema de la estructura de viriones del Gen. *Flavivirus*. (Imagen modificada de Viralzone) (Hulo et al. 2011 [44]). **(B)** Esquema de la estructura de viriones del Gen. Pestivirus. Glicoproteínas (púrpura), proteína lipídica (verde), envoltura lipídica (amarillo).

A pesar de que estas dos especies lipídicas mayoritariamente seleccionadas tienden a formar balsas lipídicas en las bicapas, no parece que haya afinidad de estas por las glicoproteínas de la envuelta [104]. Queda por tanto determinar si los lípidos y glicoproteínas de BVDV guardan alguna relación de coherencia entre sí o por el contrario su distribución es completamente aleatoria.

1.6.1- Virus *Bovine viral diarrhea virus* (BVDV)

1.6.1.1. Descubrimiento de BVDV y su patología

El primer registro referente al virus de la diarrea viral bovina (BVDV) tiene su origen en los años 40 del siglo XX en Canadá, cuando una enfermedad aún por identificar provocaba casos intermitentes de baja morbilidad pero alta letalidad con una sintomatología característica en bóvidos. El ganado afectado presentaba diarrea sanguinolenta acuosa, erosiones de las membranas mucosas, fiebre y otros síntomas que más tarde serían asociados a la presencia del virus. En 1946 T. Childs [106] es el primero en publicar sus observaciones al respecto, denominando el conjunto de afecciones observadas como `enfermedad X´.

Ese mismo año, P. Olafson identifica y publica [107] de forma independiente en Estados Unidos indicios de la misma enfermedad, añadiendo nuevos síntomas como la afección respiratoria, leucopenia, menor producción de leche y mayor tasa de abortos. En este caso, al agente causante de la enfermedad se le identifica como Virus de la Diarrea (VD).

En 1953, F.K. Ramsey[108] acuñó el término `Enfermedad de las mucosas" al observar casos en los que el ganado vacuno presentaba erosiones severas del tracto alimenticio, como ulceraciones y hemorragias internas, además de diarreas abundantes, provocando la emaciación del animal.

Todos estos cuadros clínicos fueron asociados a distintos agentes y enfermedades hasta que en 1987 J.C. Baker [109] determinó el origen común de todos ellos, siendo este un único virus ,de ahora en adelante denominado diarrea viral bovina/enfermedad de las mucosas (DVB/EM), simplificado a menudo como virus de la diarrea viral bovina (BVDV).

Un año después, en 1988[110,111], se publicó la secuencia genómica del virus y sus proteínas víricas asociadas, así como el sistema de replicación que empleaba, lo que permitió clasificar al virus BVDV en la familia *Flaviviridae*[112], de la cual todavía forma parte. Con el inicio de la década de los 90, de nuevo en Canadá y Estados Unidos, surgió un nuevo biotipo del virus mucho más infeccioso y con una alta tasa de mortalidad[113]. Esta variante, a pesar de su alta similitud con el biotipo original presenta variaciones antigénicas de las proteínas de la superficie viral, además de un genotipo distinto, por lo que son consideradas dos especies distintas [114]. En la actualidad se han identificado numerosas cepas del virus con afecciones variables, siendo difícil realizar un diagnóstico preciso incluso una vez se ha identificado en el animal[115].

1.6.1.2. Epidemiología de BVDV

La principal especie infectada por el virus BVDV y por tanto su reservorio más importante es el vacuno doméstico (*Bos Taurus*). Sin embargo el virus también se ha encontrado en otras especies de la familia *Bovidae*, tales como búfalos, cabras y ovejas, e incluso en otras familias del orden *Artiodactyla* como camellos, llamas, girafas, ciervos o cerdos [116]. No se han documentado casos de personas afectadas, aunque existen estudios donde se ha demostrado la infección en humanos y la generación de anticuerpos específicos contra el virus [117]. No obstante la inocuidad del virus está en entredicho y actualmente se están llevando a cabo investigaciones que sugieren una participación

conjunta del virus del zika y BVDV en la aparición de malformaciones congénitas en neonatos [118]. También se han realizado estudios donde se ha confirmado la persistencia de BVDV en cultivos de células humanas [119]. Por todo ello BVDV está enmarcado dentro del nivel de bioseguridad 2 (BSL-2).

Atendiendo a la distribución del virus su presencia es prácticamente mundial, pero con grados de afectación muy variables dependiendo del país[120]. En los países con programas de control y erradicación de la enfermedad el número de ganaderías infectadas se ha reducido, aunque la erradicación del virus aún se antoja lejana[121]. Mientras en la mayoría de países europeos el número de infecciones disminuye, la tendencia en el resto de continentes es hacia un mayor número de infecciones. Pese a las cifras alarmantes de algunos países, con niveles de infección de hasta el 100 % , países con grandes cabañas ganaderas han comenzado a elaborar planes de control [122].

1.6.1.3. Transmisión de BVDV

El estudio de la patogénesis en animales infectados con BVDV es un área difícil de abordar, ya que la transmisión del virus puede originarse de varias maneras, las cuales afectan de diferente modo al hospedador y dependen también de la naturaleza concreta del virus.

En primer lugar cabe aclarar que ya desde los años 60 del siglo pasado el biotipo del virus ha cobrado una alta importancia a la hora de analizar la transmisión del virus[123]. Por un lado existe el virus no citopático, que como su nombre indica no provoca cambios bioquímicos ni morfológicos aparentes en el cultivo, observándolo mediante microscopia óptica. Por otro lado existe el biotipo citopático, que imposibilita la viabilidad del cultivo terminando en muerte celular[124]. Los dos biotipos pueden ser muy similares tanto en su genoma como en los antígenos producidos, la única diferencia determinante está relacionada con la facultad de expresar la proteína no estructural 3 (NS3) en el caso del biotipo citopático[125]. Esta proteína se origina a partir de la proteína NS2-3, presente en los biotipos no citopáticos, que debido a diferentes motivos se divide en las correspondientes NS2 y NS3 [126].

En segundo lugar es importante identificar el modo en el que el virus se transmite de un animal a otro. Existen dos formas de transmisión: la vertical y la horizontal [126].

En la transmisión vertical una hembra gestante puede pasar el virus al feto por transmisión transplacentaria. Las consecuencias son muy distintas dependiendo de la fase del embarazo en la que nos encontremos. Durante el embarazo temprano (días 0 a 40), el feto suele morir provocando un aborto temprano. Si la infección ocurre en el ecuador del embarazo (días 40 a 120) la hembra sí es capaz de generar anticuerpos en contra del virus pero el feto aún no ha desarrollado un sistema inmune capaz de reconocer al virus como un antígeno exógeno y lo acepta como uno propio. Cuando esto ocurre el feto puede ser abortado, pero en muchas ocasiones el parto llega a término y nace un ternero persistentemente infectado (PI), el cual elimina grandes cantidades de virus a través de sus fluidos y secreciones, tales como la saliva, las heces, la leche y el semen, siendo por tanto el principal foco de infección para el resto de animales sanos. Por último si la infección tiene lugar en la etapa final de la gestación (días 120 a 280) el ternero nacerá inmunizado ante el virus, pero con alta probabilidad presentará malformaciones y un crecimiento lento. La transmisión vertical sólo se ha observado en biotipos no citopáticos.

En la transmisión horizontal el causante de la infección puede ser tanto un animal PI como un animal transitoriamente infectado (TI), que durante 4-10 días también es fuente de infección hasta que normalmente su sistema inmune genera anticuerpos en contra del virus[126]. Los síntomas provocados por el virus, no citopático, son los clásicos reportados por la enfermedad de la diarrea viral bovina, destacándose un cuadro diarreico leve y una bajada de las defensas. Sin embargo cuando un animal PI se infecta de nuevo con un segundo biotipo, esta vez citopático, el virus se propaga hasta el epitelio gastrointestinal, provocando la denominada enfermedad de las mucosas (EM), que provoca erosión y ulceración en el tracto digestivo, finalizando con la muerte del animal. Para que esto ocurra los dos biotipos (cp y ncp) deben ser homólogos, es decir, antigénicamente similares. Contrariamente a lo esperado, la doble infección por parte de un biotipo ncp y otro cp homólogos es muy común, ya que los animales PI suelen sufrir mutaciones del virus ncp, generando un nuevo biotipo cp que desencadena la enfermedad en el propio individuo y en todos los demás individuos virémicos.

Por último existe un síndrome hemorrágico severo, asociado únicamente a alguna cepas del biotipo ncp en el genotipo BVDV-2 [127], que pueden transmitirse sea por vía horizontal que vertical, causando hemorragias internas en el tracto digestivo e incluso la muerte[113] [128].

1.6.1.4. Tratamientos actuales

Conseguir la erradicación total de la enfermedad provocada por el virus BVDV es una propuesta demasiado ambiciosa con los medios empleados actualmente, ya que su naturaleza sigilosa, la cantidad de especies animales que pueden actuar como reservorios del virus y su alta diversidad genética hacen imposible un control absoluto sobre la enfermedad. A pesar de esto, la incidencia del virus se ha reducido en los lugares donde se han aplicado tres pilares fundamentales: El uso de vacunas, análisis sistemáticos y programas de bioseguridad.

La vacunación es una herramienta que mitiga los efectos de la enfermedad y previene la infección de individuos sanos, pero es necesario conocer de antemano la cepa concreta del virus para asegurar su efectividad. Existen distintos tipos de vacunas, que pueden ser el uso de virus inactivado o de virus atenuado.

Las vacunas inactivadas contienen cepas inactivas del virus o partes del mismo. Su efecto protege al animal de manera segura, con muy bajo riesgo de infección, pero la respuesta inmune que provoca es baja y por tanto su efecto es poco duradero. El feto no se encuentra debidamente protegido y es susceptible a una infección transplacentaria.

Las vacunas vivas atenuadas contienen virus viables que, sin embargo, se modifican de tal manera que no desencadenan un proceso de replicación virulenta. Estas vacunas confieren una protección duradera, también en el feto en casos de embarazo, sin embargo no es totalmente efectiva y puede ocasionar infecciones transplacentarias similares a las del virus silvestre, naciendo terneros PI.

En este mismo sentido, existen otras vías de investigación que tratan de descubrir nuevas terapias más efectivas contra el virus. Por un lado se están desarrollando antivirales cuyo objetivo es la unión específica a los dominios I y II de la glicoproteína viral E2, bloqueando de este modo la entrada del virus[129]. Estos antivirales serían suministrados junto a inhibidores ya conocidos contra la polimerasa del virus[130] [131] . Otro sistema en estudio consiste en la creación de un cóctel de adenovirus recombinantes que expresan antígenos de BVDV. El adenovirus actúa como vector para la liberación de los genotipos de E2 más comunes, los cuales son reconocidos por los anticuerpos del animal[132]. La ventaja de este sistema es el nulo riesgo de infección por parte del virus BVDV. Por último se está desarrollando una vacuna a base de probióticos por vía oral, utilizando la

bacteria *Lactobacillus casei* como portador del antígeno E2 de BVDV[132]. Como se puede apreciar, el conocimiento de la estructura y uniones específicas de las glicoproteínas de BVDV resultan de vital importancia para asegurar una mejor inmunización en el animal.

La segunda medida a tener en cuenta para reducir el impacto de BVDV es la realización de análisis rutinarios en la cabaña bovina. Lo más común es realizar ELISA comerciales para la detección de anticuerpos del virus en suero o leche [133]. Para detectar los animales PI, los cuales no generan anticuerpos contra BVDV, el diagnóstico más utilizado es el test de PCR en tiempo real, que detecta el RNA del BVDV en muestras de sangre fresca[134]. Ambas pruebas son altamente fiables, pero la aparición constante de nuevas cepas del virus y el elevado costo de un control rutinario exhaustivo impiden la identificación total de individuos infectados [135 136].

La tercera estrategia trata de reducir la extensión del virus mediante medidas de seguridad, minimizando mediante diversos protocolos la exposición de individuos sanos al virus y eliminando los principales focos de infección. Algunas de estas prácticas son el aislamiento de animales transitoriamente infectados, la eliminación de aquellos PI, el tratamiento de residuos contaminados y el control en el transporte de los mismos de un grupo a otro [137].

1.6.1.5. Ciclo de vida

Aunque el virus BVDV es capaz de sobrevivir desde pocas horas hasta varias semanas en un ambiente extracelular[138], su subsistencia requiere de un hospedador donde desarrollar su ciclo viral. Este consta de las siguientes etapas (Figura 1.8):

Fijación: En esta etapa las glicoproteínas estructurales de BVDV Erns, E1 y E2 participan en la unión a receptores celulares, enfrentando las membranas del virus y la célula[139].

Endocitosis: Se produce la internalización del virus mediante endocitosis dependiente de clatrina, generando un endosoma temprano.

Fusión: Gracias a un descenso del pH del endosoma, los motivos de fusión de las proteínas estructurales de las dos membranas se fusionan y comiénza la síntesis viral: El RNA monocatenario positivo ((+)ssRNA) del virus se libera en el citoplasma y es traducido en una poliproteína, procesada por proteasas celulares y víricas.

Replicación y traducción: Una vez puesta en marcha la maquinaria vírica, en el retículo endoplasmático (RE), el RNA se replica mediante transcripción de moléculas de RNA de cadena negativa complementarias, que actúan como plantillas para la síntesis de nuevas moléculas de RNA genómico[139].

Ensamblaje: También en el RE, las glicoproteínas virales acumuladas se organizan y empaquetan el RNA formando viriones.

Liberación: El virión ya formado se secreta a través de una vía mediada por el aparato de Golgi que termina con la exocitosis[140].

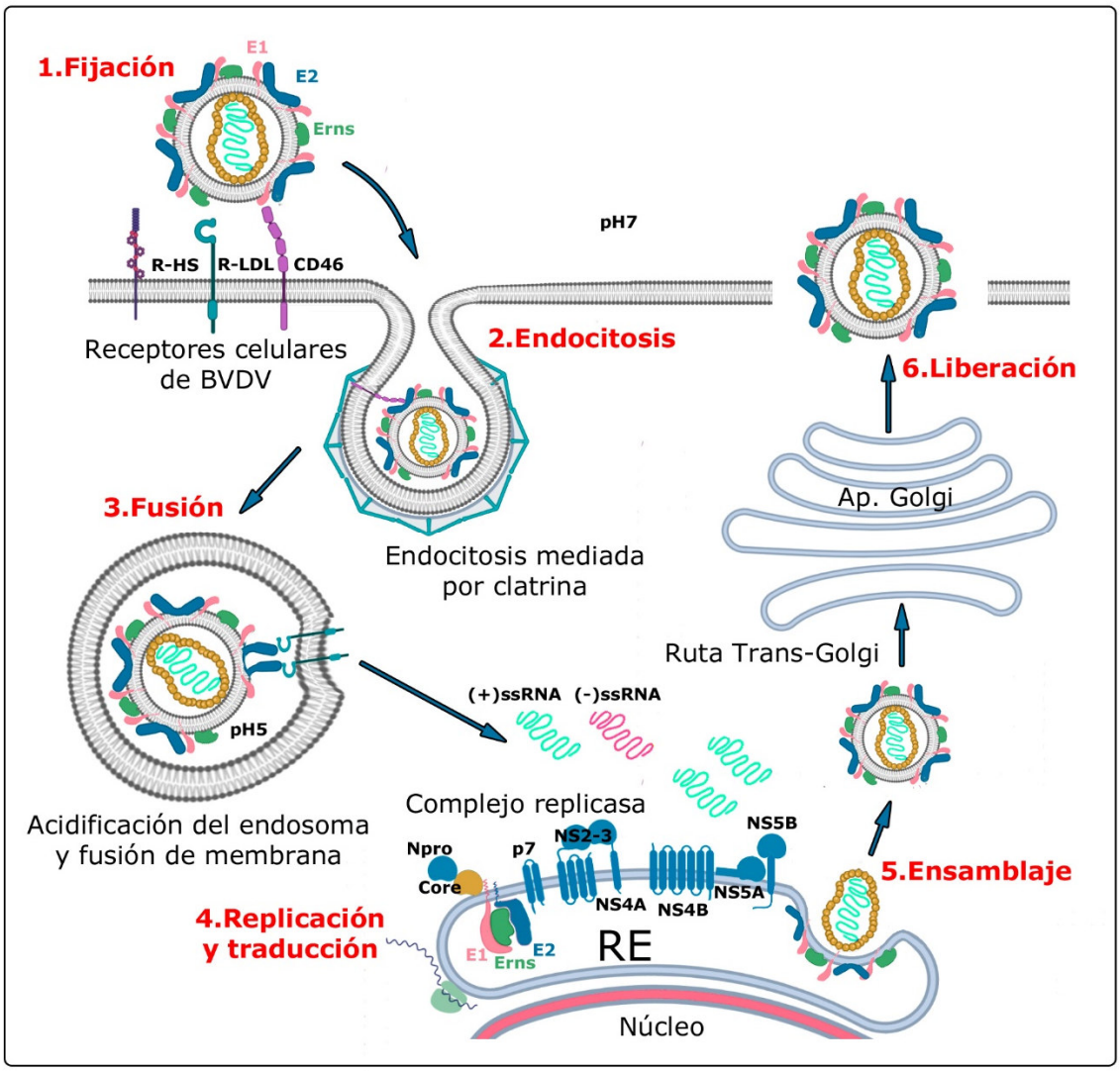


Figura 1.8. Representación esquemática del ciclo de vida viral de BVDV. En ella se muestran las diferentes etapas del virus. El ciclo del virus comienza con la fijación del virión mediante la unión de sus glicoproteínas a los receptores primarios. A continuación el virus es internalizado mediante endocitosis mediada por clatrina y unión a recptores secundarios. Ya en el endosoma, el descenso de pH provoca cambios en las glicoproteínas del virus que facilitan la fusión de la membrana vírica y la del endosoma, liberando el

RNA del virión. La transcripción y traducción del genoma viral ocurren en el citoplasma celular, en asociación con el retículo endoplasmático (RE). A partir de la membrana lipídica del RE asociada a las glicoproteínas virales acumuladas se produce el ensamblaje del virión. Finalmente, el virus es liberado por exocitosis a través de la vía trans-Golgi.

1.6.1.6. Estructura

Actualmente no existe ninguna estructura a alta resolución que defina la morfología exacta del virión de BVDV. A pesar de esto sí que es posible describir a grandes rasgos su tamaño ,organización y conformación[104]. En primer lugar BVDV es un virus pleomórfico y por lo tanto no presenta ninguna simetría aparente (Figura 1.7). El diámetro del virus es variable, oscilando entre 40 y 60 nm. Externamente presenta una envuelta lipídica derivada de membranas intracelulares del hospedador. Las 3 glicoproteínas del virus, (Erns, E1 y E2) están insertadas externamente en esta envuelta, existiendo asociaciones hetero y homodiméricas entre ellas. Una cuarta proteína estructural denominada core se encuentra desordenada bajo la membrana, uniéndose al ARN interno con una baja afinidad y sin una especifidad clara[141].

1.6.1.7. Proteínas estructurales

Como ya se ha mencionado, el virus BVDV contiene 4 proteínas estructurales, las cuales se detallan a continuación:

Proteína core: Es la más pequeña con 14 KDa y 102 aa. No tiene ningún plegamiento específico, encontrándose desplegada en forma nativa. El alto contenido en residuos básicos actúa de agente condensante del RNA vírico por medio de interacciones inespecíficas [141].

Proteina Erns: Tiene una masa de 48 KDa y 227 aa. Su dominio catalítico ha sido cristalizado (Figura 1.9, A) [142]. Es la única proteína exclusiva en pestivirus. Posee un dominio catalítico con actividad ribonucleasa [142]. Su dominio C-terminal forma una hélice anfipática que actúa a modo de anclaje a la membrana[143]. Además también participa en la unión del virus al receptor[143] . También se une entre sí por puentes disulfuro formando homodímeros.

Proteína E1: Con 25 KDa y 162 aa, esta es la proteína con menos información estructural disponible: No existen mapas sobre sus estructura y los animales infectados no generan anticuerpos contra esta proteína [144]. En el extremo C-terminal contiene dos hélices

anfipáticas y una hélice transmembrana que hacen de anclaje en la envuelta lipídica del virusn[95]. En su extremo N-terminal se ancla al receptor celular. También crea heterodímeros con E2 unidos por puentes disulfuro que son esenciales para la fusión.

Proteína E2: Tiene 55 KDa de tamaño y 407 aa. Al igual que E1 se ancla a la membrana del virus por el C-terminal mediante una secuencia hidrofóbica. A partir de su estructura cristalográfica se ha observado que su plegamiento es distinto a las proteínas de fusión de clase II de otros virus de la familia *Flaviviridae* (Figura 1.9, B). El ectodominio de E2 muestra cuatro dominios claramente definidos (DA, DB, DC y DD), siendo los dos más distales a la membrana los encargados de la unión al receptor celular[145].

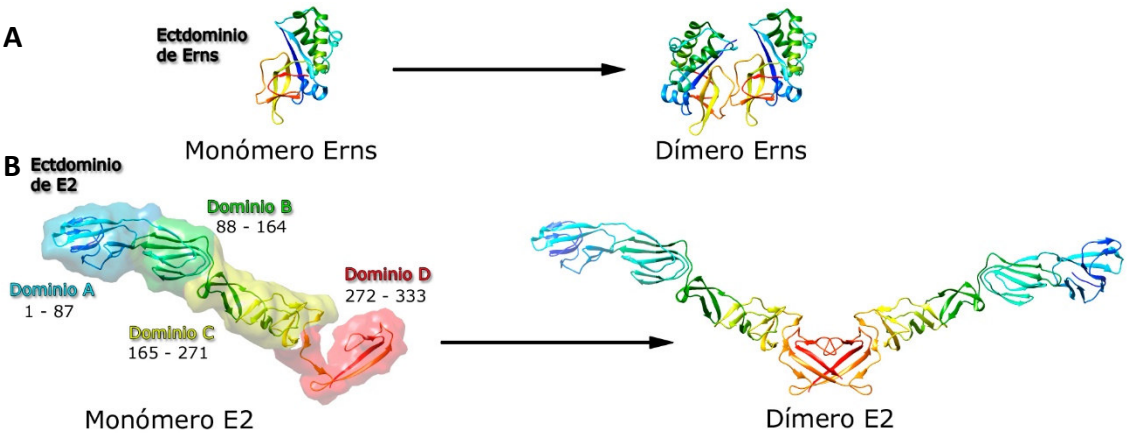


Figura 1.9. Representación gráfica de los modelos atómicos de los ectodominios de las glicoproteínas Erns y E2 de BVDV, obtenidos por cristalografía de rayos X. **(A)** Ectodominio de Erns (residuos 3-161), cristalizado en forma de homodímero. Código PDB del modelo atómico: 4DW5 (Krey et al. 2012 [142]). Erns también se encuentra en forma de monómero en la envuelta lipídica de BVDV. El extremo C-terminal forma una hélice anfipática que no ha sido cristalizada. **(B)** Ectodominio de E2 (residuos 2-333), cristalizado en forma de homodímero. Código PDB del modelo atómico: 2YQ2 (El Omari et al. 2013 [145]). El extremo C-terminal forma una región transmembrana que no ha sido cristalizada. El ectodominio E2 se encuentra a su vez organizado en 4 dominios (A, B, C y D) (El mapa de densidad y modelo atómico de cada dominio se han coloreado del mismo modo en cada caso). Los dominios A y B son los más distales a la membrana y probablemente los más expuestos en la superficie del virus.

1.6.1.8. Genotipos

En los últimos años, el número de especies del Genero *Pestivirus* reconocidas por la ICTV ha ido en aumento, catalogándose en la actualidad por letras en orden alfabético (Pestivirus A-K), aunque existen muchos aún por clasificar. El criterio de esta

clasificación se basa principalmente en la similitud de secuencia, pero la relación antigénica y el hospedador de origen también son tenidos en cuenta.

En base a lo citado podríamos hablar de BVDV en términos de tres especies distintas de pestivirus. La principal especie, por su impacto en el ganado estabulado y por ser la que mayor número de cepas distintas presenta es BVDV-1 (Pestivirus A). La cepa CP7, perteneciente a su vez al subtipo b (U63479 – 1985 (USA) Corapi et al., 1988) ha sido la utilizada en nuestro estudio estructural, por ser una de las más presentes y mejor investigadas. La segunda especie por número de animales infectados es BVDV-2 (Pestivirus B), con 4 subgenotipos reconocidos. Por último en 2004 se identificó una tercera especie de pestivirus que infectaba a bóvidos[146], llamada Hobi-like virus o BVDV-3 (Pestivirus H), similar a las anteriores y también con al menos 4 subgenotipos [147 148].

En los 3 casos la organización del genoma y el procesado de la poliproteína es el mismo, tan sólo varía ligeramente la secuencia de nucleótidos, especialmente en la región 5' no codificante (5' NCR). El genoma codifica para 11 genes, dando lugar a 4 proteínas estructurales y 8 no estructurales (Figura 1.10).

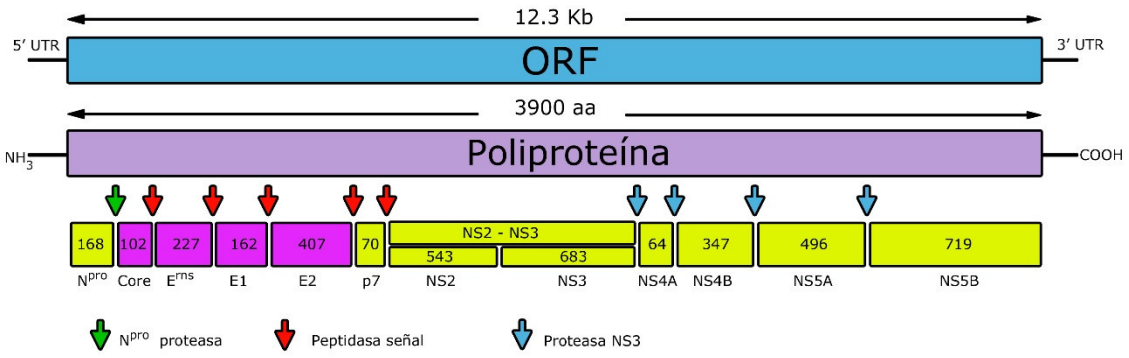


Figura 1.10. Esquema de la organización del genoma de BVDV y su traducción. El genoma contiene un único marco de lectura (ORF) (en color azul), flanqueado por las regiones no traducidas (UTR) 5' y 3'. El ORF codifica para una poliproteína única (en color morado claro), la cual se escinde y madura generando proteínas estructurales (en color violeta) y no estructurales (en color amarillo). La traducción de la proteína comienza desde el N-terminal, codificando la proteasa Npro, con actividad cisteína proteasa, que provoca su autoescisión (flecha verde). Las proteínas estructurales y la proteína p7 son probablemente escindidas por peptidasas señal de origen celular (flechas rojas). El resto de proteínas se escinden gracias a la actividad serin-proteasa de la proteína NS2-NS3 (flechas azules). Esta última proteína puede ser escindida en las proteínas NS2 y NS3, hecho que sirve como marcador en el biotipo citopático (cp).

1.6.2. Importancia económica de BVDV.

La infección causada por BVDV tiene un impacto mundial, afectando no solo al ganado vacuno sino también a otros como el bovino o el caprino. Las múltiples manifestaciones clínicas hacen difícil su diagnóstico, provocando grandes pérdidas económicas en el sector primario. Resulta complicado calcular las pérdidas económicas a nivel global, puesto que dicha cuantía en ocasiones ha sido estimada en base a las pérdidas generadas por animal, en otros casos por rebaño y en otros a nivel de producción[149].

Tomando en cuenta diferentes estudios, los costos directos derivados de la infección por BVDV en ganados vacunos lecheros es de entre 21 y 135 € anuales por animal[150]. El número de cabezas de ganado vacuno es superior a 1.400 millones en la actualidad[151], con un significativo porcentaje de animales infectados por el virus, lo que se traduce en la pérdida de decenas de miles de millones de euros anuales.

Por último hay que añadir el costo que supone realizar una campaña de control efectiva para su erradicación. En varios sitios donde ha existido este control se ha comprobado que la relación costos/beneficios es desfavorable al inicio, recuperándose la inversión al cabo de pocos años[152].

Chapter 2:

OBJECTIVES

Archaeal viruses HCIV-1 and HHIV-2

1) **Structural characterization of the HHIV-2 and HCIV-1 virions at high resolution by cryo- electron microscopy.**

   a) Atomic model of the major capsid proteins (MCPs)
   b) Atomic model of the rest of the relevant proteins for the viral capsid assembly.
   c) Structural modeling of significant protein junctions for viral capsid assembly.

2) **Characterization of the cellular receptor binding complexes of HHIV-2 and HCIV-1 virions.**

   a) Obtainment of the atomic model of the complex proteins.
   b) Structural modelling of the receptor complex on the 5-fold axis junction.

3) **Evaluation of the structural similarities and differences of HHIV-2 and HCIV-1 (both belonging to *Alphasphaerolipovirus* genus) and their context within the PRD1-Adenovirus lineage.**

Animal-infecting envelope virus BVDV

1) **Morphological characterization of the BVDV virion by Cryo-electron tomography (Cryo-ET).**

2) **Morphological characterization of conformations other than the mature BVDV virion.**

3) **Description of the organization of the viral lipid envelope.**
   a) Organization and distribution of the Erns and E2 glycoproteins in the BVDV lipid envelope.
   b) Structural modeling of the specific binding of Erns glycoprotein to the WB210 monoclonal antibody.
   c) Structural modeling of the specific binding of glycoprotein E2 to monoclonal antibody WB214.

Chapter 3:

MATERIALS AND METHODS

3.1. Sample preparation methods

3.1.1 Growth and purification of the viruses

Archaeal viruses HCIV-1 and HHIV-2

For HCIV-1 and HHIV-2 viruses, the processes of production, purification and measurement of virus concentration were conducted at the University of Helsinki by D. H. Bamford's group, a long-term collaborator of Dr Abrescia's group.

3.1.1.1 HCIV-1 production

The host of the virus is *Haloarcula californiae* (ATCC 33799), which takes 4 days to grow aerobically at 37º C in liquid culture of MGM broth. The "over-4-nights" host is inoculated in 8-12 flasks to obtain enough cell production (about 2 l), which takes one more day to reach absorbance values of 0.6 (OD = 550 nm).

At that point, assuming that the concentration of the host culture at 0.6 is 6 x 10^8 CFU/ml, the culture is infected with a MOI = 15. Then, it takes another 2 days to get a virus lysate with an absorbance of 0.3-0.4. (OD = 550 nm) [70].

3.1.1.2 HCIV-1 purification

The cell debris is removed from the liquid cultures (F12 BioSeal rotor, at 10,900 g, 30 min, 4°C). Right after, polyethylene glycol (PEG 6000) and NaCl are added to a final concentration of 10% (w/v) and 0.5 M respectively. Then, the viral particles are precipitated from the supernatant by a constant agitation at low temperature (4º C), achieving the aggregation of the viruses after 30 min approximately and centrifuging again (F12 BioSeal rotor, at 10,900 g, 30 min, 4°C). This previous step allows a more effective centrifugation, getting more amount of precipitated virus, which is resuspended in the same buffer and stored overnight on ice.

Then, aggregates are removed (Eppendorf 5415D centrifuge, at 15,800 *g*, 10 min, 4°C), and the sample is subjected to rate-zonal centrifugation in linear sucrose gradients, made by gradient mixer devices. These gradients are very useful to purify enveloped viruses such as HCIV-1. The particles travel through the gradient until they reach the point in the gradient at which their density matches that of the surrounding sucrose, appearing clearly visible in a blue scattering band. Firstly the sample is added on the top of 5-20 % (w/v)

sucrose gradient in appropriate buffer, getting the 1X-band after centrifugation (Sorvall AH629 rotor, at 103,600 g, 75 min, 15°C)..

The 1X-band is collected and further purified by equilibrium centrifugation to obtain the 2X band. This time CsCl gradient (ρ = 1.4 g/ml in HCIV-1 buffer) is chosen, which leaves cleaner virus preparations, but requires longer centrifugation. (Sorvall AH629 rotor, at 72,000 g, 16-20 h, 20°C).

The 2X band is diluted 2-fold in HCIV-1 buffer, in this case using 18% salted water without NaCl). Later on, the sample is centrifuged at high speed (Sorvall T647.5 rotor, at 32,000 rpm, 3 h, 20°C) to obtain a small pellet with high concentration of the virus.

Finally, the pellet is resuspended in a minimum amount of buffer and put on ice at least for 3 hours before the sample could be shipped, keeping continuously the sample at low temperature until its use in the laboratory.

Buffer used for HCIV-1:

1M NaCl, 70 mM MgCl2, 20 mM KCl, 1 mM CaCl2, 50 mM Hepes (pH 7.2). In some purifications 50 mM Hepes was changed for 50 mM Tris–HCl, remaining the rest of the buffer at the same conditions.

3.1.1.3 Concentration determination of the purified HCIV-1:

Once produced and purified, the concentration of the sample was determined using 2 different methods:

Bradford assay: This consists in preparing a standard curve from the stock solution of BSA (Bovine Serum Albumin), adding Bradford reagent into each sample, and measuring the protein concentration by a Multiskan reader. Straightaway, the sample concentration was determined interpolating different concentration values in the standard curve. [153]

Tricine–SDS-PAGE (Shägger gel): This technique, in addition to separate proteins according to their electrophoretic mobility, allows the evaluation of the quantity for each protein according to the thickness of each band. The presence of tricine increases the resolution for proteins smaller than 30 kDa.

3.1.1.4 HHIV-2 production

The host of the virus is *Haloarcula hispanica* (ATCC 33960), which takes 3 days to grow at 37º C in solid agar. For reasons still unclear, the growth in liquid medium for HHIV-2

is lower than expected. Hence, the virus stock is obtained directly from semi-confluent plates for the preparation of the virus lysate, taking the top infected layer of a large number of plates to achieve high production volumes (3 ml of MGM broth are added per plate, obtaining about 1 L of Lysate).

Next, the lysate was shacked during 1.5 h at 37°C and immediately centrifuged (Sorvall SLA3000 rotor, at 10,900 g, 20 min, 4°C), removing the cell debris.

At that point, taking that the concentration of the host culture at 0.8 absorbance is 6 x 10^8 CFU/ml, the culture is infected with a MOI = 10. To reach that concentration the culture was incubated for one day with aeration at 34°C, until get a virus lysate with an absorbance of 0.3-0.4, (OD = 550 nm), enough to start the purification process.

3.1.1.5 HHIV-2 purification

The cell debris of the lysate was removed by centrifugation (F12 BioSeal rotor, at 10,900 g, 30 min, 4°C), This step and the forthcoming ones follow the same methodology as that described for HCIV-1 with a few exceptions:

The use of PEG 6000 and 5-20 % (w/v) sucrose gradient does not change, just the buffer. However, to obtain the 1X band, the virus centrifugation changes slightly (Sorvall AH629 rotor, at 83,000 g, 20 min, 20°C)

The 1X band is layered on top of CsCl gradient (in HHIV-2 buffer solution) and centrifuged (Sorvall AH629 rotor, at 83,000 g, 16 h, 20°C), getting the higher purified 2X band.

Finally, the pellet is resuspended on ice for 3 hours in a minimum amount of buffer

The host of the virus is *Haloarcula californiae* (ATCC 33799), which takes 4 days to grow aerobically at 37º C in liquid culture of MGM broth. The "over-4-nights" host is inoculated in 8-12 flasks to obtain enough cell production (about 2 l), which takes one more day to reach absorbance values of 0.6 (OD = 550 nm). [83]

3.1.1.6 Concentration determination of the purified HHIV-2:

The concentration measuring methods are the same that those used for HCIV-1, using this time a SDS-Polyacrylamide Gel Electrophoresis without tricine. The electrophoresis was prepared in an ATTO gel (small gel) with 16% separation gel and 5% stacking gel.

Animal-infecting envelope virus BVDV

For BVDV virus, the processes of production, purification and measurement of virus concentration were conducted at the Friedrich-Loeffler-Institut in Germany, carried out by M. Beer's group.

3.1.1.7 BVDV production

The host of the virus is *Bos taurus,* and the cell line used for virus production are esophagus bovine cells, KOP-R cells, obtained from the Collection of Cell Lines in Veterinary Medicine (RRID: CVCL_D289). This cell line is the one with the highest replication efficiency among those tested for BVDV strain CP7 [154] .

KOP-R cells are inoculated with CP7_Erns_flagII at a MOI = 0.2 in liquid culture of MEM (Minimal Essential Medium) without FCS. After 48 h of infection, cell culture supernatant is harvested and cleared through centrifugation (4,000 rpm, 5 min, 4°C).

3.1.1.8 BVDV purification

Polyethylene glycol (PEG 6000) and NaCl are added to the collected supernatant getting a final concentration of 7% (w/v) and 0.4 M respectively (pH7.2). Next, the mixture is constantly stirred under low temperature (4°C), accomplishing the aggregation of the viruses after 4 h and immediately centrifuged (10,000 rpm, 30 min, 4°C). The pellet is resuspended in NTE (100 mM NaCl, 10 mM TrisCl, 1 mM EDTA, pH7.2) and incubated overnight at 4°C.

The resuspension is then purified by anti-FLAG M2 magnetic beads. These beads contain 4 % agarose bound with the anti-FLAG M2 (mouse monoclonal) antibody. The M2 antibody recognizes the FLAG sequence at the N-terminus, Met-N-terminus and C-terminus, allowing the detection and capture of our viruses, which contain a FLAG peptide sequence.

For that, 4 ml of the resuspension are mixed with 2x 200µl of NTE-washed magnetic beads and incubated for 1 h at room temperature under stirring conditions. The supernatant is collected and the beads are washed three times with 1 ml of NTE buffer to remove any possible cell protein residues. To elute the virus from the beads, 2x 250 µl of synthetic FLAG peptide is added to the beads at a final concentration of 100 mg/µl in

NTE buffer and incubated for 1 hour at room temperature under stirring conditions. The elute fraction is harvested and 2x 250 µl of synthetic FLAG peptide are added again for 30 min. Next, all elute fractions collected are put together and centrifuged (45,000 rpm, 2h, 4°C). Finally, the resulting pellet of purified virus is resuspended in a minimum amount, about 50 µl, of PBS buffer.

The purified sample is then stored at 4 ° C in the laboratory, it is kept refrigerated on ice during shipment and stored at 4 ° C again until be analyzed by electron microscopy.

3.1.1.9 Concentration determination of the purified BVDV

At every step of the purification, aliquots were collected and the titer (TCID50/ml) was determined. This titer is based in an endpoint dilution method in which only the number of infectious and replicating particles are assessed.

First the virus suspension is diluted in a proportion 1:10, doing 6 replicates per dilution step. These dilutions are plated in a 96 well plate with KOP-R cells. After 3 days of incubation the wells are screened for typical cytophatic effects. Aditionally, as a double check, the wells are stained with an antibody against NS3 that can be detected by immunofluorescence. Finally, the number of positive and negative wells are counted and the concentration is calculated.

3.1.2 Sample preparation for Cryo-EM

The different tests for the preparation of optimal grids for visualization in transmission electron microscopy were performed 'in house' at the Abrescia's lab at the CIC bioGUNE laboratory (Spain)

3.1.2.1 Negative Staining

Negative staining is an electron microscopy technique that allows samples to be checked thanks to the presence of heavy metal salts solutions, in which the sample is embedded. The contrast is generated between the metallic solution, which strongly disperses the electron beam and the volume occupied by the specimen, where the electron beam is less scattered. The problem is that the resolution is limited by the size of the heavy salt grains to about 15-20 Å.

This technique has been used for an initial visualization of our samples due to several reasons:

a) Its preparation is quick and allows us to obtain preliminary information on the integrity of the sample.

b) The radiation damage of the sample is very low, since the organic molecules are protected by the heavy metal layer, which allows a higher electron dose.

c) In the case of our viruses, they appear in greater number on the carbon surface of the grid than when the sample is vitrified, which allows us to quantify in an objective way the amount of virus in the sample.

All samples are prepared on copper grids (QUANTIFOIL® Carbon, 300 mesh), to which a glow discharge is first applied (10 mA, 2 x 10^{-1} mbar in air, 35 sec, in a Glow-discharge/Carbon evaporation MED 020 BALTEC). The generated plasma causes the ionization of the surface of the carbon film, which is negatively charged, favouring the adsorption of the sample.

Immediately, a drop of 8-10 µL of the purified sample is deposited on the grid over the side where the plasma is applied. The drop is incubated for 2 min and then is washed once in the corresponding buffer to remove the excess of protein that has not been attached to the grid surface.

Then, the drop of buffer is dried with a piece of filter paper until a thin layer of liquid is left on the grid surface. Finally, the grids are stained with a drop of 8 µL with either 2% uranyl acetate (w/v) or 2 % methylamine vanadate (w/v) (Nanovan ®). After 1 min the excess of liquid is also removed with filter paper and the rest of the stain solution is dried at room temperature.

3.1.2.2 Vitrification

Vitrification is a technique used in Transmission Electron Cryomicroscopy (CryoTEM), in which the sample is quickly frozen in a container with cryogenic liquid. The rapid freezing prevents the formation of ice crystals, creating an amorphous solid, called vitreous ice. This solid preserves the sample in its native state, with no denaturation, obtaining images of the sample closer to the real conformation. Finally, the vitreous ice section must be less than 500 nm thickness, as in thicker sections the electrons would be unable to pass through the sample due to the resulting dispersion events.

Before vitrification all the grids are hydrophilized in high vacuum coating system (7-8 mA, 2 x 10-1 mbar in air, 35-40 sec, in a Glow-discharge/Carbon evaporation MED 020 BALTEC). This time different grids were used depending on the sample.

For HCIV-1 and HHIV-2 200-mesh Quantifoil R 2/1 holey-carbon copper grids were used. The concentration for purified HCIV-1 was 1.0–1.2 mg/ml while for purified HHIV-2 was 0.9–1.2 mg/ml. The amount of pipetted sample per grid was 3.5-4.0 µl in both cases.

Regarding purified BVDV wt, the concentration was 1-4x10^8 TCID50/ml. This sample was used in different incubation tests together with specific monoclonal antibodies until the most suitable conditions, observed by TEM, were found.

The first preparation consists of mixing BVDV wt with a specific monoclonal antibody (mAb) for glycoprotein ^{E}rns, called mAb WB210. The amounts used in the incubation (for 1-2 grids preparation) were 5.5 µl BVDV (1-4x10^8 TCID50/ml) + 1.2 µl WB210 (1mg/ml) kept for 3.5 h at 5 ºC. The grids chosen for vitrification were Quantifoil 3.5/1 on 200 copper mesh, coated with an extra carbon layer added manually. In the grids finally used for data collection, 7 µl of sample were utilized (the sample was added for about 20 min as it evaporated to finally be vitrified at half of the initial volume added, thereby increasing the number of viral particles per volume).

The second preparation is analogous to the first one, but this time the mixture consists of BVDV wt incubated with a specific monoclonal antibody (mAb) for glycoprotein E2, called mAb WB214. The amounts used in the incubation (for 1-2 grids preparation) were 5.5 µl BVDV (1-4x10^8 TCID50/ml) + 1.2 µl WB210 (1mg/ml) kept for 6 h at 5 ºC. The grids chosen for vitrification were QUANTIFOIL® Carbon, 300 mesh (in cases where the attachment to the surface of the viral particles with other grids was challenging) or 200-mesh Quantifoil R 2/2 holey-carbon copper grids

In the third preparation only BVDV wt, called 'BVDV bold' was used due to the absence of antibodies bound to the lipid envelope. This time 4 µl of BVDV (1-4x10^8 TCID50/ml) were used per grid.

All the grids were vitrified using a FEI Vitrobot™ Mark III, except for one vitrification of BVDV + mAb WB214, where a FEI Vitrobot™ Mark IV was used. For that aim to be achieved the next steps are followed:

First a tank is filled with liquid nitrogen (−195.8 °C), keeping the rest of the tools cold during the entire vitrification process. Inside this tank a smaller container is filled with liquid ethane, which remains in a liquid state while cooling down until it reaches its solid state (-182.8°C). The sample must be vitrified with the ethane in a liquid state, at a temperature close to its freezing point.

Second, the grid is held by tweezers at the bottom of a plunger, inserted inside a camera where the temperature and the humidity are controlled and the sample is added to the grid inside the camera, where it is incubated for the desired time before freezing.

Third, the blotting of the sample is done on both surfaces of the grid pressing it against filter papers, adjusting the liquid sample layer based on the selected settings. Immediately, the plunger drops the grid into the liquid ethane.

Lastly, the vitrified grids are introduced into the gridbox and stored in liquid nitrogen until visualization

3.2 Data Collection Methods

To obtain images of our vitrified samples it is necessary to use a transmission electron microscope. The principle is based on an electron source, from which the electrons are led to the sample by electromagnetic lenses. When electrons hit the sample, some of them get through it and others are dispersed. The electrons that can pass to the other side of the sample (and conserve the initial energy) are captured by a detector thus giving rise to an image.

On the above basis, we can use this technique in different ways depending on which method is most suitable for our objective. In this work two different techniques (Cryo-EM and CryoET) have been used that rely on transmission electron microscopy, probably the most relevant ones today.

Before using these techniques, vitrified samples were inspected in our ‘in house’ microscopes to verify quality and conditions.

In a first testing the sample was treated with negative staining and visualized under a JEM-1230 transmission electron microscope (JEOL, Japan), equipped with an Orius SC1000 (4008×2672 pixels) cooled slow-scan CCD camera (GATAN, UK).

Similarly, a second test was carried out in a JEM-2200FS/CR transmission electron microscope (JEOL, Japan), equipped with an UltraScan 4000 SP (4008×4008 pixels) cooled slow-scan CCD camera (GATAN, UK), as a previous evaluation before the samples were sent to a high-end microscope.

3.2.1 Cryo-EM Single particle collection

Single-particle Cryo-EM is a technique consisting on the collection of 2D electron micrographs, snapping individual particles in each one of them. As the particles contained in the sample are frozen in random orientations, different 2D projections of the same particle are obtained. Getting a large number of 2D views for each orientation, each projection vector inserts the Fourier values into the correct plane of a Fourier volume, getting an initial 3D map. The different conformations of 3D maps can also be classified. Finally, the resulting maps will be constantly refined and validated until the final structure of the particle is obtained at the best possible resolution.

This technique is the most suitable for the three-dimensional reconstruction of the capsid of our viruses under study, HCIV-1 and HHIV-2, since it offers us certain advantages over other 3D reconstruction methods. On the one hand, icosahedral viruses, due to their high symmetry, can be treated as 20 trimeric building blocks, totalling 60 subunits that can be treated as equivalent for each virus. Therefore the structural unit of the virus is redundant and is called icosahedral asymmetric unit (IAU). On the other hand, these viruses are very large, so is relatively easy to align them and the applied defocus is not determinant for their identification. It is not necessary to select a big defocus to generate the necessary contrast required to identify the particle, and therefore choosing a lower range of defocus the high resolution is preserved.

The acquisition of images in the case of HCIV-1 took place at the electron Bio-Imaging Center (eBIC) inside the complex of Diamond Light Source (UK). The provided microscope was a state-of-the-art Titan Krios, with 300 Kv accelerating voltage. 3,218 micrographs were collected, containing each one a movie of 27 frames, adding up to a total electron exposure of 36 $e^{-}/Å^{2}$. A Falcon II™ Direct Electron Detector camera (4096 x 4096 pixels, 14 *μm pixel)* recorded the images at a nominal magnification of ×59,000, leading to a pixel size of of 1.40 Å. The selected defocus ranged from 0.6 μm to 3.9 μm.

In the same way, the acquisition of HHIV-2 images for single particle Cryo-EM was held at the Netherlands Center for Electron Nanoscopy (NeCEN). The supplied microscope was also a Titan Krios with 300 Kv accelerating voltage, equiped with a Falcon II™ Direct Electron Detector camera and a Cs-corrector. 2,786 micrographs were collected, containing each one a movie of 7 frames, adding up to a total electron exposure of 35 e^{-} /Å². The nominal magnification of the recorded images was ×60,000, leading to a pixel size of of 1.34 Å. The selected defocus ranged from 0.7 µm to 3.0 µm.

3.2.2 Cryo-electron Tomography collection

Electron cryotomography (Cryo-ET) is a technique consisting on the collection of 2D electron micrographs from the same individual particles at different tilt angles, with the aim of combining all the collected views to represent a 3D reconstructed volume, called tomogram.

The principle is based on applying low doses of electrons on the sample (to preserve it from radiation damage) while the sample is tilted, rotating orthogonally to the incident electron beam, obtaining images at different angles.

As the sample is placed in the holes between a latticed copper support of the grid, there is a physical limit for tilting the sample, due that the incident beam ends up interfering with the metallic lattice at high tilts. However, empirically the tilting of the sample is even lower, since the increasing sample thickness acts as a barrier for the incoming electrons, which lose their energy due to inelastic scattering. Then, the tilt range is restricted to ±60 or ±70°, losing the high angles up to 90°. These absent region is called missing wedge due to its shape and causes lack of information in the Fourier space, generating ray artefacts, distortion effects and structural elongation when the tomogram is reconstructed.

A second major challenge is related with the signal-to-noise ratio (SNR). The electron dose per tilt has to be low to prevent radiation damage in the sample and therefore the resolution is compromised as the particles details are slightly appreciable from the background. Again, the increasing thickness of the sample at high tilts produce lower contrast in the projections and the SNR also gets worse.

These two physical impediments make this technique achieve an absolute resolution lower than the single particle Cryo-EM. Nevertheless Cryo-ET is the most appropriate technique for the study of pleomorphic viruses, such as BVDV and polymorphic interactions, such those between BVDV glycoproteins and their specific antibodies. Furthermore, repetitive or frequent appearing elements within these pleomorphic structures can be averaged together, creating a higher resolution map.

All the Cryo-ET data were acquired at the electron Bio-Imaging Center (eBIC) inside the complex of Diamond Light Source (UK). The chosen microscope was a Titan Krios, with 300 Kv accelerating voltage, equipped with a Gatan Quantum K2 Summit Direct Electron Detector camera [155] (3838 x 3710 pixels, 5 μm pixel), a Volta phase plate (which enhances the contrast close to focus) and a 'On-axis Rotation Tomography Holder'.

The sessions intended for the obtainment of a set of tomograms in which BVDV is mixed with monoclonal antibodies against Erns glycoprotein presented the following settings:

A total of 76 tilt-series were collected, with a range from -60° to +60° degrees inclination. Altogether 41 images resulted in one tilt series, with an angular step size of 3 degrees, starting from 0° and alternating each 12 degrees from positive to negative tilt in a semi dose-symmetric scheme. The dose per frame was 0.25 $e^{-}/Å^2$, for a total dose of 80 $e^{-}/Å^2$ per tomogram. Each of the tilts contains a movie of 8 frames. The magnification of the acquired images was 18,050.5 giving a pixel size of 2.77 Å at the specimen. The selected defocus was maintained in 0.1 μm underfocus during all the process.

A comparable procedure was applied when BVDV mixed with anti-E2 antibodies was collected, although this time some settings were improved trying to get higher quality images:

The images have been collected in super-resolution mode, converting a physical pixel of the detector in four 'super-resolution' pixels. These pixels are only virtual, but are based in a real physical detection. Technically is possible to discern each individual electron event thanks to the rapid counting of the detector. As the interaction area of the electron is slightly bigger than the size of one physical pixel, it is detected with a different degree of signal in several adjacent pixels, making possible to detect the center of the event with a sub-pixel precision.

Also the tilt-scheme was optimized using a dose-symmetric mode in which the positive and negative alternation was each 6 or 9 degrees (depending on the session), starting at 0°. This option maximizes the high-resolution transfer in the lower tilts, where the SNR is better before the radiation damage starts burning the sample. Regrettably, the optimal dose-symmetric setting (alternating the tilting direction each 3 degrees) was not possible due to time constraints.

In addition, the dose per tilt is slightly increased, passing from 8 to 10 frames per tilt, which provides two extra images to correct the motion drift without excessively increasing the total dose, just from 100 $e^{-}/Å^2$ to 116 $e^{-}/Å^2$

With these settings mentioned above, a total of 35 tilt-series (more in process) were collected analogously to what was done for the BVDV-anti-Erns data. The dose per frame was 0.28 $e^{-}/Å^2$, for a total dose of 116 $e^{-}/Å^2$ per tomogram. Each of the tilts contains a movie of 10 frames. The magnification of the acquired images was 72,411.3 and 80,000, giving a pixel size of 0.69 Å and 0.625 Å respectively. The defocus chosen was 0.1 µm or 0.3 µm underfocus when the first was not possible.

Besides these BVDV-antiE2 tomograms with improved settings, 38 tomograms more were collected with intermediate settings between those of BVDV-antiErns and these of BVDV-antiE2 explained above. In this case the acquisition was also in super-resolution (7676 X 7420 pixels images) but acquired at a lower magnification of 36,443, giving a pixel size of 1.372 Å. The number of frames per tilt was increased to 15, with a dose per frame of 0.23 $e^{-}/Å^2$, so the total dose was 141 $e^{-}/Å^2$. The reason for increasing the total dose was due to the higher background of the grid, which was compensated by a higher dose per tilt. Most of the imaged were acquired at 0.1 µm underfocus.

Specialized software was used for the tomography collection. SerialEM was the preferential software, although Tomography 4.0 software was alternatively used. Both programs have the phase plate and dose-symmetric acquisition implemented.

3.3 Processing & Computational methods

3.3.1 Processing of single particle Cryo-EM data

3.3.1.1 Motion correction

Once the sample collection process in the microscope is finished**,** the main objective is to keep the maximum amount of information present in the frames throughout the whole computational processing. This is one of the basis for obtaining a final high resolution map of our structures.

The main factors that cause image blurring during the frame-recording are the instability of the sample stage and the mobility produced by the incident electron beam on the sample. These factors are mitigated by fractioning the acquiring of the image into a stack of several frames called movie, which offers certain advantages. On the one hand the movement caused by the beam is minimized when taking shots with low exposure time and on the other the stage movement is reflected in each frame. Nonetheless, the signal-to-noise ratio (SNR) remains constant in each frame, reducing the signal per frame. Then this strategy can only benefit if the microscope camera readout is nearly noiseless and efficient in counting, such as the used K2 camera.

Then MotionCorr [156] program was used to align the frames of the movie within only one output micrograph. This software is based in an algorithm which is able to align the frames of a movie exploiting the redundant information between nonadjacent frames, getting a high accuracy of the common position of the pixels and therefore increasing the resolution.

In the case of HCIV-1, the movies contained a stack of 27 frames, of which 26 were included for motion correction. Similarly, HHIV-2 movies were formed by 7 frames, taking 6 of them for motion correction. In both cases the first frame of the stack was rejected as it showed high drift respect to the rest of the frames.

3.3.1.2 CTF estimation and CTF correction

The contrast transfer function (CTF) provides the phase changes of the scattered electrons when they interfere with the sample respect to the electrons continuing a direct beam, represented in a 2D image. Therefore, the CTF describes graphically in reciprocal space the credibility of the real image based in mathematical parameters that are defined in the 2D CTF image.

The ideal CTF should contain the information for all the spatial frequencies of the real image, but this is not possible due to several factors, some due to the physical properties

inherent to the microscope and others generated manually by the user. These are the most important factors:

Acceleration voltage: At higher voltage the multiple scattering is lower, as the electrons have more energy (higher wavelength), gathering less contrast but extending the resolution limit of the CTF.

Spherical aberration (C_S): It happens when the electrons pass by the objective lens. The electrons crossing peripherally the lens have a different focal plane from those crossing at the center.

Defocus: When the electron beam is unfocused (under focused) the number of phase shifts is increased valued by the user. Thus, when defocus is applied the higher resolution is lost but the contrast is increased,

These two last parameters define the contrast, so if lens were aberration free at focus 0 the CTF would not have any ring.

Amplitude contrast: This contrast arises due to the difference in the amplitude between the elastic and inelastic scattering waves of the electrons. It depends on the thickness and the mass of the sample, so in our case (biological samples of viruses and thin thickness) is really low (around 0.07).

Astigmatism: Is an aberration that affects to the symmetry of the CTF rings in X or Y. It is caused when the beam is exposed to an uneven magnetic field surrounding the lens.

Therefore, both defocus and astigmatism are the two factors essentially changing in each of the individual images collected in a session under the same conditions. Thence, each individual CTF parameters can be measured by different softwares and be represented compared to the expected CTF in the named ”CTF estimation”. Then, the defocus of the real CTF is tried to be fitted to the theoretical one, correcting the phases and the amplitudes in a process called “CTF correction”.

In the case of the micrographs of HHIV-2, the CTF was estimated and corrected using the software CTFFIND3[157], whereas for HCIV-1 a new version of the software, CTFFIND4[158], was utilized. The differences between both softwares are not so pronounced, just CTFFIND4 is faster in the processing and may give better CTF estimations as it includes all parts of the micrograph for that.

3.3.1.3 Picking and extraction of the particles

Coordinating all the necessary steps for image processing in single-particle Cryo-EM can result a tedious task. In consequence, Relion [159] software was chosen. This program encompasses the whole process tracking all the performed jobs, as it links the output from one job when it is used as an input for a new job. Moreover, the workflow is easily followed thanks to a GUI (Graphical User Interface).

Therefore, once the movie of frames has been corrected from beam induced motion, the resultant micrograph is included on the list under Relion. Then, the micrographs of the list are manually examined, excluding those that have not our interested sample (in the case of our viruses it can be checked by eye as they are of big size and perfectly recognizable) or those useless (surface with big cracks, broken or superposed particles, dirtiness, etc.). This step is called "micrograph inspection" in Relion.

Next, the previously mentioned CTF estimation software is run from Relion, choosing the best settings from the GUI, discarding also the micrographs with a non-suitable CTF.

Following that, the next step to compute is the "manual picking", in which some viruses are manually picked from the micrographs to generate particle references, which are used as template to recognize the viruses from the rest of the micrographs this time using the "auto-picking".

Finally the picked particles have to be extracted in order to get a set of cut viruses to be classified and aligned between them. This step is called "particle extraction". Accordingly, 4,584 particles of HCIV-1 were extracted with a box size of 768 × 768 pixels. In the same way, 14,877 particles of HHIV-2 were extracted with a box size of 760 x 760 pixels. The boxes are dividable several times by 2 to facilitate the calculation of the algorithms used during the next steps of the processing, reducing the computation time. In both cases the square box only included the icosahedral structure of the virus, which its radius limit is defined by the external capsid, decreasing also the computation time. The receptor complexes of the viruses were excluded from the box, as they followed a different reconstruction method called "Localized reconstruction".

3.3.1.4. 2D classification

After picking all the putative particles that will take part in the next processing steps, we used the 2D classification tool in Relion. This option allows you to group the selected

particles according to their similarity into several classes. To do this, it takes as references those that were generated manually to carry out the autopicking. Among the generated classes, different rotational and translational orientations of the same particle appear. Depending on the morphological heterogeneity of the sample, even more classes will appear with their respective orientations. Each class contains a variable number of particles, the most populated being those that represent the most common morphological type of the particle observed from its preferential orientations. The groups with less members are those that include less common morphologies, non-preferential views and false positives of auto-picking. Using the maximum likelihood estimation algorithm, after some iterations, the final 2D averages for each class are obtained as output, selecting those that will be used for the 3D classification.

On these terms, the 4,584 HCIV-1 particles were run through 2D classification selecting different number of classes (2, 4 and 10). From the 10 classes, 9 displayed views referring to the same morphological particle, the highly packaged one, while just one view formed by 84 particles was morphologically different, showing a virion class-average in which the membrane diameter was clearly smaller. Hence, the first 9 classes were merged together, totaling 4500 particles while the last class was kept separated. Finally both classes were analyzed by 3D classification.

On similar terms the 14,877 HHIV-2 particles were also run through 2D classification, obtaining similar results to those on HCIV-1. Although differentiated 2D classifications were also performed, this time it was decided to send the whole set of particles directly to 3D classification. This would not be possible with other samples, but in the case of our viruses the particles have been visually inspected and all the particles are classified in some valid class, not existing junk classes.

3.3.1.5. 3D classification in symmetrical viruses

Once the 2D classification of the particles is done it is necessary to apply the correct settings before starting the 3D classification.

First it is important to know what type of symmetry our virus presents. There are two basic patterns of capsid symmetry on viruses, which are the helical (h) or the icosahedral symmetry (I). A set of rotational operations can be applied for each symmetry, obtaining equivalent positions (but unique) in each virus. The icosahedral viruses for instance, have

60-fold symmetric places for particle. On many occasions this is not known in advance and a cyclic 1-fold symmetry around a single axis (C1) is applied to obtain the initial asymmetric volume. In our case we already know that both, HCIV-1 and HHIV-2, are arranged in an icosahedral symmetry (I). In particular the particles were aligned in I2 symmetry, which is an orientation convention for x, y, and z axes along the 2-fold axes of the icosahedron.

Second, a 3D reference structure is required to star the 3D classification. A good reference is crucial for a successful classification, although it cannot be really detailed, as it has to be a model of consensus that in a way represents all the different structures existing in the heterogeneous dataset[160]. In the absence of a 3D reference, if there is a reasonable distribution of 2D viewing directions of the particles, a *de novo* 3D model can also be obtained. In the case of HCIV-1 virus a suitable model was obtained from the HHIV-2 high resolution map, filtered to 60 Å by a low pass filter. This references worked because both viral capsids are practically of the same size and the Triangulation (T = 28) is also the same. For HHIV-2 the initial reference was directly taken from a low-resolution Cryo-EM HHIV-2 map, which was obtained in a previous project performed with in-house collected data.

Third, to get the optimal orientation for each of the particles an angular sampling is applied. The aim is to obtain the best Euler angles for those particles. For that the reference map is used to assign the 3 particle orientation parameters and 2 position parameters. This is done by matching the projections obtained from the reference map with the experimental images of the particles based in the maximum-likelihood scoring criterion. Therefore, the size of the angular sampling determines the number of 3D orientations that are searched during 3D classification. In our data we started with an angular sampling of 7.5 degrees, quite big but enough for a geometrical particle to be oriented. At each iteration the orientational assignments change and are based in the previous iterated model, improving the resolution and allowing and smaller angular sampling.

Another element that is crucial for 3D classification is the use of a proper mask. The concept of masking implies a multiplication with maps. The particle map is multiplied by a map with values of zero for the non-interested regions. In this way, the SNR is reduced and the particles align better, as the more variable regions of the virus can be excluded.

Since 2D classification, Relion uses a spherical mask outside of the given particle diameter. In 3D classification an extra mask is added inside the virus structure, masking part of the inner membrane and the RNA, which are more disordered than the capsid lattice.

Finally, as in 2D classification, it is really important to choose the proper number of classes to get the different homogeneous groups of particles and discard those groups with an ambiguous shape. In HCIV-1 2 classes were given for classification. The most populated one, with 3414 particles improved its resolution as the iterations progressed, while the other class was stuck during the process, showing poorly defined structures. For HHIV-2, the 14887 particles were distributed in three classes, being the majority class the most accurate one, with 11446 particles. The best classes of both viruses were sent to 3D refinement.

3.3.1.6. 3D Refinement and Postprocessing

When a class of particles is sufficiently homogeneous, this can be used in the step called "3D auto-refine", where the aim is to align at finer angles. To do this, the following must be done:

First, the homogeneous subset of particles must be splitted into 2 random halves with the same number of particles. These halves will be refined independently one from each other, and then used for the gold-standard FSC assessment (see below).

Second, the angular sampling will be automatically changed by the software until convergence, which in our case reached an angle smaller than 1 degree.

Finally, to prevent the bias in the high frequency components, the map obtained in the 3D classification is filtered again at low frequency to be used as reference. In our case the maps were filtered to 50 Å.

The independent refinement of the 2 halves prevents the iterative build-up of noise, as when both reconstructions are compared using the Fourier Shell Correlation (FSC) the true pixels that exist in common in both halves are detected, avoiding the adding of noise and the over-refinement of the map. The FSC is also used to estimate the resolution of the reconstructed map. The measured points are standardized at a threshold of 0.5, which means a cross-correlation of 0.5 between the 3D volumes and the same procedure is done

at 0.143 cross-correlation. Depending on at what spatial frequency this cross-correlation is reached, the final resolution will be higher or lower.

The FSC calculation of the volume is performed on the unmasked whole map after the convergence of 3D refinement. Therefore the true resolution is in some way underestimated. Thus, some postprocessing calculations are necessary to get a higher final resolution map:

First, a tight mask based in the threshold is performed. This masks is slightly thicker than the confidence boundary of the particle. The densities of the features in the map are not affected, the difference is that during the FSC calculation the less cross-correlated pixel of the map are excluded.

The second factor is related with the modulation transfer function (MTF). This function is related with the ability of the camera to emphasize the real contrast of the object in the acquired images at each resolution. If the MTF of the camera is know it can be deconvoluted, obtaining a higher resolution volume.

Third, a b-factor is applied to the map. This factor increases the weighting of the higher resolution data of the structure. The b-factor makes a sharpening in the map that can be adjusted "ad-hoc" until the resulting map certainly displays an interpretable density. The b-factor applied in the HCIV-1 and HHIV-2 maps was -40 Å^2 in both cases.

Finally, using Relion-3, the last version available of the software, the Ewald correction was also applied to the maps, as it is recommended for big size structures . This concept is associated with the defocus. During the whole processing a single defocus value has been calculated for each 2D particle image. Therefore, if it is assumed that the 2D image has different defocus values along Z, it is possible to correct the defocus in blocks fitting the local mean defocus. This correction significantly increases the final resolution in big particles with long Z. In fact, this made possible to increase de resolution in 0.4-0.6 Å in our maps.

3.3.1.7 Localized Reconstruction method

HCIV-1 and HHIV-2 viruses have a capsid formed by highly ordered proteins distributed in a clear icosahedral symmetry. However, these do not occur in their specific receptors, from now spikes, which are attached to the icosahedron in each of the 5-fold in a flexible manner (Figure 3.1). Therefore, the alignment of the spikes from images of the entire

particle is not a correct strategy to obtain high resolution data. On the one hand the spikes of the same particle will have a heterogeneous variability greater than the capsid, since they are more flexible and may contain more stable conformations. On the other hand, the symmetry of the spikes is different from icosahedral, which is directly observable in the 2D images (Figure 3.1).

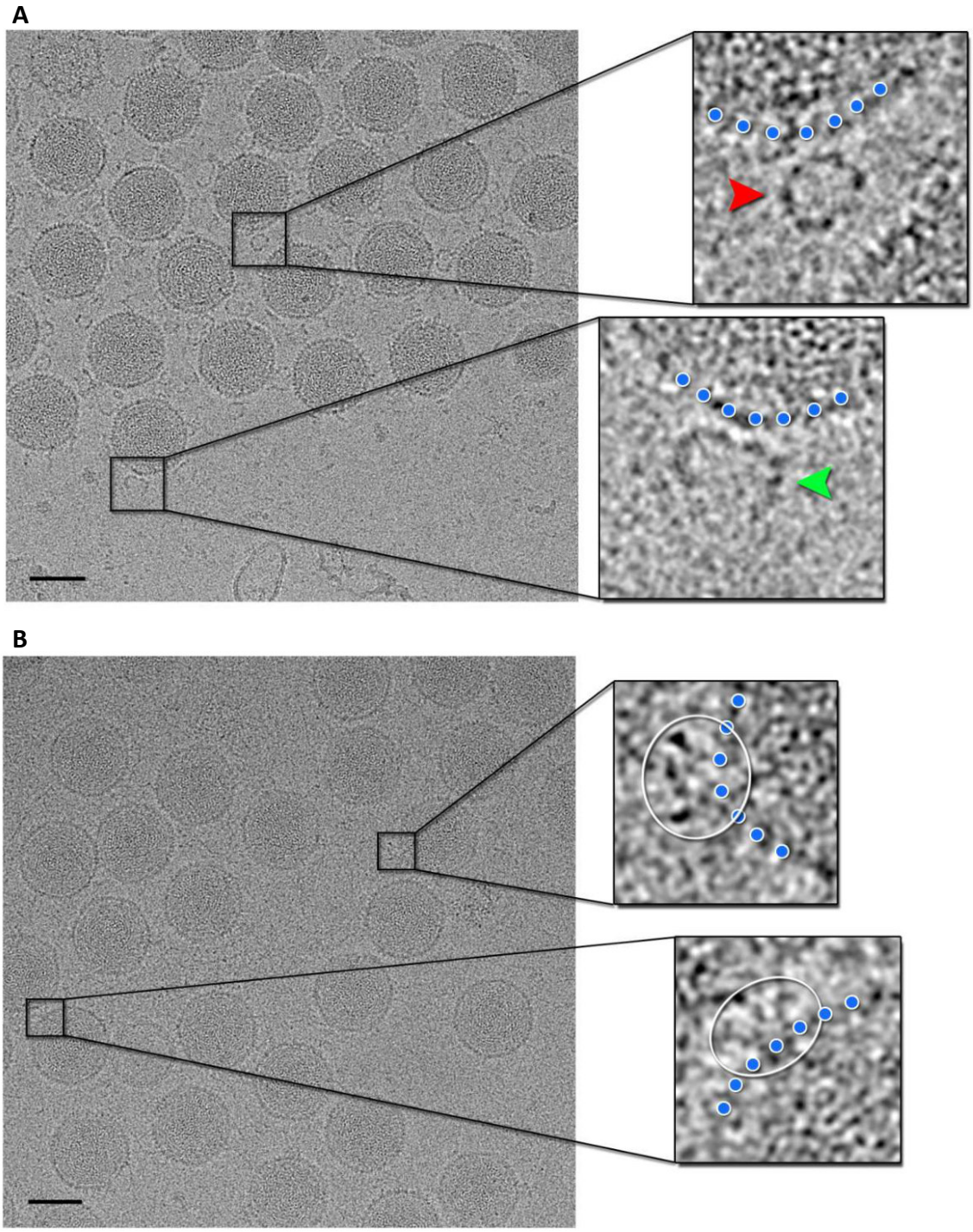


Figure 3.1. Raw images of vertex complexes. (**A)** Overall view of a 2D cryo-image of HCIV-1 sample (estimated defocus: -2.5 µm) with black square marking some of the vertex complexes. Insets, enlarged views of the dimeric vertex complex in close- (red

arrow) and open-conformation (green arrow) respectively, blue dots schematically mark the capsid profile; insets have been gaussian blurred to facilitate their recognition. **(A)** As (B) but for HHIV-2 virus (estimated defocus: -1.5 μm) in this case the vertex complex is pentameric and surrounded by white circles. Scale bar, 50 nm.

Faced with this problem, another alignment strategy was chosen to solve the structure of the spikes, which is called "Localized Reconstruction" [161] . The approach of this technique does not differ greatly from that proposed for the capsid, but some new steps are introduced:

In the first place it is necessary to have a file with the best particle orientations and origins of the whole virus. This can be obtained from the last 3D refinement performed.

In the second place, a 2D image of the whole particle containing also the area occupied by the spikes is required. Therefore the same 2D images used for the icosahedral reconstruction are re-extracted, this time with a box size of 1056 × 1056.

In third place, a script is run using "relion Localized Reconstruction in python". In this script we use the re-extracted particles and the best particle orientations together with a vector and a length, which define the location of the spikes relative to the 3D reconstructed particle. The output of the script are a new set of 2D images, called subparticles, existing 12 per particle, referred to each of the spikes positioned in the 5-fold vertices.

This set of subparticles can be align together by other script to create a draft starting reference for 3D classification. At the same time a soft mask is created in Bsoft [162] from the map of the whole particle, giving a value of 1 to all the pixels showing a density. Then, the density of the reference subparticle is deducted from that map multiplying the mask by the subparticle with a value of 0. The result is a mask that includes the whole virus except the area of the spikes.

After that, the previous script of "relion Localized reconstruction" is re-run, but this time adding this new mask. The result is a new set of subparticles in which the densities that do not correspond to the spikes have been subtracted, which makes easier the subsequent reconstruction.

At this point the scheme is analogous to that used for single-particle reconstruction in the whole particle in relion, although the symmetry is different.

In HCIV1, a total of 40,968 subparticles resulted from the 3414 particles used for the 3D refinement of the capsid. The rest of spikes present in viruses of other classes were not taken into account since their coordinates in the extraction of subparticles could vary slightly.

To know the degree of heterogeneity of the spikes, a 2D classification was first assigned by testing different numbers of classes. Using a 2D classification of three classes, two of them contained a high percentage of lateral views of the spikes, while the third contained frontal views and was discarded. This is due to the fact that the frontal classes are oriented in an axis different from that of the lateral spikes and also have a high background since the capsid of the virus is in the same area of vision, either in front or behind the spike. The two classes with high amount of lateral spikes were merged (56%; 22,804 particles) and a new 2D classification of 5 classes was made. Then, the three most defined classes were merged, summing-up to a total of 12,034 particles (29% of the total subparticles). Finally, a last 2D classification was performed assigning 10 classes (Figure 3.2, B).

To know the 3D structure of the spike of HCIV-1, first all the subparticles were taken (40,968) and a 3D classification of just one class was performed. In this classification the prevously generated draft starting reference was used again as reference, and a C2 symmetry was imposed (Figure 3.2, A). The resulting volume was filtered to 60 Å and used as a new reference. After that, again a 3D classification was executed designating 4 classes in a C2 symmetry. In 2 of these clases (48%; 19,664 spikes) any structure was deduced, just heterogeneus density. In the other 2 classes (24 % and 28%) the volumes had a structural homegeneity, representing two different conformational states of the opening of the spike. Finally both classes were refined independently by 3D refinement.

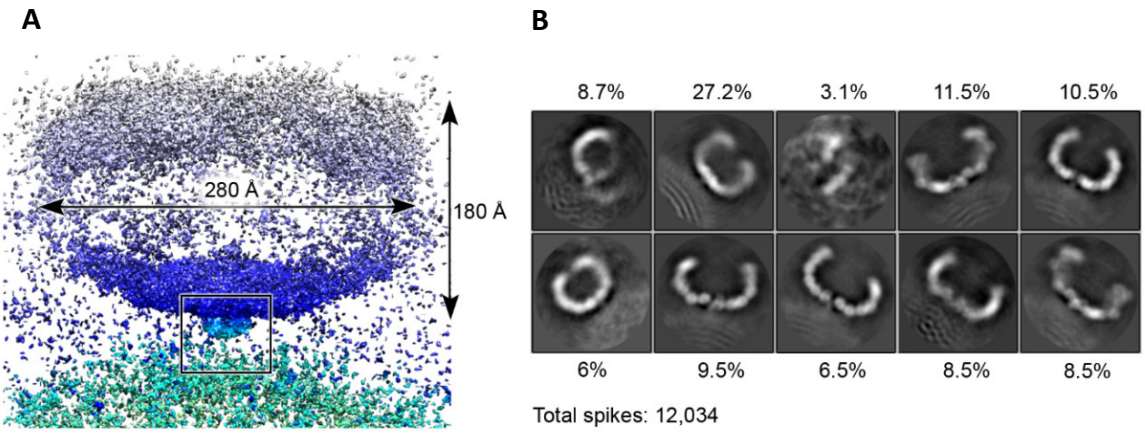


Figure 3.2. Processing of HCIV-1 vertex complex reconstruction. **(A)** Overall view of the HCIV-1 spike before localized reconstruction; black rectangle marks the density corresponding to the region stemming from the penton of the capsid to the 'horn' (site of the symmetry mismatch). **(B)** 2D classification processing, giving ten distinct classes of the dimeric horn-like spikes attached to the five-fold vertices showing their structural flexibility. The percentages indicate the class abundance over the total number of spikes analysed.

In HHIV-2 a total of 137,352 subparticles resulted from the 11,446 particles used for 3D refinement of the capsid. Another advantage of "Localized reconstruction" is the possibility of knowing the real number of spikes that are present per virus, which is called occupancy. By treating each spike independently you can find out the total number of volumes in which the spike is present or absent. In a classification with icosahedral symmetry, however, absent spikes would simply reflect a lower average signal at that location.

The spikes of HHIV-2, unlike those of HCIV-1, maintain a vertical axis at the 5-fold vertex, which makes their orientation taken from the extracted particles to be quite accurate and similar in all of them. Therefore, a 3D classification was carried out in which the alignment was skipped, turning out in 3 classes that were clearly differentiated by the degree of occupancy. The spike was absent in the 35.5 % of the cases, while other 22.5 % had lost the fiber. The class average showing the entire spike represented a 42%. This last class was sent for 3D refinement alignment, applying a C5 symmetry.

3.3.2 Processing of Cryo-ET data

3.3.2.1 Motion correction

The factors that affect the sample collection in CryoET are the same as for Single-particle alignment, but these do not affect exactly the same way. Therefore the correction of the movies also varies:

First of all the frames acquired in each movie contain a much lower SNR. This is because the dose per frame in Cryo-ET is much lower (in the order of 5-7 times lower in our case) than in single-particle. In addition the thickness that has to cross the incident beam in high angle tilts is also greater, increasing the SNR.

Secondly, not all frames receive the same dose of electrons, since frames with less tilting expose a smaller surface area and therefore receive higher doses of electrons than more

tilted frames, where the incident electrons are distributed between the surface of interest and the surrounding margins that are not present in all movies.

Thirdly, defocus is quite consistent in low-angle tilts, but as the inclination increases, a raising defocus gradient is created in the frame, making correction difficult.

Fourth, it should be noted that the frames collected last have suffered greater degradation of the sample since the mobility produced by the incident ray has been accumulated, as well as burns in the particles of interest. This clearly affects the correction between frames.

Finally, as the sample stage must move for each of the collections at different tilts the observed drift is somewhat greater, although an appropriate time is expected before the acquisition of images in each change to avoid this effect.

As a novelty, BVDV tomograms were corrected with a new software, called MotionCor2 [163]. This program performs a correction that is not limited to the global motion but it is also able to correct local motion. The local motion divides the frame in patches of smaller size that are aligned pixel by pixel independently and then this patches can be overlapped, correcting smoothly the movie. All the set of BVDV tomograms were corrected dividing the frames in 5X5 patches with an overlap of 20 % between them. Furthermore, in the movies corresponding to tomograms of BVDV mixed with anti-E2 the anisotropic magnification and the dose weighting of the frames was also corrected.

3.3.2.2 Tomogram reconstruction

If a single 2D image is observed, it would be impossible to determine the exact position of the object in the 3D space. Therefore, the images of each tilt *per se* do not provide three-dimensional information if they are not combined in the correct way.

There are several specialized softwares that carry out the correct alignment of 2D images to finally apply a reconstruction algorithm that fits as realistically as possible to the original shape of the object. For the collected data of BVDV, IMOD [164] software was selected, specifically etomo [165] , which is a 3D reconstruction program with a friendly GUI to progress along the whole processing.

The reconstruction of the tomogram begins with a pre-processing, where in each of the 2D images the extreme pixel values that could cause artifacts are eliminated. These values

can be caused by artifacts from the CDD camera, such as putative X-rays hitting it or physically broken pixels. In addition, the average gray values in tilts at high angles have a lower signal than in the central tilts, so all 2D images are normalized together to facilitate alignment.

In the next step the 2D images are related to each other for the first time by a coarse alignment. This makes a cross-correlation that gives a crude X and Y translation of the successive images along the axis. In this way all 2D images are related forming a continuous stack at different angles. At this point, a more accurate alignment is possible thanks to the use of nanogolds, also called fiducials. The use of fiducials allows a very precise alignment, since it offers a series of additional advantages to a marker-free alignment. On the one hand, the fiducials are spherical, so their shape does not change depending on the angle of the tilt, expressing an accurate translation of the euler angles along the tilt axis. On the other hand the fiducials have a higher SNR, as they are made by gold and display a high contrast under the microscope. Normally the fiducials embedded in the sample are selected at the zero-degree tilt, trying to represent the largest possible area in X and Y, thus achieving distal reference points between the fiducials that will help with a proportional reconstruction of the tomogram. The same nanogolds are then tracked along the stack, creating a fiducial model in which each fiducial creates a thread that connects the anchor points in each of the 2D images of the stack. In the tomograms of BVDV mixed with anti-Erns 50 nanogolds per tomogram were selected, while for BVDV mixed with anti-E2 25 fiducials were chosen.

After that, a new adjustment, called fine alignment, transforms the images to represent the projection of a solid body around the tilt axis, as well as to refine the projection angles. To achieve this, each image needs to be properly rotated and translated, maintaining also the correct magnification. So local alignments are accomplished by making decisions again under the fiducials.

In the next step the objective is to define the tomogram positioning, with the aim of making the reconstruction as flat as possible and fit into the minimum achievable volume. For that a horizontal representation viewed from the Y axis (the tilt axis) is showed. Then, based on the manual selection of 3 boundaries at 3 different levels of the Y axis, the sections outside the region of interest are discarded from the later 3D reconstruction. The

output of this step results in the final aligned stack, which can be used for the creation of a 3D volume by tomogram reconstruction.

Imod proposes the two most used methods in tomogram generation, which are based on 2 different reconstruction algorithms:

The most used method is known as weighted back-projection (WBP) [166]. This is basically based on the Central Section Theorem (CST), which asserts that a projection series on real space corresponds to a series of central sections in Fourier space. So, an inverse Fourier transformation of the aligned images results in a 3D reconstruction of the sample. This function would perform an exact reconstruction if all the values of the projections along the angular variable were present, but the number of projections is limited (41 in our data). In addition the angular range is not 360 °, since there is a missing wedge, where the information is not restored. The term weight refers to the suppression of overrepresented frequencies due to the overlapping of the projections in the Fourier space. In addition, the low frequencies are better represented than the high frequencies, since in the central area of the FFT the radii of the projections are closer than at higher frequencies, away from the central point, causing blurring of the image. For this reason weighting summation is also compensated according to the distance from the origin in the Fourier space. This method was used for the reconstruction of all our tomograms.

An alternative method for tomogram reconstruction is based on the iterative reconstruction algorithms. The most representative one is the Simultaneous Iterative Reconstruction Technique (SIRT)[167] . This method involves the creation of a starting trial tomogram by simple back-projection from the tilt series. The new reconstructed tomogram is reprojected and compared with the original projections tilt angles. This operation is repeated in consecutive iterations, reducing the error between the re-projections and the measured projections. The term simultaneous refers to the fact that the corrections of all projections are calculated at the same time. The advantage of this method over WBP is better SNR and contrast. Instead, SIRT uses more computing time and performs reconstruction in real space, enhancing the low frequencies but losing the high frequency information. This method has been used to contemplate the details of the structure of the virus, but for the technique of "subtomogram" averaging WBP was used because it retains more information.

In any case, both WBP and SIRT methods present lateral shadows in the X-Y planes and X-Z artifacts due to the effect of the missing-wedge, so it must be taken into account when interpreting the data.

3.3.2.3 CTF estimation and CTF correction

As explained before, CTF depends on defocus, among other factors. The CTF causes the periodic inversion of the signal depending on the spatial frequency range, missing the information when its value is 0. Therefore, to recover the information at frequencies greater than the first zero-crossing it is necessary to perform a CTF correction. To perform this correction it is fundamental to obtain a good CTF estimation, which is achieved by fitting the theoretical CTF in the power spectrum calculated for the image. This process is usually simple for micrographs from single-particle data collection, but it is difficult in Cryo-ET data [168] . The main factors are the low electron dose used and the thickness of the sample, which increases to high tilts. Moreover, in these high tilts the defocus values vary since they are at different heights with respect to the incident beam.

Therefore, a solution to achieve high resolution information in Cryo-ET is to obtain the first zero-crossing shifted at very high resolution frequencies, omitting in this case the CTF correction. For this the defocus has to be low, but in that case contrast would be lost. Fortunately there is a device, called phase plate, which is able to modulate phase contrast without applying any defocus[169] . This device is formed by an amorphous carbon that accumulates a Volta potencial when is pre-irradiated by the electron beam. In the immediate collection this potential introduces a phase shift between the central electrons and the diffracted ones, generating a bigger phase contrast [170] . Therefore, this was the *modus operandi* in our collection, as the contrast of the antibodies linked to the viral glycoproteins must be high. Otherwise it is really difficult to recognize these links.

3.3.2.4 Picking of the particles in tomograms

The concept of particle picking in Cryo-ET is analogous to that of single-particle. The only big difference is that in Cryo-ET a volume, called subtomogram, is selected instead of an area, and therefore this region needs to be delimited by 3 coordinates (X, Y, Z). There are some packages for tomogram visualization that make easier the manual picking of the desired elements. In our case the chosen software was dynamo [171] , due to the variety of tasks available, both for picking and visualization and also for the subsequent subtomogram averaging process.

The first step is to create a catalogue [172] with all the reconstructed tomograms. From the catalogue the tomograms will be displayed one by one to perform the picking of all the desired subtomograms. Secondly, due to the big size of the tomograms (up to 52 Gb), these are binned to alleviate the maneuverability of the volume in 3D. Thirdly, the binned volume is opened by the 3D browser dtmslice, where other parameters that improve the visualization of the elements of interest are adjusted. Among these settings are the contrast control, which excludes contrast values from the total range; and the bandpass filtering, which apply a lowpass excluding high frequencies.

Finally, the elements of interest are identified from all the possible planes and picked. The picked points are saved as models, which are recorded in a data pool. Then these models are used as a central point to extract a volume or subtomogram of the dimensions necessary to include the element of interest. In our case, several lists of subtomograms were created depending on the structure selected. The size and number of subtomograms selected in each case are detailed below:

1) Subtomograms of the Standard BVDV virions.

2) Subtomograms of the inflated BVDV virions.

3) Subtomograms of the empty particles.

4) Union events between anti-Erns monoclonal Antibodies (WB210) and the viral glycoprotein Erns.

5) BVDV RNA genome.

Apart of these sub-boxed structures, other striking conformations were also labeled in models. One of these are the tethering membranes, where the envelopes of two different viruses are merged. Another structure was the solely nucleocápside of BVDV, with no RNA or envelope related.

3.3.2.5 Subtomogram averaging

The technique based on the acquisition of tomograms by single tilt axis tomography comprises a number of limitations that generate a missing wedge impossible to remove. This fact means that part of the Fourier space is not present in the subtomograms extracted from the tomogram. Assuming that all subvolumes extracted are referred to the same structure, if these are randomly oriented, in each one the absent Fourier space will be

different. Therefore the missing details in one subtomogram will be present in another and vice versa. Thus, the common information shared by the subvolumes can be put together to get a more detailed structure, closer to the real object.

This is the principle of "subtomogram averaging" technique, in which the analysed subvolumes are properly aligned and combined, generating an average subvolume with an increased signal-to-noise-ratio. This process is iterative[173] , starting when a set of subtomograms is aligned against a reference. The reference is rotated into different orientations, comparing each of them with the subtomogram[174], creating alignment scores depending on the cross-correlation. Then, the alignments with the higher score of the different subtomograms are averaged together creating a structure that will be used as reference for the following round of alignment.

The BVDV subtomograms classified into the different sets are averaged to obtain the best maps for each model. The strategy to perform the average is summarized in dynamo in an "alignment project", adding all the necessary settings and parameters by a graphical user interface (GUI).The alignment project is organized in 2 different main sections.

The first section serves to choose the input files that will participate in the process, which are 4:

1) Particles: It links a data set containing the desired subtomograms to the project.

2) Table: It is a matrix that describes the properties of each subparticle. It identifies the subvolumes and gives their initial orientation and angles. Thus, it is the main element to perform mathematical operations.

3) Template: It is the initial density, which is rotated at different angles against each one of the subvolumes of the data sate. If there is no previous reference of the model, it can be created artificially to start the project.

4) Masks: In the alignment project various mask are involved. The alignment mask constrains what part of the template takes part in the alignment against the data subvolumes. The classification mask is also important if multireference analysis are performed.

As mentioned before, subtomogram averaging is an iterative process, so in each iteration a new template will be generated and the values of the table will also change.

The second section serves to choose the input settings that will define the alignment process, which are 2:

1) Numerical parameters: These parameters define the mathematical approach of the project. The program is not autodidact, so the values entered are decided by the user from the beginning to the end of the process, being possible to change these values in each iteration. Some important values are:

Cone aperture: It rotates the template defining the vertical axis. For that, Z axis and X axis are rotated a specified range of degrees, defining the best Euler coordinates.

Azymuth rotation: It defines the rotation around the new vertical axis established, finding the third Euler point.

High pass and low pass: They define the lower and upper limit of frequencies that are not excluded, respectively. It works as a filter to align better the template against the data set.

Symmetry: A determined symmetrization can be applied to the reference before starting each round.

Shift limits: It determines the search area from a given point, normally the center of the particle.

Threshold parameter: The aligned subvolumes are taken into account depending on the CC value against the template.

2) Computing environment: The processes can be run on dynamo using different systems. Depending on the system chosen, the processing time can vary considerably. In our case, we tried, as far as possible, to use GPUs in parallel. The use of 2 to 8 GPUs equals hundreds of single CPU cores, thus reducing processing time at reasonable scales.

3.3.3 Other testing methods in subtomograms by Dynamo

For the characterization of the BVDV-Antibody complexes additional strategies were carried out. These were the main approaches performed:

1) Discard of possible antibody-glycoprotein bindings not included in the subtomogram averaging:

To verify that all antibody-glycoprotein unions had been included in the data set, the envelope of all standard viruses was cut radially into 80 pixel subtomograms (the same size chosen to extract the Glycoprotein-Ab unions). In this way 135,000 subtomograms came out, of which all those less than 60 pixels (radius + ½ radius distance of the subtomogram) away from the central point of the subtomograms already identified with glyco-mAb unions were excluded. The remaining subtomograms with apparently no spikes nearby were aligned together to see if any density was still recognizable.

2) Search for densities embedded in the membrane: The 135,000 cuts made were aligned by adding a mask on the region outside the membrane to find any possible density in the intermembrane space.

3) Randomization of the inplane angle: Once the Glyco-mAb alignment is performed, the output model is vertically and azimuthally oriented. A good method to demonstrate that such alignment has not been the result of chance is to create new templates from the particle data set. These templates will retain the vertical axis but will be azimuthally randomized to create unpredictable templates. These templates are used to create a new alignment. After several iterations, if the output model is similar to the one obtained before randomization it is a good proof that this model is reliable.

4) Noise adding: Another way to demonstrate the reliability of the aligned model has been to create several templates from this model by adding random noise of value 3. As a result, each template has a different density and is somewhat more unspecific than the original model. Again, if the alignment is re-run with this noisy models the output model should be similar to the original.

5) Creation of a cross-correlation matrix (ccmatrix) to be used in a principal component analysis (PCA):

The PCA is a technique used to describe the analysis of our “data set” through a series of significant characteristics that are represented in different volumes, called “eigenvolumes”. These volumes highlight differentiating characteristics of our model, and therefore the contribution of all “eigenvolumes” represents all the morphological variability in our data set.

6) Spatial distribution of the Erns glycoprotein in the viral lipid envelope:

This process has been performed in order to know the separation between the Erns glycoproteins of the lipid envelope in each individual virus. Taking the coordinates of the central points of each subtomogram (sub-boxes cut from the specific Erns-Ab union), a matrix of distances between the individual points of each virus has been made. On the other hand those same linking points (their coordinates) have been used, assigning them random points on the surface of the same viruses. This has been repeated in a thousand different tests, which serves to demonstrate whether our selected points follow a random pattern or if on the contrary there is a specific and characteristic disposition of our Erns-Antibody junctions.

3.3.4 Nano-LC–MS/MS for protein determination

Unfortunately, not all the structural proteins present in the HCIV-1 and HHIV-2 virions can be identified by fitting of the sequences into the densities achieved in the final maps of the virus model. Nevertheless, there are alternative techniques with high power resolution that allow the identification of candidate proteins to occupy these densities.

One of these techniques is Nanoscale liquid chromatography coupled to tandem mass spectrometry (nano LC-MS/MS)[175] , which is indicated to solve complex samples coupling the analytical capabilities of two different techniques, which are liquid chromatography and mass spectrometry.

In the final maps of HCIV-1 several protein densities were not identified. On the one hand, 5 homopentamers located in the 5-fold membrane and associated with the vertex complex and on the other hand 2 different proteins associated with the membrane and located under the protein capsomers were not associated with any sequence.

Although these proteins are located in the membrane and presumably possess transmembrane elements, their close relationship with the vertex complex and with the capsomers respectively makes the formation of soluble protein clusters very likely. Thus, the following procedure was carried out to identify these protein clusters and analyze the protein composition of them. These protocol was accomplished by the Proteomics Platform led by Dr Elorza at the CIC bioGUNE (Bilbao, Spain).

First, 45 µl of HCIV-1 sample (7.5 mg/ml) contained in a specific buffer (1 M NaCl, 70 mM $MgCl_2$, 20 mM KCl, 1 mM $CaCl_2$, 50 mM Tris–HCl, pH 7.2) were incubated at RT for 30 min.

Second, the 45 µl were filtered in a Vivaspin® 500 centrifugal filter 1000 kDa MWCO (Beckman Coulter Microfuge 22R F241.5P fixed-angle rotor, 12,000 rpm, 7 min, 20°C).

Third, the flow-through was collected to be used in a Blue Native Polyacrylamide Gel Electrophoresis (BN-PAGE)[176] , which is a native type gel that is able to separate the different membrane complexes by molecular weight without producing denaturation and identifying the bands thanks to Coomasie blue (Gradient bis–Tris gels, 4–16% acrylamide)

After that, gel bands were excised and washed with ultrapure water to immediatelly be subjected to in-gel digestion. Then, In-gel-digested peptides were resuspended in 10 µl 0.1% formic acid and sonicated for 5 min.

Next, the peptides were separated by online nanoliquid chromatography (nLC) and then analyzed by electrospray tandem mass spectrometry (MS/MS) (nanoACQUITY UPLC system (Waters) connected to an LTQ Orbitrap XL ETD mass spectrometer). For reliable protein identification, at least two peptides were required to pass the $p < 0.01$ filter (high confidence). Briefly, protein relative abundance was inferred from spectral counting (defined as the total number of spectra identified for a protein in a sample) where the number of spectral counts (SpC) identifying a protein are divided by protein's length (L), and further divided by the sum of SpC/L for all N proteins in that particular sample. Therefore, relative abundance values are expressed as the % of the total normalized signal that corresponds to each protein. These calculations were carried out over the data obtained in each of the three bands detected in a total of four native gels prepared with samples derived from two different virus preparations. Average NSAF values were calculated for each protein in each gel band, and these values were visualized in a hierarchical clustering heatmap using Perseus software [177] .

3.3.5 Model building and fitting

Archaeal viruses HCIV-1 and HHIV-2

The final objective of a structural work cannot be concluded without a proper model building and refinement of the atomic models of the proteins. This model has to be in

accordance with the experimental observations, having a consistent physical, chemical and biological sense. The folding of a specific protein can be approximated by two different strategies. When similar proteins are already described, a building by homology can be performed, which consists of picking from the structural data base 3D structures of related homologous proteins that can be used as templates, as they probably have similar folding and motives. If unluckily there are not homologous models, it is necessary to use a *de novo* protein structure prediction, which by algorithmic processes predicts a tertiary structure from the protein primary sequence. Once the protein model has been broadly docked to the map density a refinement process is performed. This is an iterative process, which based in different modelling parameters, improves the fitting of the molecular model to the density of the experimental data. The model building and refinement was different for each virus and protein:

For HCIV-1 the atomic models of the MCPs, VP7 and VP4, were manually built by protein homology from the crystal structures of VP16 and VP17 of P23-77 virus. This virus is from the same structural family and displays the same β-barrel cores in its MCPs. Then poly-ALA full models were generated and refined in the graphical software Coot [178]. The initial refinement for VP4 and VP7 was initially performed in two MCPs of the IAU, and afterwards the individual density of each MCP comprised in the IAU was properly fitted. With the poly-ALA models adequately fitted in the densities, they were refined in PHENIX [179] software by phenix.real_space refinement [180], which by several secondary structures restrains further adjusts the real position of the atoms. The model was refined again, this time for the whole capsid MPCs, by Rosetta [181] , detecting and correcting automatically the placement errors of the atomic models. Finally the models were refined for the last time in PHENIX v1.13.

In the case of VP9 (gene *27*—Uniprot code: A0A1C7A3R7), the good quality of the map made possible to identify the protein in the density of the 5-fold. The sequence was manually registered from residues 44 to 146 and the rest was modeled as poly-ALA. The rest of the procedure was followed as in the MCPs.

For the proteins underneath the capsomers (VP10 and VP12) and the 5-fold vertex (VP13), the density of the map did not allow the recognition of the side chain of the proteins, but most of the secondary structure was identified. The secondary structure prediction performed by I-TASSER [182] software and the length of the candidate proteins

were taken into account. Furthermore, in VP13, a homopentameric protein with a similar transmembrane regions and sequence similarity was found, fitting also properly in the density.

Finally, the proteins forming the spike (VP2, VP3 and VP6) did not have enough resolution to fit the backbone of the sequences. Therefore the secondary structure prediction and modelling was interpreted by I-TASSER.

In the same way, the model building and refinement was carried out in HHIV-2. As the sequence identity between the MCPs of HCIV-1 and HHIV-2 is higher than 70 %, the MCPs of the first were used as homologous models to manually fit the backbone and the side chains. Then, the models were also refined in PHENIX and Rosetta.

The penton protein VP9 was localized among all the putative genes (gene *22*—Uniprot code: H9AZX8) due to the same relative locus position presented in HCIV-1 genome. The procedure was then analogous to the other VP9, fitting manually the atomic model except for the first 40 residues, which were disordered.

In the proteins underneath the capsomers, the density pattern in HHIV-2 was clearly similar to HCIV-1, represented by the same equivalent proteins. Unfortunately, the quality of the map was slightly lower in this areas and only the VP12 could be modelled in the map.

Lastly, one of the spike vertex proteins of HHIV-2, VP16, showed discernible individual β-strands in the internal domains, where model building was performed creating *ab initio* ALA-models in COOT. Another non-identified polypeptide showing 5 copies was recognized in the 5-fold density and the Cα backbone was built. The rest of the spikes proteins (VP2 and VP17) could not be reliably traced but previous studies and the map densities identified the position of these.

Animal-infecting envelope virus BVDV

In BVDV virus the model building and refinement of the different averaged subvolumes was more challenging as the resolution obtained by Cryo-ET was considerably lower than from the maps achieved by single-particle Cryo-EM.

For the docking of the glycoprotein Erns into the corresponding CryoET densities, the available crystal structure of the Erns dimer (PDB code: 4DW5) was used as template. The specific antibodies for each glycoprotein were taken from a crystal structure of a similar IgG monoclonal antibody (PDB code: 1IGY). All these models were manually fitted first in chimera [29] and then by rigid body fitting in Coot. Finally, the Erns areas putatively in contact with the Fab regions of the antibodies were identified and interpreted consequently.

3.3.6 Segmentation of Cryo-EM volumes

The aim of segmentation in the Cryo-EM tomograms is to identify characteristic regions in the density of the reconstructed volume that correspond to recognizable individual components. The high pleiomorphy of BVDV particles allows the visualization of many different morphologies of the virus in a same tomogram. Also in the same tomogram it is possible to see different states of assembly of the virus as well as virions other than the standard, with viral RNA released, empty or fused vesicles, etc.

Therefore, Survos [183] assisted 3D segmentation was performed in some of the tomograms reconstructed presenting high variability of BVDV particles. SIRT reconstruction has been preferred for the segmentation as the contrast of the tomogram is better defined, enhancing the difference between the viral elements.

The first step carried out in Survos was a preprocessing of the data. Denoising filters were applied to the tomogram to improve the quality and after that some textural filters were used to enhance the relevant features.

Secondly, the elements with common characteristics inside the volumetric data are represented based in a 3-layer hierarchical determination. These layers are voxels, supervoxels and megavoxels. Each one groups at different level nearby pixels that show common features, forming in that mode boundaries and regions of pixels grouped by similarity.

Finally, once the data have been segmented, measurement and analysis tools have been performed in the corresponding models. Measures such as volumes occupied for each structure, it shape and average intensity have been extracted from the data and represented in graphics.

The volumes were also segmented by the segger tool in Chimera, where it is really easy to segment the volumes at a defined density threshold. The volume dimensions where compared to those obtained by Survos to verify the correct measurement of the volume encompassed by the different BVDV structures.

Chapter 4:

RESULTS IN HCIV-1 & HHIV-2 RECONSTRUCTION

4.1 Organization of the viral capsid

After completing the process of 3D auto-refinement and post-processing of HCIV-1 and HHIV-2 viruses, two maps were obtained in which the atomic models of several viral proteins were built and refined, obtaining a global framework of how is the capsid organized.

The first observable feature is that both viruses display a caged polyhedral arrangement formed by 12 pentagons at the 5-fold vertex and 270 hexagons covering the rest of the facets and defining the 3-fold and 2-fold vertices of the icosahedral capsid. This organization follows the Caspar-Klug Theory [57] of the quasi-equivalence, which affirms that the protein subunits forming the capsid are in slightly different environments, existing only certain numbers of subunits able to be packed into a regular lattice. These distinct subunits are geometrically defined as equilateral triangles inside the pentagons and hexagons. Thus, if we attend to the triangulation number[184] we obtain that exist strictly 28 equivalent protein positions on the icosahedral capsid (T = 28) in both viruses (Figura 4.1). Attending to the lattice coordinates, this viral capsid is right-handed oriented, as the integers *h* and *k* indicate. (h = 4 k =2) showing therefore, a *dextro* capsid configuration. This subset of quasi-equivalent proteins define the icosahedral asymmetric unit (IAU), which is repeated 3 times in each facet. This pseudo T = 28 *dextro* symmetry is the same found in the rest of the *Sphaerolipoviridae* family viruses, such as SH1[185] and p23-77 [74 186].

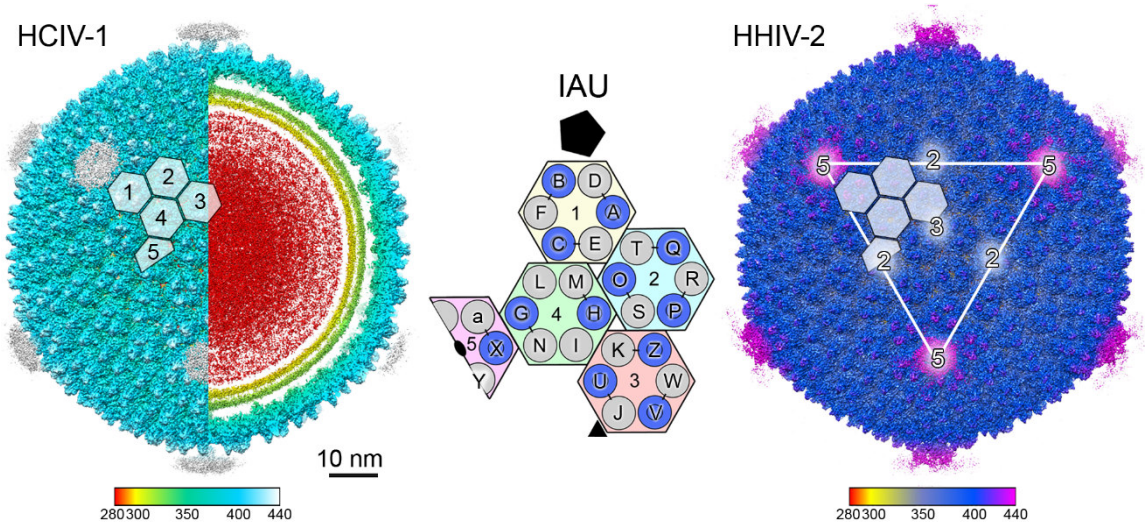


Figure 4.1. Cryo-EM density maps of HCIV-1 and HHIV-2 and MCPs. Overall view of Cryo-EM density maps of HCIV-1 (left) and HHIV-2 (right) color-coded by distance from the center (legend below). HCIV-1 is rendered to display the capsid shell (left-half) and the particle interior (right-half), genome in red and inner and outer membrane leaflets in yellow and yellow-lime (vertex complexes have been omitted (see methods). White-transparent hexagons on top of the capsid densities mark the capsomers (numbered in HCIV-1 Nos. 1–5) forming the icosahedral asymmetric unit (IAU); the white triangle and numbers on HHIV-2 surface mark a facet of the virion and the icosahedral symmetry axes, respectively. At the center, schematic of the capsomers organization with the three-tower capsomers (No. 1, light-yellow; No. 2, cyan; No. 3, pink), composed of three copies of MCP VP4 (blue circles) and three copies of MCP VP7 (light-gray circles) and with the two-tower capsomers (No. 4, light-green; No. 5, light-magenta; the latter sitting on the icosahedral two-fold axis) composed of two copies of VP4 and four copies of VP7; the five triangles composing the black pentagon represents the five copies of the penton protein plugging the vertices. Characters (A–Z and a) identify each MCP subunit within the IAU and black short-lines joining the circles identify the VP7–VP4 heterodimers.

Then, the IAU contains 12 copies of the MCP VP4, another 15 copies of the MCP VP7 and 1 copy of VP9, this last one located at the 5-fold vertex (Figure 4.2, A). The MCPs VP4 and VP7 are arranged in six-coordinated capsomers shaping a pseudo-hexameric conformation. Instead, protein VP9 is coordinated in groups of 5 homologous proteins to create a pentamer shape structure at the 5-folds.

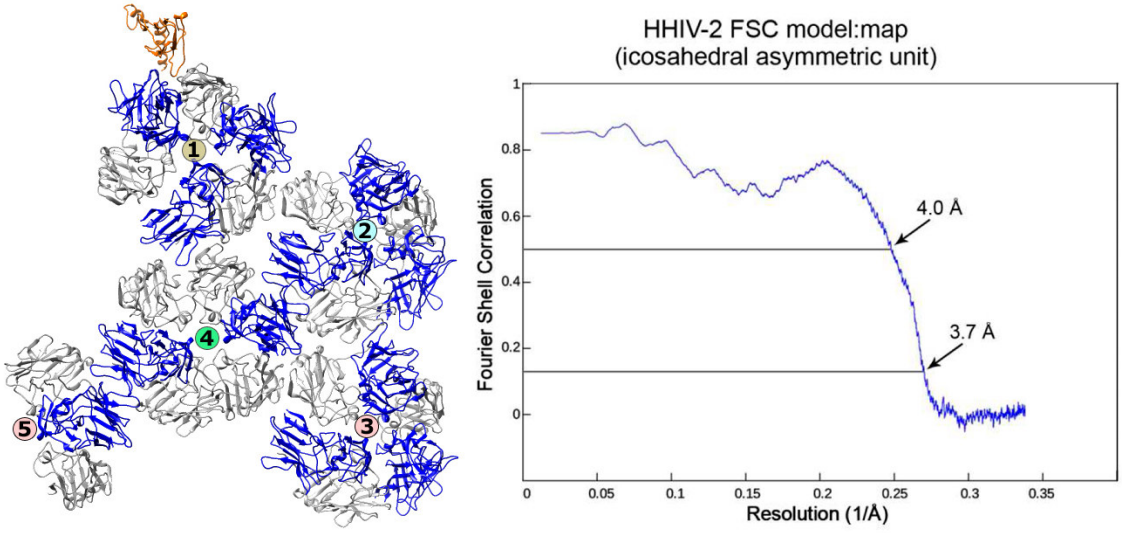


Figure 4.2. Icosahedral Asymmetric Unit (IAU) of HCIV-1. **(A)** Cartoon representation of the HCIV-1 IAU (PDB code: 6H9C). The MCPs VP4 (blue color) and VP7 (light-gray) are arranged in different manners depending on the capsomer (numbered 1-5 as in Fig.4.1), forming three-tower (1-3) and two-tower (4 and 5) capsomers. The protein VP9 at the 5-fold vertex (orange color), also belongs to the IAU. Underneath the mentioned proteins one and half copies of GPS-II, one copy of GPS-III and one copy of an unknown protein fragment complete the IAU (not in the picture). **(B)** Masked FSC calculated between the atomic model of the IAU of HCIV-1 and the corresponding boxed map using phenix.mtriage [180].

Although pseudohexamers are always formed by VP4 and VP7, there are two different conformations of capsomers in the IAU. The first group is made up of 3 capsomers (numbered as 1, 2 and 3), which comprise alternate copies of VP4 and VP7, displaying a crenelleted shape (Figure 4.2, A). The taller VP4 has 3 copies remembering towers or turrets, while other 3 copies of VP7 are disposed as crenells between the towers. Capsomer 1 is the closest to the 5-fold, being the only one in making contact with the VP9 protein through two sub-units of the capsomer. Capsomer 2 occupies a central position between the 5-fold and 3-fold axes. Its structure is equivalent to capsomer 1, but the subunits are rotated 60 degrees respect to the first one, so that the positions occupied by the vp7 subunits are exactly at the place where VP4 subunits were located and vice versa. Capsomer 3 displays the same subunit orientation that capsomer 2 and is the closest to icosahedral three-fold axis.

The other pseudohexameric conformation is characteristic of capsomers 4 and 5, which are built by 2 distal VP4 with 4 VP7 inbetween, resembling a 2-fold symmetry. In regard to capsomer 4, it is the only one connected to all the rest of capsomers conforming the IAU. Likewise, capsomer 5 shows the same MPC arrangement as the latter, but this time the pseudohexamer is shared by 2 adjacent IAU at the 2-fold axis (Figure 4.1). Thus, this capsomer is splitted into two identical halves, belonging to each IAU the VP4 subunit closest to it and the flanking Vp7 subunits.

4.2 Description and interactions of the main capsid proteins

The IAU of HCIV-1 and HHIV-2 capsids is formed by 28 subunits which display non-symmetry-related environments, according to the quasi-equivalence theory [57]. Thus each MCP shows minor distortions in the folding, especially at the N and C terminals, and also at the outermost regions and connecting loops involved in MCPs packing within and across the IAU (Figure 4.3). Despite the singularity of each subunit, the unions between MCPs across the capsid follow a distinctive pattern of links. In this respect, there are two different connections between subunits. On the one hand there is a subunit packing across the IAU between neighbouring capsomers (Figure 4.3-4), and on the other hand there is a closer assembly union between the VP4-VP7 heterodimers within the capsomers (Figure 4.9).

Packing of the subunits VP4 and VP7 in the context of the neighbouring capsomers

The pseudohexamers that form the capsid lattice in HCIV-1 and HHIV-2 form junction points with their neighbourhood capsomers in clusters of three capsomers that leave an empty space of an equilateral shape triangle inside. In these unions of 3 capsomers, the junctions do not occur directly between the 3 MCPs that converge at the center, but between one of the MCPs of the axis and another MCP of the adjacent capsomer that does not participate in that axis (Figure 4.4). Thus, in each union of 3 capsomeres, not only 3 MCPs participate, but 6 MCPs form a binding triangle that strengthens the fit of the 3 capsomeres. The bindings occur via adjacent strands and connecting loops with some ARG residues forming H-bonds or salt bridges with neighboring subunits (Figure 4.3, B).

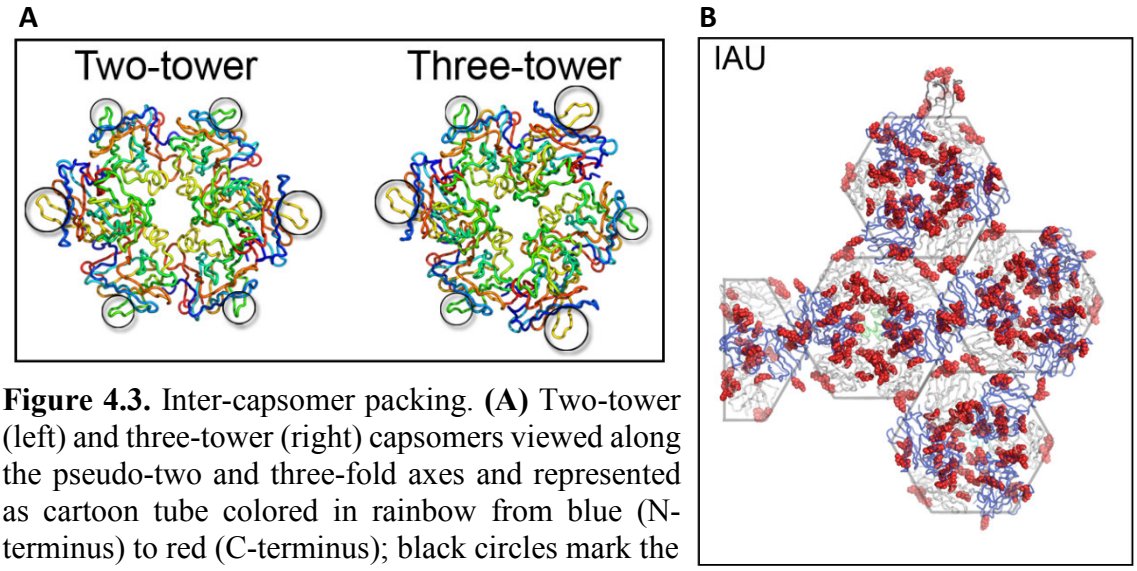


Figure 4.3. Inter-capsomer packing. **(A)** Two-tower (left) and three-tower (right) capsomers viewed along the pseudo-two and three-fold axes and represented as cartoon tube colored in rainbow from blue (N-terminus) to red (C-terminus); black circles mark the outermost regions and connecting loops involved in MCPs packing within and across the IAU (GPS-II and III proteins omitted for clarity). **(B)** Location of ARG residues as red-spheres within the IAU, some of them at the periphery of the capsomers are conceivably involved in H-bonds and salt-bridges across subunits.

Therefore, all MCPs bind, in addition to their heterologous MCP within the capsid (if they have any) to an MCP of an adjacent capsomer. In this context there are 3 different packing motifs depending on the participating MCPs (Figure 4.4). First we have a pseudo two-fold of related VP4 residues. Second we also have a pseudo two-fold of related VP7 residues. Third we have a heterologous packing between VP4 and VP7 related residues. Spatially there are 9 points of union between different capsomers (7 inter IAU and 2 intra IAU), in which all the MCPs of the IAU participate except for the VP7 subunit closest to 5-fold. Among these different combinations, there are 13 different inter-capsomer packing motifs, 3 belonging to the first type, 4 to the second and 6 to the third. In all 6

MCPs binding triangles there is at least one VP7-VP7 junction except in the 3-fold, where all joints are heterologous. Regarding the VP4-VP4 junctions, 2 of them are intra IAU, while the third maintains cohesion in the pentasymmetrons around 5-fold. Finally, the VP4-VP7 junctions are present in all 6 MCPs binding triangles, being the only packing motif in the 3-fold.

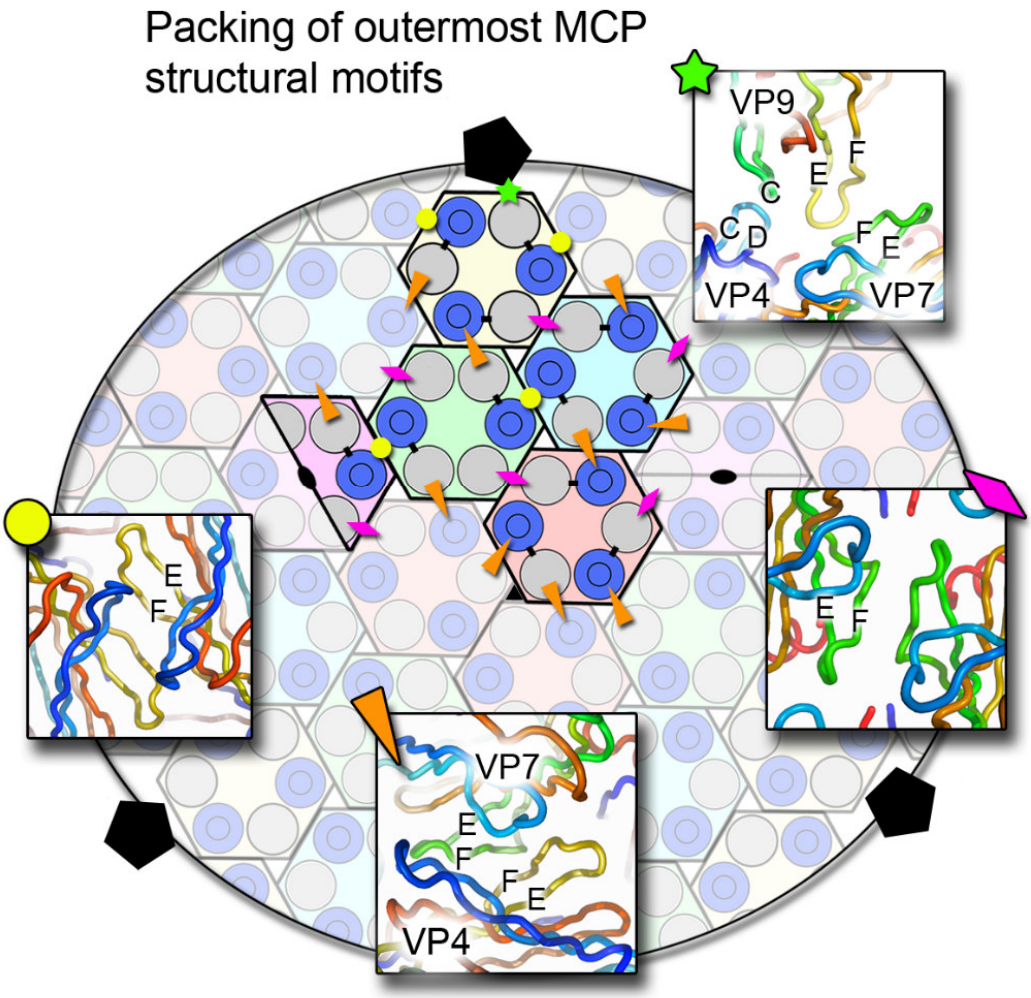


Figure 4.4. Packing of outermost MCP structural motifs. Schematic representation of the IAU in the context of the neighbouring capsomers with packing motifs depicted by a yellow circle: pseudo two-fold related VP4 residues 28-34 (leading to strand C) and 148-166 (in strand F); a magenta rhomboid: pseudo two-fold related VP7 residues 81-87 (connecting strands E-F); an orange triangle: VP7 residues 31-34 with VP4 residues 33-36 and VP7 residues 81-89 with VP4 residues 156-161 (leading to corresponding strands F); the green star marks the location of interaction between protein VP9 and adjacent peripentonal VP7-VP4 heterodimers.

Heterodimer VP7–VP4 as the capsomer building block

The MCPs VP4 and VP7 of HCIV-1 and HHIV-2 are really similar not only in sequence, but also structurally the folding of these proteins is almost identical. The length and amino acid similarities enables this same scheme (VP7 (184/185 residues and 85.9 % aa similarity) and VP4 (232/232 residues and 90.1 % aa similarity) for HCIV-1 and HHIV-2 respectively [71]). Therefore, the structural description of both MCPs has been defined from Cryo-EM-derived density maps, where de novo atomic models were built for both viruses (Figures 4.6 and 4.9). However, to avoid a redundancy of information, the atomic models described here correspond to the map whose resolution density allowed a better fitting of the building blocks, being in this case the map of HCIV-1, at 3.74 Å resolution (0.143 criterion). In the case of HHIV-2, the resolution obtained for the capsid was 3.78 Å resolution (0.143 criterion) (Figure 4.5).

Regarding VP4 protein, with a molecular weight of 26 KDa, it is composed by two vertical single jelly-rolls, one standing on top of the other, each one formed by eight beta strands, which display two four-stranded antiparallel beta sheets that are packed together throughout a hydrophobic interface (the two four-stranded sheets are named as BIDG and CHEF). The upper jelly-roll is smaller than that on the base. There is also a short α_1 helix between the F and G strands (residues 169–175), which is preceding the outer loop of the protein (residues 181-184). In this protein the N-termini and C-termini (GLN4 and LEU232) are spatially really close and complete the anti-parallel BI β-sheet (Figure 4.6 and 4.9).

Concerning VP7 protein, with a molecular weight of 20 KDa, it is composed by just one jelly-roll motif (BIDG and CHEF strands). Out of the β-barrel, is noteworthy the formed loop before the region of the I strand (residues 149–154) and also a short α_1 helix in a similar position to that of VP4. At the C-terminus, when ordered, it forms a short α_2 helix of about two turns (residues 167–174). The N-terminus is the outermost region, linearly stretched out from the jelly-roll, except in the case of the N-termini of the three ‘unpaired’ VP7 subunits in capsomers 4 and 5, which are disordered (Figure 4.6 and 4.9).

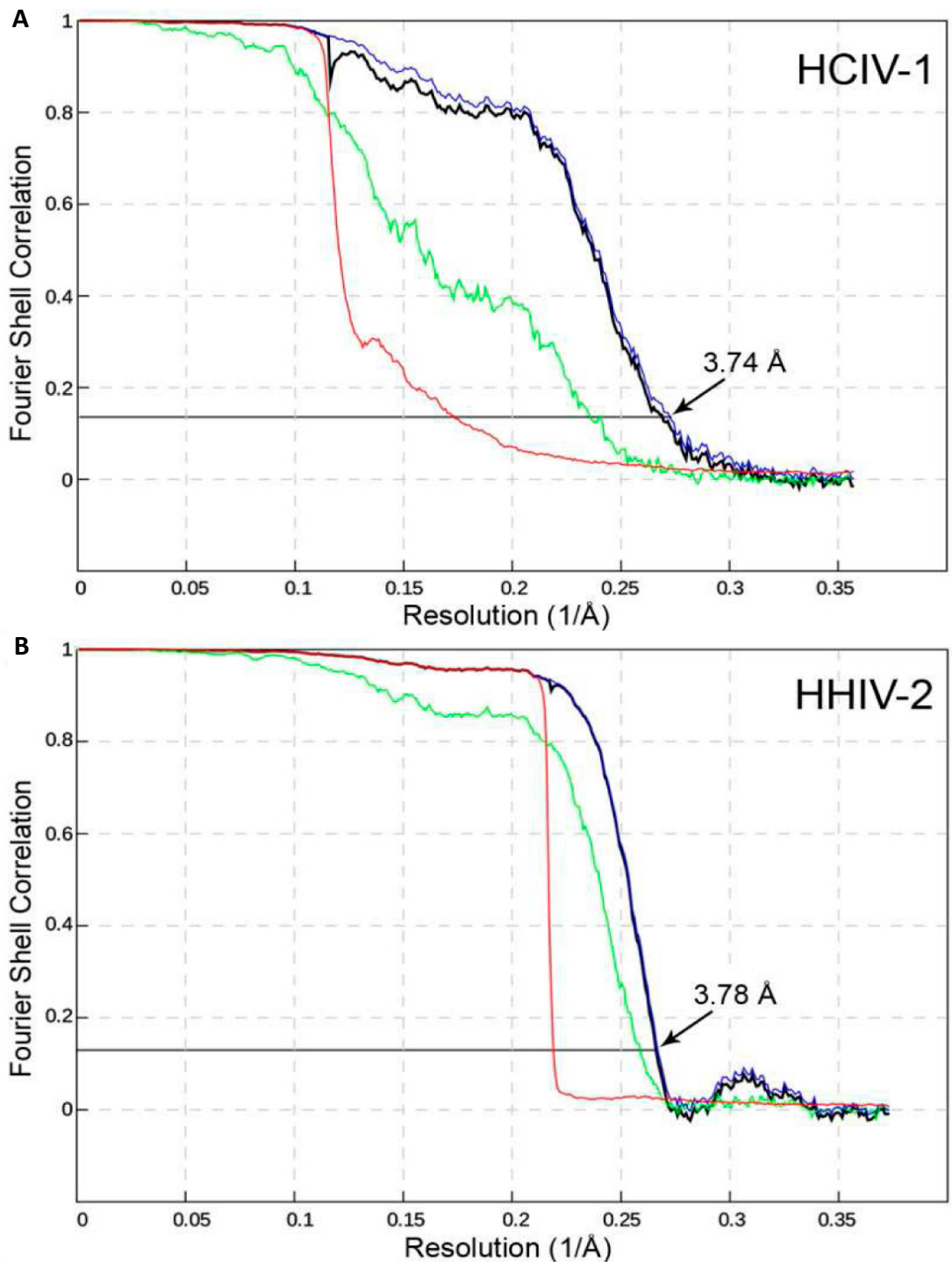


Figure 4.5. Fourier Shell Correlation (FSC) of the reconstructed particles. **(A)** FSC curve of HCIV-1 with reported resolution of 3.74 Å at the 0.143 criterion (black line: corrected map). Other colored curves represent FSC for: blue line, masked maps; red line, phase randomized masked maps; green line, unmasked maps. **(B)** FSC curve of HHIV-2 with reported resolution of 3.78 Å at the 0.143 criterion. Lines colored as in (A).

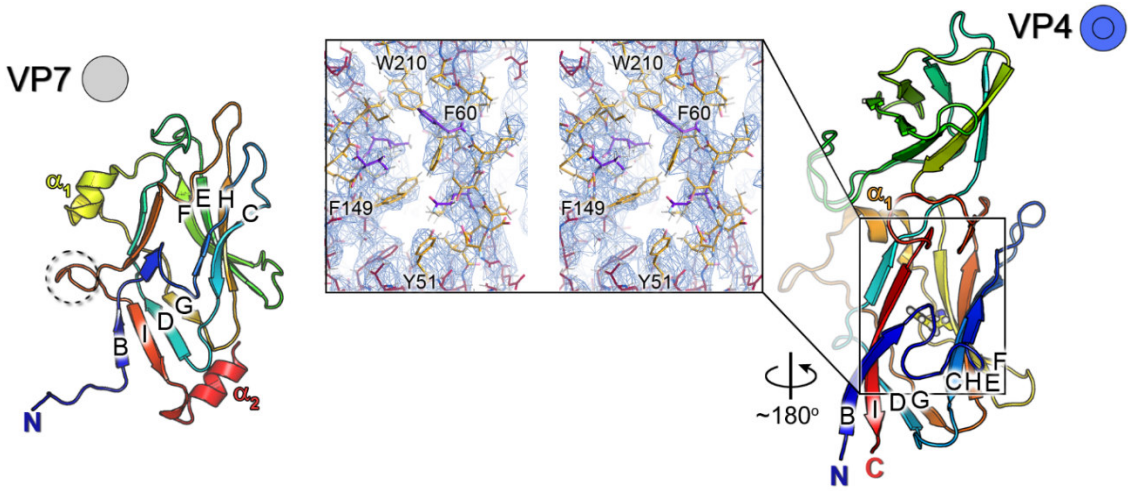


Figure 4.6. Cartoon representation of the HCIV-1 MCPs VP7 (left) with dashed-circle marking the loop with residues 149–154, and VP4 (right) with residue F149 colored in yellow and represented as stick; inset, stereoview of the region marked by a black rectangle with the VP4 atomic model (as stick and differently colored) fitted into the corresponding 3.7 Å resolution density map (blue mesh contoured at ~3σ in COOT [178]) some residues including F149 have been labelled.

Relating to the building block VP4-VP7, the only previous model of a similar union between two single β-barrels MCPs is that of the virus P23-77, where the building block VP16-VP17 displays a similar arrangement [74]. Although the amino acid sequence similarity is quite low between both viruses (VP7/VP16 (184/173 residues and 29.7 % aa similarity) and VP4/VP17 (232/291 residues and 24.2 % aa similarity) for HCIV-1 and P23-77 respectively [71]), both protein pairs (VP7/VP16 and VP4/VP17) share a stunning structural homology (2.1 Å and 3.2 Å rmsd, respectively). Therefore, the atomic models of the HCIV-1 VP7 and VP4 MCPs were manually built taking as baseline templates the core β-barrels derived from the crystal structures of P23-77 MCPs (Figure 4.7).

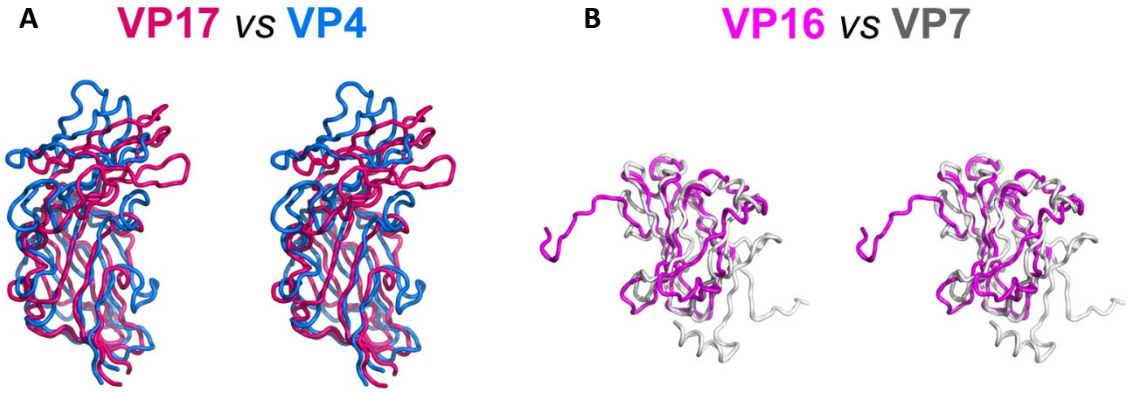


Figure 4.7. Cartoon Superimposition of the MCPs across the single vertical β-barrel proteins. **(A)** Left, stereoview of Thermus phage MCP VP17 (dark-magenta; 232 residues out of 291 modelled) (PDB code:3ZN6; (Rissanen et al. 2013) superimposed onto HCIV-1 MCP VP4 (marine-blue; subunit A in Fig.4.1 centre, 229 residues out of 232 modeled), 3.2 Å rmsd, 194 Cα equivalences. **(B)** Right, as left but with Thermus phage MCP VP16 (magenta; 150 residues out of 173 modeled) (PDB code:3ZN6; (Rissanen et al. 2013) onto HCIV-1 MCP VP7 (light-grey; subunit D in Fig. 4.1 centre, 174 residues out of 184 modeled), 2.1 Å rmsd, 115 Cα equivalences.

This structural homology also resembles that of bacteriophage PM2, which is the prototype of the PRD1-adeno viral lineage [28]. Although PM2 presents a single MCP with a double β-barrel motif, the relative angular orientation of the individual β-barrels of the heterodimer VP7 – VP4 fairly fit with the homodimer (4.2 Å rmsd) (Figure 4.8).

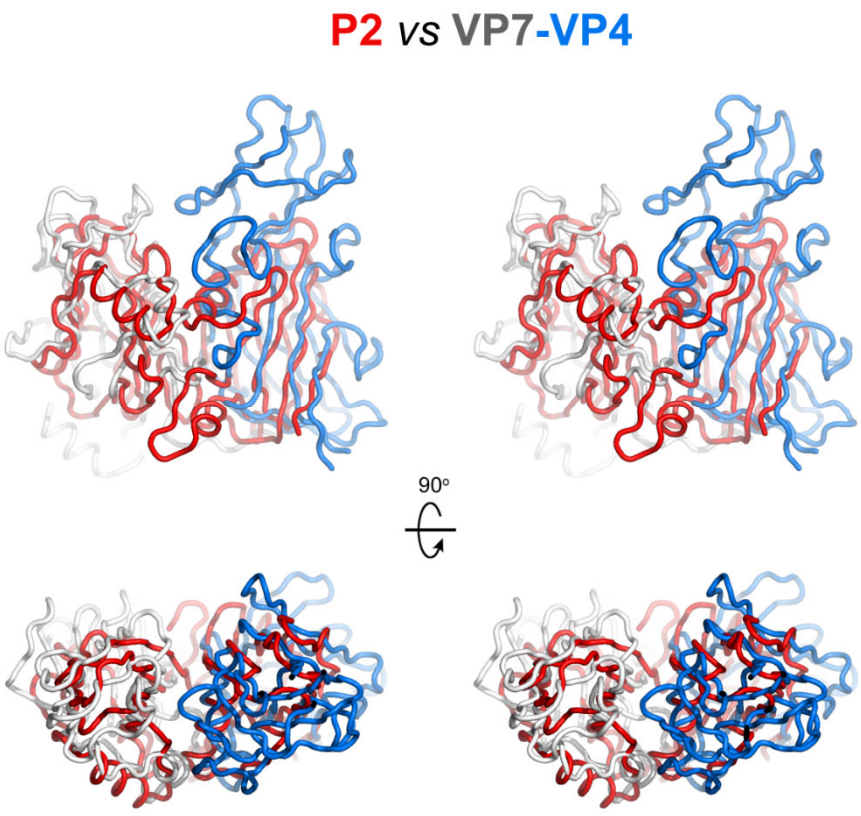


Figure 4.8. Stereoviews of HCIV-1 MCP VP7-VP4 heterodimer (403 residues) onto the double β-barrel MCP P2 of bacteriophage PM2 (red; 269 residues) (PDB code: 2VVF: (Abrescia et al. 2008), 4.2 Å rmsd with 189 Cα equivalences; the relative angular orientation between the individual HCIV-1 VP7-VP4 β-barrels practically replicates that of double β-barrel MCP. (Up) View of the heterodimer parallel to the membrane. (Down) Top view of the heterodimer.

The main unions between VP7 and VP4 that make possible the dimer stability are mainly two (Figure 4.9):

1) First, at the N-terminal of protein VP7, an extension of 4 residues is linearly stretched under the β-barrel of the neighbouring VP4 placed clockwise. As a result, VP7 ILE4 and LEU7 amino acids get buried, while ASN3 and GLY5 amino acids interact with VP4 THR191 and SER189 residues via carbon–oxygen hydrogen bonding, respectively. In addition, this union is reinforced by the VP7 − VP4 protein interfaces, since the interface area of VP7 with VP4 clockwise subunit buries ~ 1340 Å2 (n = 12), while the buried area related with the anticlockwise VP4 is much smaller (~ 760 Å2).

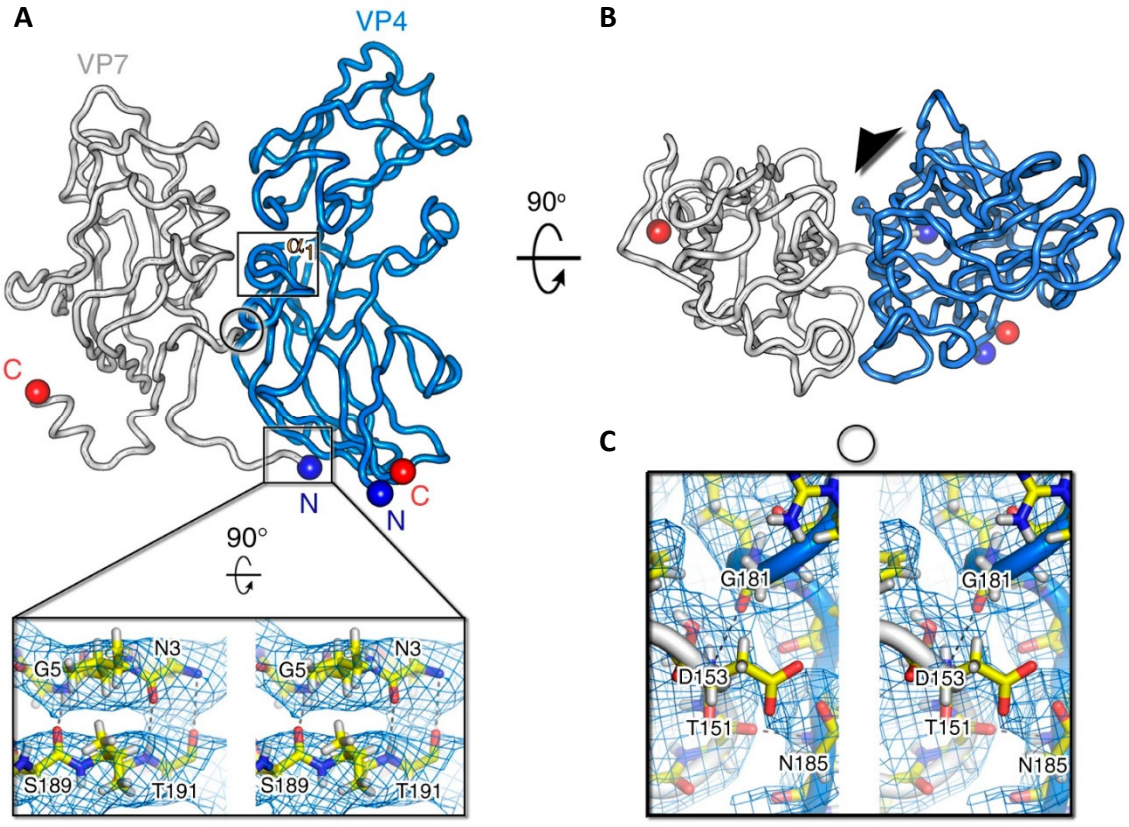


Figure 4.9. VP4–VP7 heterodimer. **(A)** Heterodimeric VP7–VP4 building block (VP7 in light gray; VP4 marine-blue) with interacting VP7 N-terminal residues. The black-circle marks the VP7 loop, residues 149–154, burying underneath the region underlying the VP4 α1 helix marked by a black rectangle, and contacting the VP4 loop, residues 181-184; blue and red spheres label the corresponding N-terminal and C-terminal. Inset, wall-eye stereoview of the main-chain hydrogen bond interactions (gray dashed lines) between the N-terminal residues of VP7 and residues of VP4 with corresponding density as blue mesh. **(B)** VP4–VP7 heterodimer top view (rotated 90° from (A)). **(C)** Wall-eye stereoview inset, suitably oriented, showing density as blue mesh and details of main-chain hydrogen bond interactions (black dashed lines) between the residues marked by the black circle in the panel (A). Figure was generated in Pymol (https://pymol.org/2/).

2) Second, the protruding loops of VP7 (residues 149–154) and VP4 (residues 181-184) are spatially faced, in close proximity. Thence, VP7 THR151 and ASP153 residues link to VP4 ASN185 and GLY 181 via hydrogen bond interactions, respectively. This interaction is favoured by the anchoring of VP7 loop underneath the VP4 α_1 helix (residues 169–175) originating a stable buried site (~60% of the loop).

In the case of the unpaired VP7 subunits of capsomers 4 and 5, the individual MCPs show the same core, but this time the N-terminal is faced against other VP7 and is disordered, losing the hydrogen bonds present in the heterodimers. The VP7 loop is present, but the closest area of the clockwise VP7 is the chain between α1 helix and G strand, which does not involve any significant interaction.

In the case of HCIV-1, heterodimers do not form disulphide bond linkages, since only VP7 contains a single cysteine. Neither does HHIV-2 heterodimer, despite containing a cysteine amino acid in each MCP, since they do not coincide in the same area of interaction.

4.3 Capsomers formation mechanism in the capsid

The icosahedrally averaged Cryo-EM maps of HCIV-1 and HHIV-2 do not only show a strong density at the place of the MCPs, but they also exhibit a further density underneath the pseudo-hexamers corresponding to additional proteins, which are in contact with the base of the capsomers (Figure 4.10, A). Thus, taking HCIV-1 map as reference, d*e novo* poly-ALA modeling was fitted into the related densities (Figures 4.11 and 4.12).

As the primary sequence of these densities is unknown and there is not enough resolution to establish unequivocally the connectivity of the Cα backbone, it was impossible to identify the protein structurally. Although the recognition of the side chains was unattainable, the secondary structural elements of the proteins are discernible at the proposed densities, identifying two different types of proteins (Figures 4.10 and 4.12). Curiously, the proteins beneath the three-tower capsomers display the same α+β-fold, while the densities below the two-tower capsomers exhibit a bundle of five helices. Both proteins are at 10–15 Å distance from the outer membrane leaflet (Figure 4.11). The presence of these proteins suggest that they may be directly involved in the capsomers formation, directing the positioning of the MCPs. For this reason, they have been named as Global-Positioning-System proteins GPS-III and GPS-II, when they are placed beneath the three-tower capsomers and two-tower capsomers, respectively.

A

B

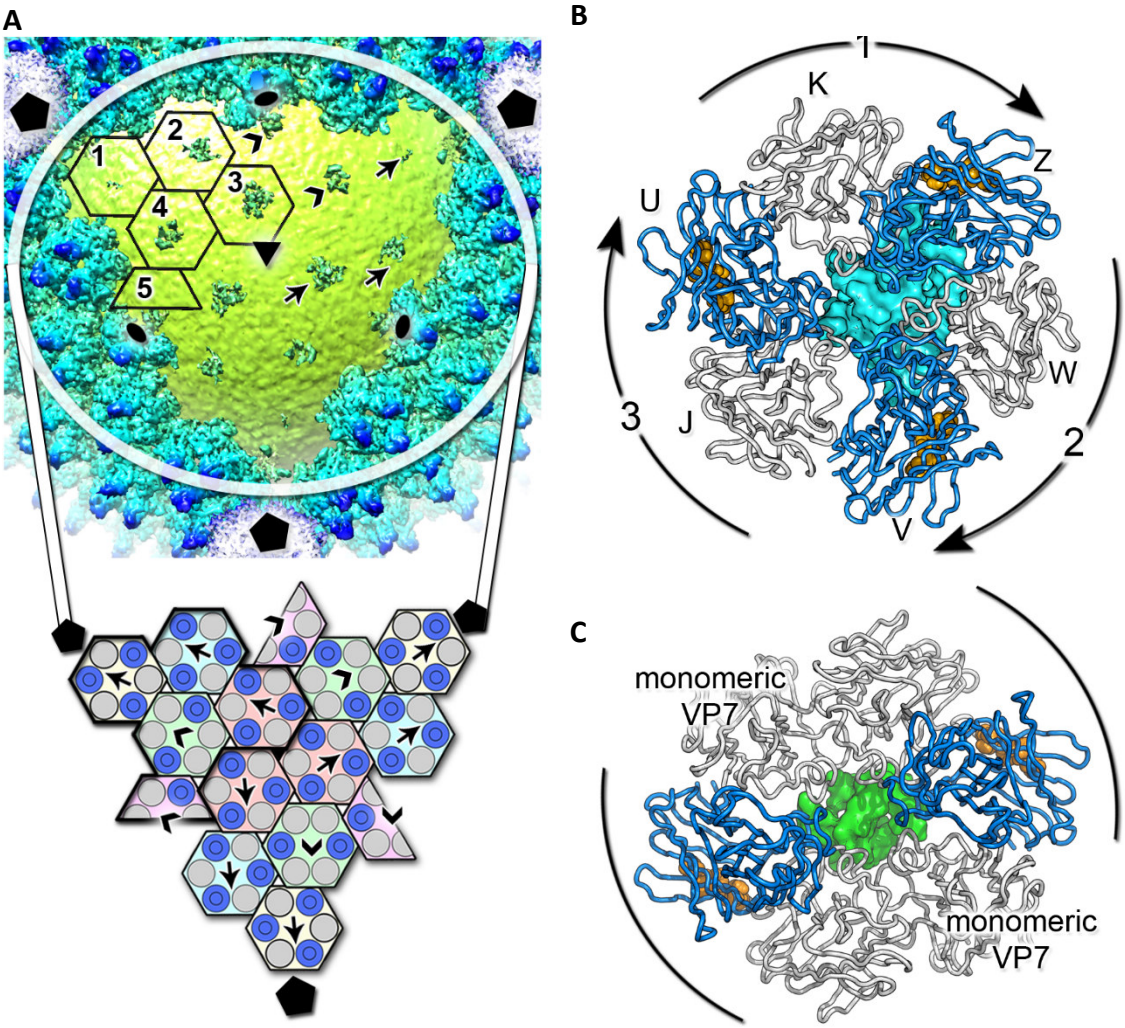


Figure 4.10. GPS proteins beneath the capsid shell. **(A)** Top, cut-through density of a virus facet viewed along the icosahedral three-fold axis; arrows and arrowheads indicate the density (shades of blue) beneath the three-tower (Nos. 1–3, see Fig. 4.1 center) and two-tower (Nos. 4 and 5) capsomers (black hexagons), respectively. The outer leaflet (OL) of the membrane is in lime-yellow. Below, schematic of the capsomers composing the facet with one IAU outlined by a thicker black line (represented as in Fig. 4.1). Pentagons, triangles and ovals mark the icosahedral symmetry five-fold, three-fold, and two-fold axes. **(B)** Top view along the pseudo-three fold axis of MCPs composing the three-tower capsomers as cartoon tube (colored as Fig. 4.7 with the off-centered GPS-III protein represented in cyan surface; labels identify the VP7–VP4 subunits and curved black arrow with numbers the putative order of docking/registering of the VP7–VP4 subunits onto the GPS-III. **(C)** As (B) but for the two-tower capsomers where the GPS-II (green surface) is centered with the α_3 spanning the central cavity of the two-tower capsomers stapling together the opposite VP7–VP4 heterodimers (curved black lines) and leaving space for the docking of monomeric VP7.

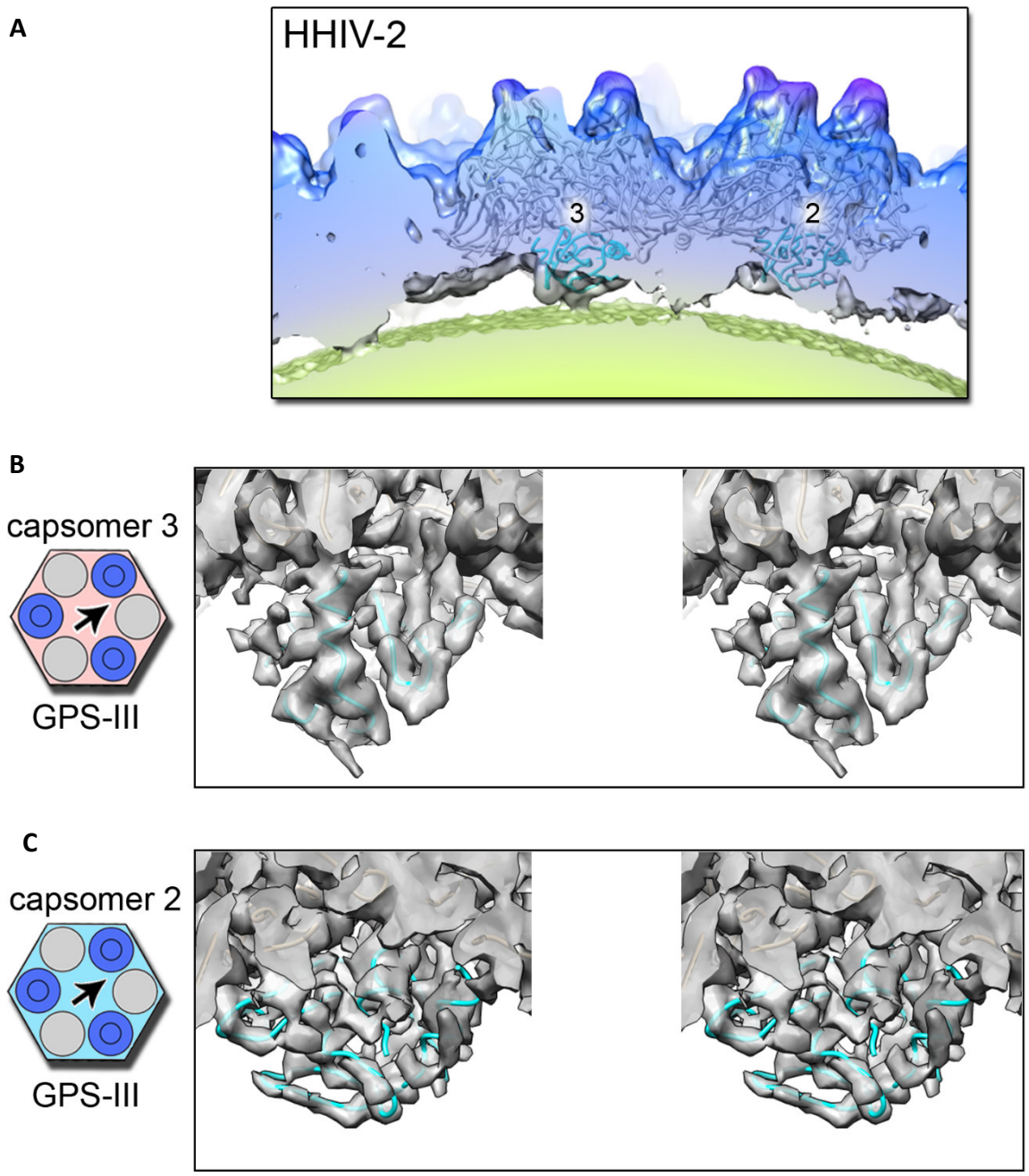


Figure 4.11. HHIV-2 GPS-III proteins. **(A)** Cut-through of the filtered cryo-EM map of HHIV-2 color-coded as Fig. 4.1 (right) with cartoon tube of the IAU model with numbers marking capsomers Nos. 2 and 3 as Fig. 4.1 (center) showing their proximity to the membrane outer leaflet (OL). **(B)** and **(C)** Wall-eyed stereoview of HCIV-1 GPS-III poly-ALA (cyan) model fitted into the HHIV-2 density (mesh grey) underneath capsomers Nos. 3 and 2 respectively, differently oriented (the density underneath capsomer 2 is weaker than that underneath capsomer 3). The density protruding from the helix in the center panel identifies the presence of large side-chains, also clearly visible in the corresponding HCIV-1 density region.

GPS-III protein

The density corresponding to GPS-III is stronger and more interpretable in capsomer 3, near the icosahedral three-fold axis, than in the capsomers 1 and 2 (Figure 4.10, A). This is probably due to the greater stability that the membrane confers near 3-fold, where the GPS-III is closer to the outer lipid layer. Towards the 5-fold, the membrane curvature increases, gradually distancing the membrane from the capsid. This provokes a weaker interaction of the GPS-III with membrane, especially under capsomer 1, where the protein is really flexible. Notwithstanding that, Poly-ALA models were fitted into the densities of the GPS-III under capsomers.

The protein backbone of GPS-III models created in HICV-1 map were also fitted into the corresponding HHIV-2 map densities (Figure 4.11). The good quality of the fitting comes out that GPS-III proteins of HHIV-2 are structurally homologous to those of HCIV-1. This is particularly evident for the GPS-III beneath the capsomers 2 and 3.

Regarding the orientation of the GPS-III respect to the upper capsomer, its position is off-centered from the pseudo-three-fold axis of the three-tower capsomer (Figure 4.10, B). This asymmetry provides the order for the register of the VP7–VP4 heterodimers onto the GPS-III during capsomer assembly, as each of the three heterodimers display a decreasing contact area (K–Z: ~350 $Å^2$; W–V: ~310 $Å^2$; J–U: ~120 $Å^2$, respectively).

GPS-II protein

The density corresponding to GPS-II proteins clearly displays a pseudo-two-fold symmetry in HCIV-1 map, as also do the capsomer 4 and 5. Hence, Poly-ALA models were fitted into the densities under capsomer 4 and half of the capsomer 5, as the latter one is sitted on the icosahedral 2-fold axis (Figure 4.12). As mentioned before, 5 α-helices were identified (2 and ½ in the case of the GPS-II under capsomer 5, due to the splitting of the protein). These models were not placed into the related densities of HHIV-2 volume, as the map signal was scarce for the fitting.

Concerning the orientation of the GPS-II, it is centered respect to the axis of the upper capsomer, with the α3 helix occupying the central cavity of the two-tower capsomers. In this case VP7-VP4 heterodimers join opposite sides of GPS-II, leaving a proper space for the accommodation of monomeric VP7 subunits (Figure 4.10, C).

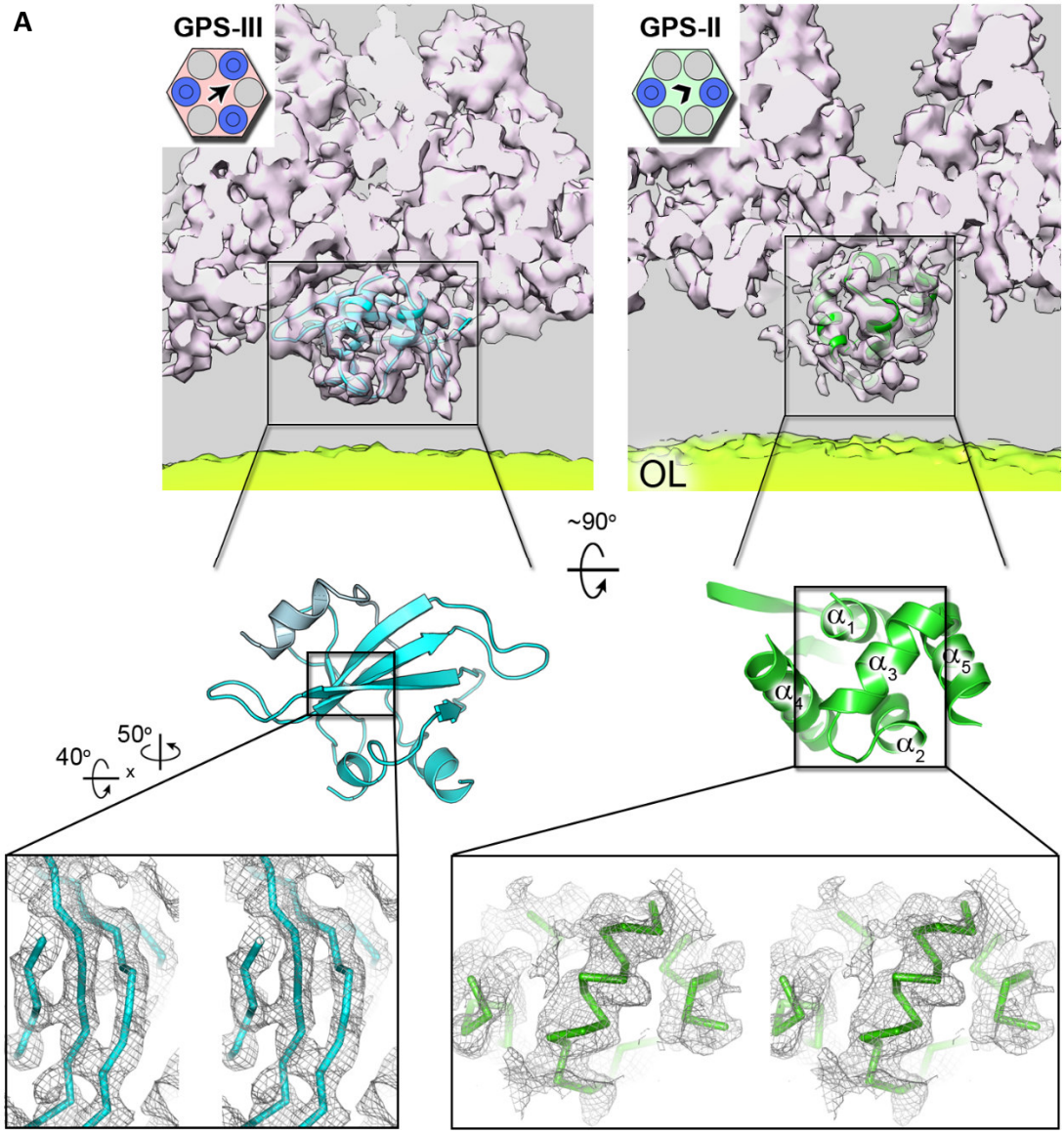


Figure 4.12. Poly-ALA models of GPS proteins fitted into the density map of the virus. **(A)** (Left) side-view of the Gaussian filtered (1.4 Å width in Chimera [29]) electron density (white 40% transparency) corresponding to the three-tower capsomer No. 3 (As Fig. 4.1 center) close to the icosahedral three-fold axis, with further density beneath marked by a black rectangle with the GPS-III poly-ALA model fitted in (cyan cartoon) and in lime-yellow the blurred membrane (OL); the same atomic model at the center shows an orthogonal view of the GPS-III protein with an inset of a stereoview of the density (gray mesh contoured at ~2σ in Pymol) corresponding to the resolved strands (black rectangle). **(**Right) The same as left but corresponding to the density of the two-tower capsomer No. 4, with the additional density at its center as marked by the black rectangle corresponding to the five-helix bundle GPS-II protein (green cartoon).

4.4 Vertex complex organization

The vertex complex organization in HCIV-1 and HHIV-2 involves the coordination of various groups of proteins, structurally different, that converge at the icosahedral 5-fold axes. At the apex of the 5-fold, completing the capsid, the density of both maps show a homopentamer formed by five penton proteins. Each of the proteins belongs to a different IAU (Figure 4.2), to which is connected by the adjacent peripentonal VP7-VP4 heterodimers of Capsomer 1 (Figura 4.4).

These penton proteins display the same single β-barrel observed in the MCPs (Figure 4.13). About eight residues preceding this motif clamp the base of the adjacent VP9 clockwise, followed by an N-terminal α-helix ~30º tilted relative to the membrane surface (Figure 4.13 and 4.14, A). Previous biochemical studies suggested that the identity of this protein coincided with the virus protein 9 (VP9) [187] [71] in *Alphasphaerolipoviridae* family. In consequence, the Cα backbone was manually built into the map density and VP9 was successfully fitted in-register in both viruses. The amino acid registration was possible due to the good quality of maps, as many side chains of the protein were identified (Figures 4.13, B and 4.20)

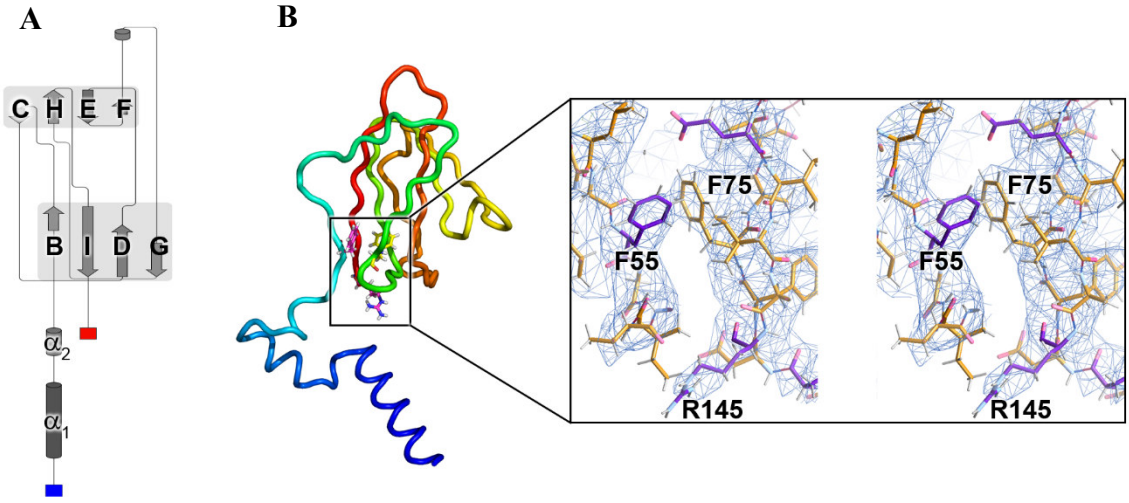


Figure 4.13. VP9 protein folding and structure. **(A)** Schematic of VP9 penton protein, displaying the 8 strands that form the β-barrel and two α-helices at the N-terminus. Red and blue rectangles indicate the C-terminus and N-terminus respectively. **(B)** Cartoon-tube representation of VP9 (rainbow coloring: N-terminus, blue; C-terminus, red) with inset showing a stereoview of the corresponding electron density region, marked by the black rectangle with in stick the residues F55, F75, Y55, and R146.

In contrast to the VP4 and VP7 MCPs, the VP9 proteins of HCIV-1 and HHIV-2 share a low amino acid sequence similarity (33.2%), despite conserving the same structural homology in the form of a vertical jelly-roll. In the case of HCIV-1, the protein length is shorter than in HHIV-2 (146/174 respectively). In addition, the secondary structure in α-helix of the N-Terminal is more evident in HCIV-1 (~34 visible residues) than in HHIV-2 (just the beginning is hardly discernible).

Following the 5-fold inwards, the five N-termini of VP9 end at the internal membrane bilayer, where they interact with a membrane protein complex. This complex occupies the entire membrane thickness and is compatible with five transmembrane helices forming a bundle (Figure 4.14), resembling the structure of A55 pentameric arrangement in *Sulfolobus* virus STIV, with a double β-barrel MCP [188].

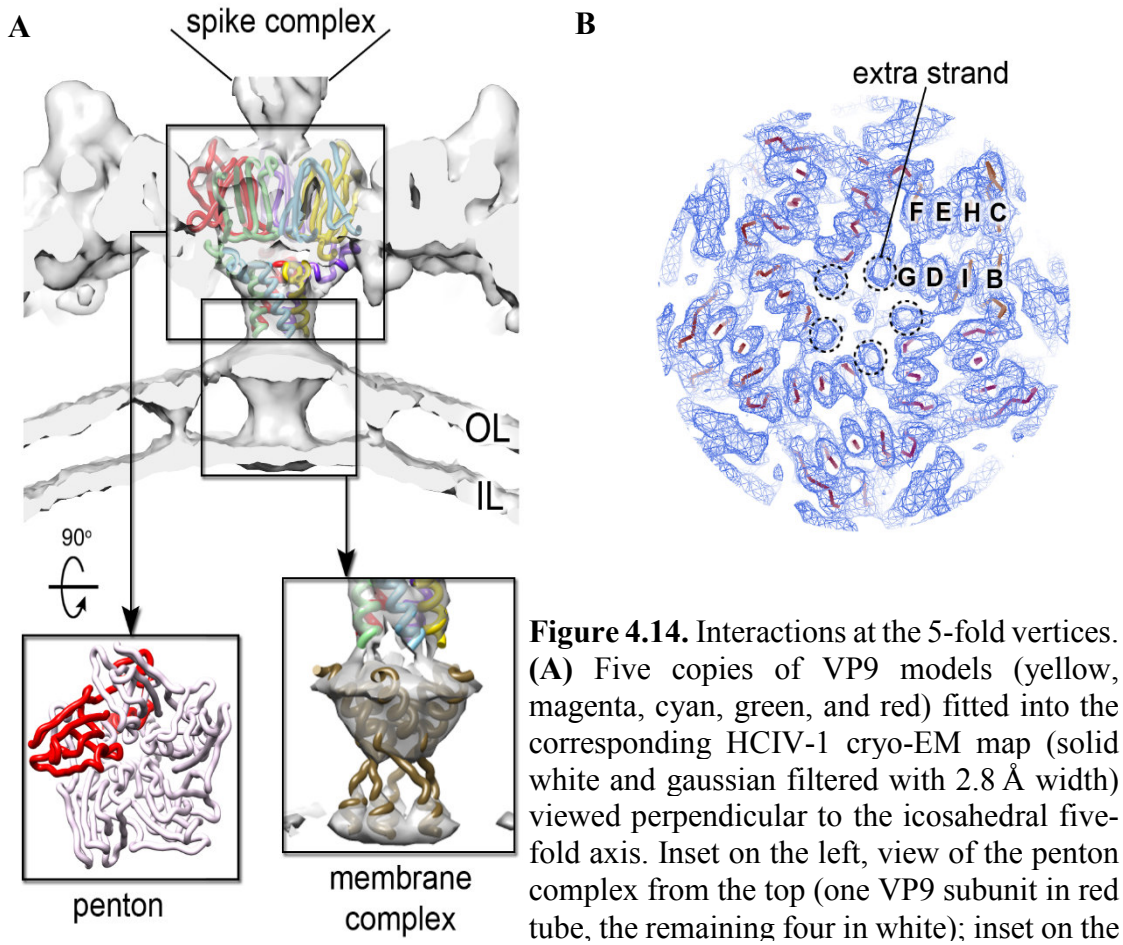


Figure 4.14. Interactions at the 5-fold vertices. **(A)** Five copies of VP9 models (yellow, magenta, cyan, green, and red) fitted into the corresponding HCIV-1 cryo-EM map (solid white and gaussian filtered with 2.8 Å width) viewed perpendicular to the icosahedral five-fold axis. Inset on the left, view of the penton complex from the top (one VP9 subunit in red tube, the remaining four in white); inset on the right, the density region corresponding to the membrane protein plugging the membrane vesicle five-fold apices (contoured at higher threshold) and compatible with five copies of the C-terminal transmembrane helix of VP13 (brown tube). **(B)** View of the cryo-EM map as blue mesh (contoured as in Fig. 4.11 inset) viewed along the five-fold symmetry axis with fitted the Cα model of the penton in red and with one VP9 jelly-roll labeled as in Fig. 4.11; the dashed black circles mark the extra β-strands deriving from the above spike complex and that interact with the VP9 G strands.

Finding the identity of the membrane complex proteins was not possible by fitting any poly-ALA model inside, as the resolution was low (outer membrane layer at 9.5 Å resolution (0.143 criterion) (Figure 4.15). Therefore, three different approaches were performed to infer which protein of the viral genome was suitable.

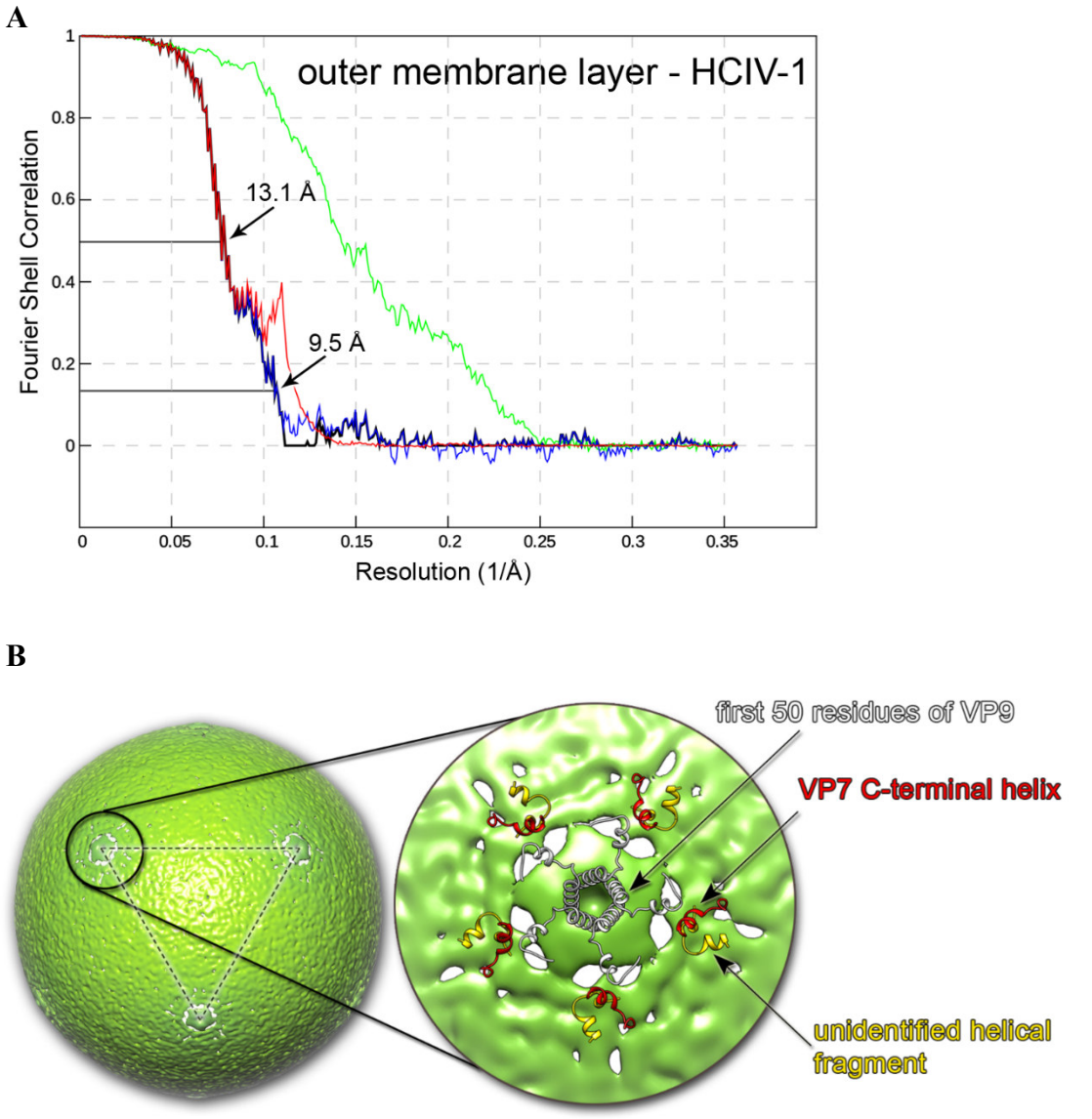


Figure 4.15. Outer membrane layer of HCIV-1. **(A)** Fourier Shell Correlation **(**FSC) curves (black, blue and red lines) correspond to the outer leaflet of the membrane vesicle of HCIV-1 with reported resolution of ~13 Å at the 0.5 criterion; green line FSC corresponds to the unmasked maps and therefore the whole virion. **(B)** Surface rendering (light-green) of the outer leaflet (OL) of the membrane vesicle (Cryo-EM map filtered to 9.5 Å). Inset, enlarged area of the density of the OL beneath the penton viewed along the five-fold axis (inner leaflet and genome removed for clarity) with some ordered structural elements populating the region around the five-fold and in between the capsid shell and the OL.

First, HCIV-1 purified sample was filtered (1000 kDa MWCO) and the flow-through was collected for BN-PAGE and tryptic digestion. Then, the three main emerging bands were analysed by nano-LC–MS/MS, resulting in soluble protein subassemblies in which a cluster of proteins composing the vertex were present. Among the proteins appearing in this cluster, apart from the expected ones, also VP13 protein came out with a similar abundance to the rest of the vertex complex proteins.

Second, a secondary structure prediction was carried out to recognise which HCIV-1 structural proteins likely contain transmembrane helices. In regard to VP13, its protein sequence predicted a transmembrane helix at its C-terminus (residues 60-77).

Third, an exhaustive search was accomplished to find any homology, sequential or structural, in other species. If we compare the amino acid sequence similarity and identity between VP13 of HCIV-1 and HHIV-2 the homology is still quite high (81/73 residues, 42.9 % aa similarity and 54.8 % aa identity, respectively). In HHIV-2 the C-terminal also predicts a transmembrane helix (residues 49-65). Regarding the similarity of VP13 to the equivalent protein in STIV, A55, HICV-1 is closer to that of STIV-1 virus (81/55 residues, 17.6 % aa identity and 29.8 % aa similarity), while HHIV-2 shares a higher homology with STIV-2 virus (73/55 residues, 23.4 % aa similarity and 31.2 % aa identity) (Figure 4.16).

Out from the viruses, a hypothetical protein predicted for *Haloarcula* sp (NCBI Reference Sequence: WP_121577218.1), which is the host genus of both, HCIV-1 and HHIV-2, maintains a significant level of similiraty with VP13 viral proteins, especially with that of HHIV-2 (73/63 residues, 37.2 % aa similarity and 46.2 % aa identity) for HHIV-2 and *Haloarcula* protein, respectively). This archaeal hypothetical protein also displays a transmembrane helix at the C-terminal (residues 44-61), according to the predictions.

From the structural point of view, the search of single-pass transmembrane proteins with transmembrane helix–helix (TMH) interactions forming left-handed pentameric coiled-coil complexes was brought off (Figure 4.16). The best characterized structure which fulfils this requirements is the phospholamban protein (PLN), a calcium regulator of the sarcoplasmic reticulum well conserved in mammals [189 190].

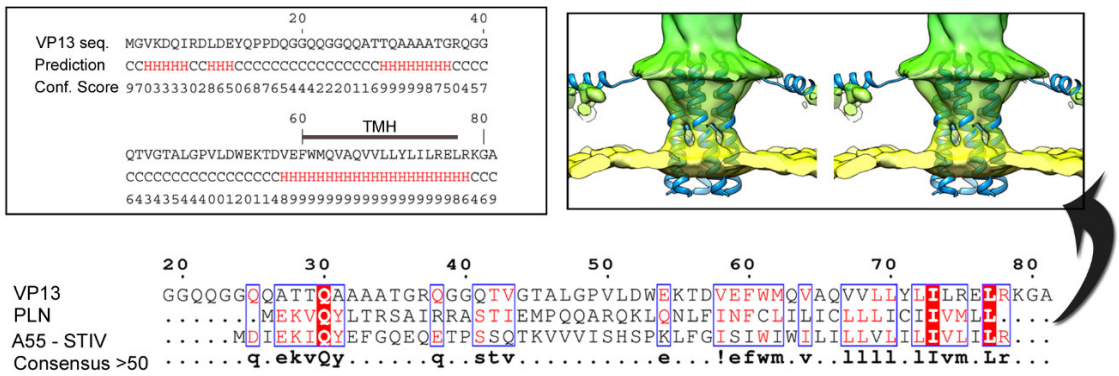


Figure 4.16. Secondary structure prediction of VP13. (Top left) Prediction using I-TASSER [182], with marked the transmembrane helix segment (TMH) as from phobious webserver (http://phobius.sbc.su.se/); coil as C, helix as H, transmembrane helix as TMH, and the confidence score from 1 to 9, with 9 being the highest. Bottom, multiple sequence alignment between VP13, human phospholamban (PLN) and STIV A55 with red box, white characters as strict identities and with white box, red characters as similarities and with the consensus sequence > 50% marked (Espript software8). Top right, stereoview of cryo-EM density (gaussian filtered with a 2.1 Å width in Chimera) corresponding to the membrane complex below the VP9 penton with fitted in as rigid-body with the pentameric structure of PLN (PDB code:2M3B (Vostrikov et al. 2013 [190])

Interestingly, the corresponding transmembrane regions of HCIV-1 VP13, STIV-2 A55 and human phospholamban showed a high consensus sequence (> 50%) (Figure 4.16). The sequence similarity of VP13 with these proteins was much higher than those found for the other transmembrane helices in VP10 and VP12, reinforcing VP13 as the protein forming the pentameric bundle at the membrane. The good fitting of the phospholamban pentamer PDB into the viral membrane density also supports this idea (Figure 4.16, up left).

Finally, the pentons at the 5-fold have to work as adaptors for the viral receptor proteins, also called spike complex or simply spike. These spikes are placed at the 5-fold vertex, externally to the capsid. In the case of HCIV-1, the spike displays a dimeric shape, causing a symmetry mismatch when linking to the capsid penton. Although it is unclear what is the exact binding mechanism between the spike and the penton, the map density shows 5 extra strands penetrating in the centre of the pentamer (Figure 4.14, B). Every strand is in close contact with the adjacent VP9 β-barrel motif, precisely interacting with strand G. Thus, this association could work as an anchor of the spike to the vertex. In HHIV-2 the spike complex preserves the same symmetry of the 5-fold vertex. In this instance, a central fiber coordinates the binding of the spike to the penton, assisted by accessory proteins, as it will be explained further on (Figure 4.21).

4.5 Structure of the receptor binding proteins

The adsorption and binding of viruses to cell surface receptors is mediated by spike complexes, which in icosahedral viruses are usually located at the 5-fold vertices. The specific adaptation of these complexes to their binding to the cell surface often causes a symmetry mismatch with the icosahedral symmetry of the capsid. Therefore, it is necessary a proper binding mechanism between the spike complex and the 5-fold vertex.

To solve the spike complex of both viruses, in both cases localized reconstruction method was implemented, in which the 12 vertex complexes of every particle were extracted to be aligned between them (see methods) [185]. In the case of HICV-1, the receptor complex shows a C2 symmetry horn-like shape (Fig. 4.17-18), which is very similar to that already described for SH1 virus [191]. The flexibility of this structure was verified by successive 2D classifications, in which different aperture degrees of the horn were observed. Both horns are anchored at their base to the 5-fold, while the ends can show a closed conformation, in which the tips of the horns touch each other, or an open conformation, where the ends are distant (Figure 3.2, B). Among the different morphologies appreciated, a spike reconstruction was performed in an open and slightly more closed conformations, at 8.9 Å and 18.2 Å respectively (FSC 0.143 CUT-OFF) (Figure 4.17).

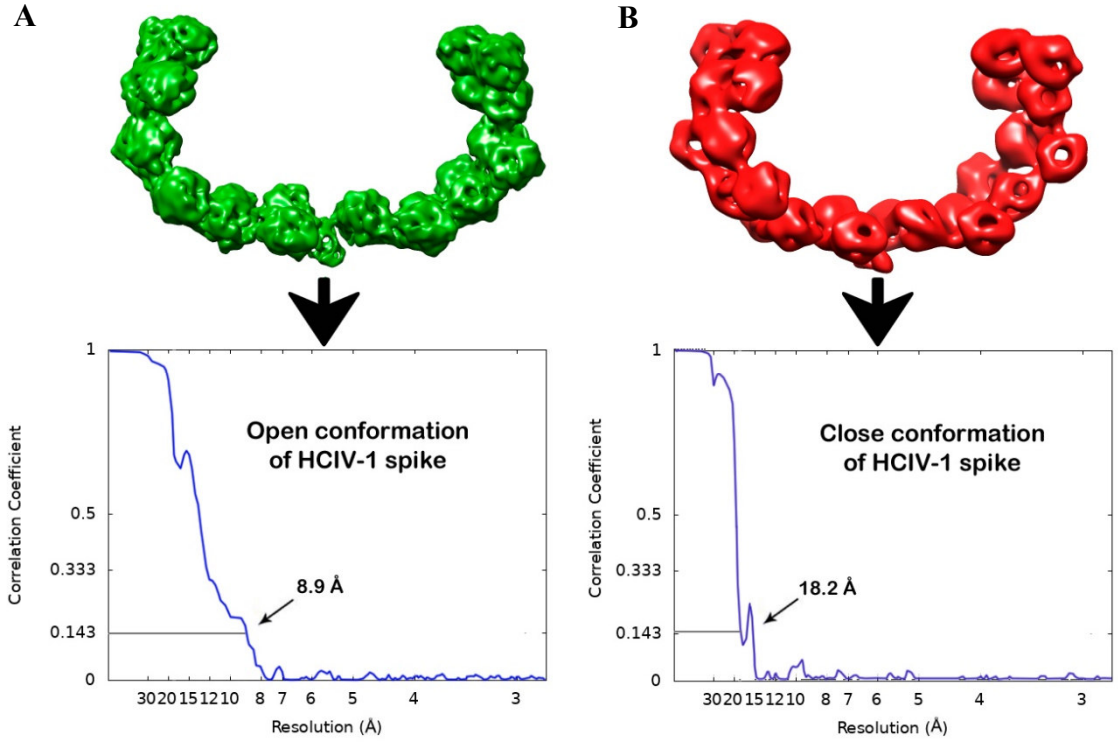


Figure 4.17. Flexibility of the HCIV-1 Spike. **(A)** Map of the 3D reconstructed spike of HCIV-1 in open conformation (colored in green). (Below) Masked-corrected Fourier Shell Correlation (FSC) with reported resolution of 8.9 Å at the 0.143 criterion. **(B)** as (A), but displaying a closer conformation of the spike (colored in red), reporting a resolution of 18.2 Å at the 0.143 criterion.

In both cases the resolution of the maps was not good enough to build de novo atomic models. Thus, to identify the structural proteins forming the spike, a combination of structural, proteomics, and secondary structure prediction was performed (Figure 4.19). The predicted model for the structural protein VP2 showed an elongated and disordered structure. This structure, taking into account the length of this protein (593 aa), can

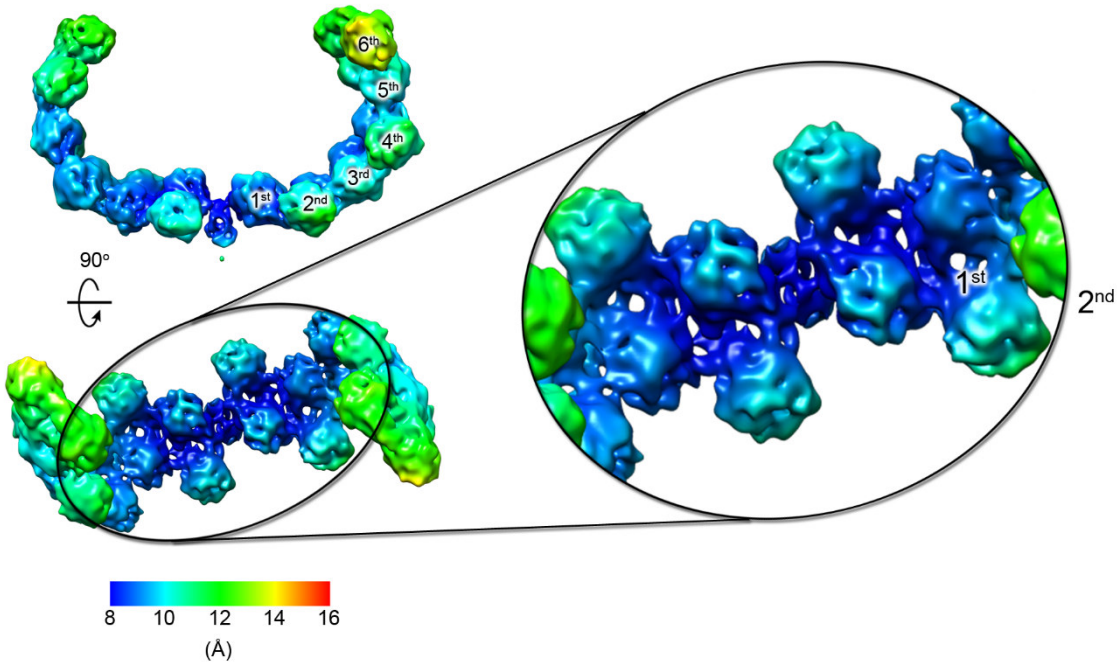


Figure 4.18. Local resolution of HCIV-1 spike (open conformation). View of the 3D reconstructed spike by localized reconstruction, orthogonal to the two-fold axis, showing its architecture as a pearl collar necklace composed by six layers (labelled from 1st to 6th). (Bottom) view of the dimeric spike along the two-fold axis displaying the arrangement of globular protein domains as shown in the oval inset (right); the density has been colored accordingly to the local-resolution estimation (see the key below).

perfectly span the dimension of a horn. The same secondary structure prediction was performed for structural proteins VP3 and VP6 (326 and 258 residues, respectively), getting in both cases globular domains as the most likely conformation (Figure 4.19). Therefore, the proposed model for the HCIV-1 spike is composed of two copies of VP2, which act as the backbone scaffolding protein of the structure, being the central axis to which the VP3 and VP6 globular proteins attach. These globular domains are placed in parallel pairs, on both sides of the backbone. Between each pair there is a lateral displacement along the backbone, creating a zig-zag-shaped structure that allows a gear-like fitting, much more stable and compact (Figures 4.18 and 4.19).

The structural prediction of these 3 proteins at the spike vertex is also supported by the amino acid similarity shared with SH1, where VP2, VP3 and VP6 were known to be located at the virion vertices [70]. Finally, in the bands analysed by nano-LC–MS/MS, VP3, VP6 and VP2 appeared in the same cluster of proteins composing the vertex complex, together with VP9 protein.

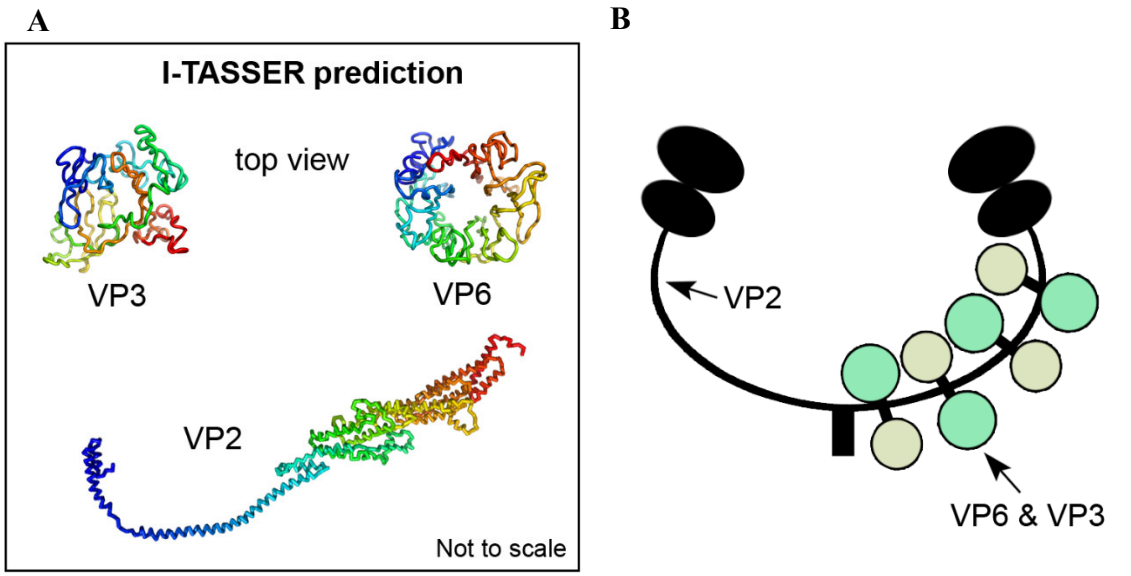


Figure 4.19. Secondary structure prediction of HCIV-1 spike proteins. **(A)** Possible model from I-TASSER server for the spike proteins. VP2 prediction showed an elongated structure (confidence of structure prediction C-score = -1.27; typically from -5 to 2, higher the better) using as template structure PDB: 5YFP (Mei et al. 2018 [203], chain F). VP2 shares only 6.4% sequence identity in the structurally aligned region. VP3 prediction showed a globular structure (PDB: 6FHJ (Moroz et al. 2018) (C-score = -4.30 (low)), although sequence identity was higher than in VP2 (19% sequence identity). VP6 prediction also showed a globular structure with similar score that in VP3. **(B)** Proposed schematic architecture of the HCIV-1 spike complex with VP2 forming the dimeric scaffolding horns with globular VP3 and VP6 arranging along its length.

In the case of HHIV-2, the spike vertex morphology notably differs from the spikes observed in the rest of alphasphaerolipoviruses. The obtained map shows a propeller-like shape, surrounding a long vertical fiber at the center (Figure 4.21). The resolution of the spike gradually decreases following the 5-fold axis outwards (the mean resolution of the map is 7.7 Å at FSC 0.143 CUT-OFF, although local resolution ranges from 4 Å at the base of the spike to 12 Å at the tip of the fiber) (Figure 4.21). Consequently, poly-ALA model was manually built only in some of the densities where the Cα chain was recognizable (Figure 4.20). As a result, 5 multi-domain proteins (four jelly-rolls arranged end-to end) were identified at the propeller-like densities, crowning to the below VP9 penton. The innermost jelly-roll of the multi-domain sits circa orthogonally on top of the

VP9 subunits and inserts one of the terminal ends in the inner region of the penton which is coaxially glued together to the central, highly flexible fiber by an unidentified polypeptide stretch (~24 residues). The outermost domain shows a weak density in which only part of the jelly-roll has been fitted.

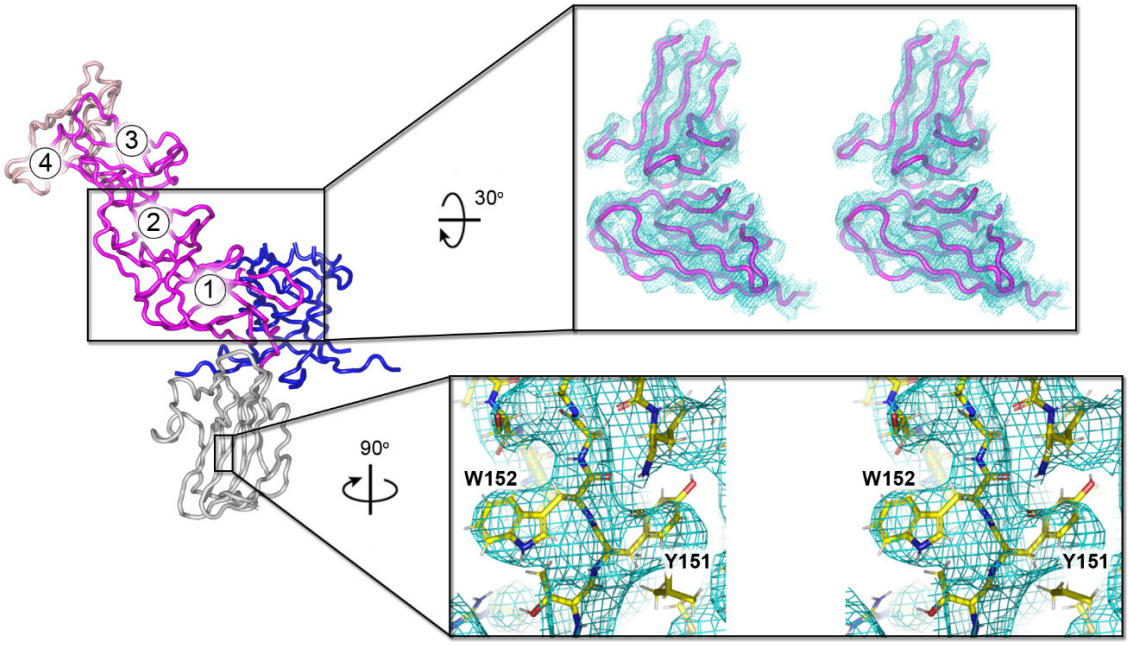


Figure 4.20. HHIV-2 spike linked to VP9 penton protein. Multidomain composition of VP16 (tube magenta; numbers 1-3 mark the jelly-roll domains) and its arrangement relative to a possible VP17 domain (tube pink and labelled as 4), to the penton subunit VP9 (light grey) and to the five copies of the polypeptide stretch (blue). The insets on the right show in wall-eye stereoview the densities (light-blue mesh) for the VP16 domains 1 and 2 closer to the vertex centre (top) and for the VP9 single β-barrel that has permitted the registering of the sequence (bottom).

The difference of HHIV-2 spike with respect to the other alphasphaerolipoviruses is also evident in the genomic arrangement. While in the rest of the virus of the genus the homology and the order of the ORFs encoding for the spike complex genes proteins is really high, there is no synteny with the spike genes in HHIV-2 (Figure 1.6) [71]. The identity of the proteins forming the spike complex have been already suggested in previous works by urea treatment [73], recognising the proteins VP16, VP17 and VP2 at the vertex complex.

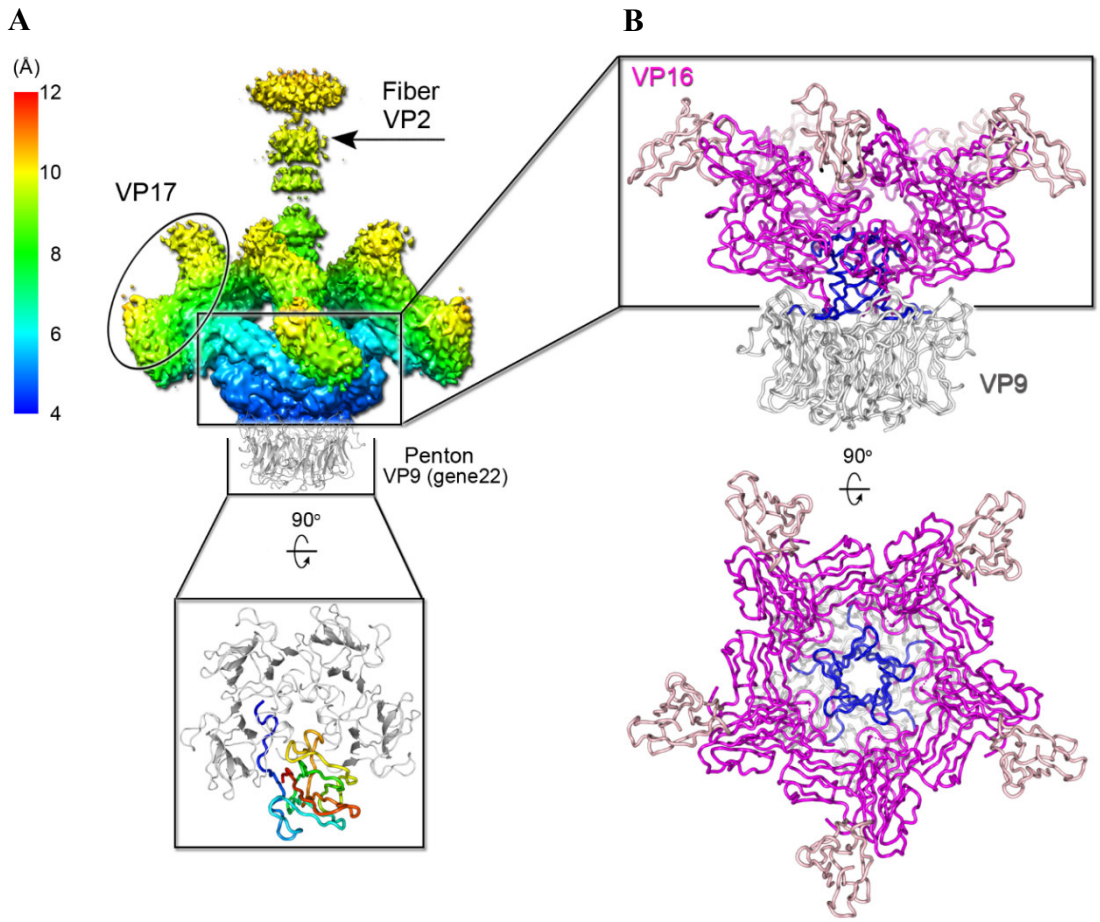


Figure 4.21. HHIV-2 vertex complex. **(A)** Local resolution estimation of the HHIV-2 spike complex density (resolution key on the left) with the penton represented in whitesmoke cartoon viewed perpendicularly to the icosahedral five-fold axis; regions corresponding to the putatively assigned fiber VP2 and VP17 are marked by a black arrow and oval respectively while the region assigned to the five copies of VP16 is defined by the black rectangle. The below inset shows a magnified view of the penton along the five-fold axis with one VP9 subunit colored from N-terminal (blue) to C-terminal (red) – the first visible N-terminal residues, about ten, clamp the base of the adjacent jelly-roll penton subunit. **(B)** (Top) poly-ALA model manually built and assigned to VP16 (about 217 residues modelled out of 253) represented as magenta cartoon tube while in pink a further possible putatively belonging to VP17; in the center in blue cartoon tube five copies of a poly-ALA stretch that coaxially cements VP16, VP9 (grey cartoon) and the central fiber (no models for VP2 and VP17 could be reliably traced); bottom, as above but viewed from the top to highlight the propeller-like appearance reminiscent of STIV and PM2 spike complexes [188 28].

The protein VP16 has been proposed to form the aforementioned multi-domain, as the 217 residues modelled in the density are really close to the length of the protein. (253 aa). Regarding VP17 (418 aa), it is thought that the outermost jelly-roll of the fitted multi-domain and probably the whole "blade" shape of the propeller on top are composed by this protein (Figure 4.21 and 4.22). Unfortunately, the high flexibility of this structure does not allow the fitting of any domain. Nevertheless, the secondary structure prediction

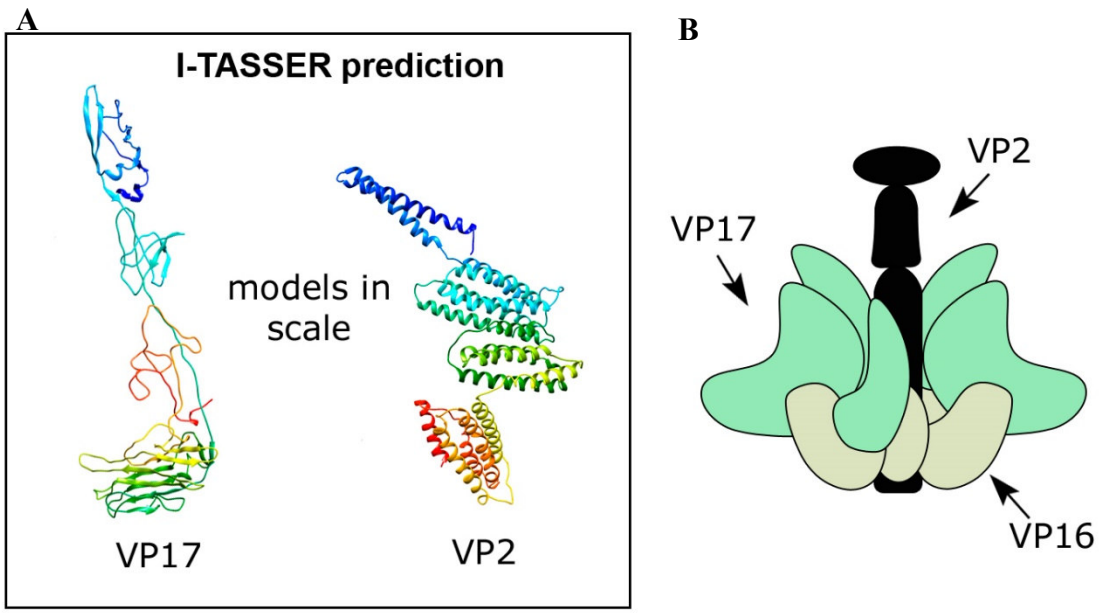


Figure 4.22. Secondary structure prediction of HHIV-2 spike proteins. **(A)** Possible model from I-TASSER server for the spike proteins. VP2 prediction showed an elongated structure with three different α-helix composed domains (confidence of structure prediction C-score = -1.15; typically from -5 to 2, higher the better) using as template structure PDB: 6R9T (Dedden et al. 2019, chain A). VP2 shares 11% sequence identity in the structurally aligned region. VP17 prediction showed a blade elongated structure made by β-strands and with a jelly-roll at the base (PDB: 5K8G (Guo et al. 2016, chain A) (C-score = -1.27), and sequence identity of 10%. **(B)** Proposed schematic architecture of the HCIV-1 spike complex with VP2 forming the central Fiber, surrounded by 5 copies of VP16 (light gray) and also 5 copies of the propeller-like VP17 (turquoise green).VP16 was not predicted as the poly-ALA model was manually built.

of VP17 shows an elongated structure with a jelly-roll at the base (Figure 4.22, A), that could be that identified in the map density

Despite the propeller-like structure of HHIV-2 is not present in the rest of the known viruses of this genus, some PRD1-like viruses conserve a reminiscent appearance. One of these is STIV1, virus, which displays five petal-like densities at the spikes [73]. Furthermore, the amino acid sequence similarity of STIV1 C381 protein with HHIV-2 proteins VP16 and VP17, even if low, is still appreciable (C381/VP16; 381/253 residues, 16.1 % aa identity and 29.1 % aa similarity and C381/VP17; 381/418 residues, 16.0 % aa identity and 26.7 % aa similarity, respectively). Also bacteriophage PM2 displays a petal-like architecture at the vertex. Comparing the amino acid sequence similarity of P1, the vertex protein of PM2, with HHIV-2 spike proteins, there is also a low but important similarity with VP16 (P1/VP16; 335/253 residues, 16.4 % aa identity and 27.8 % aa similarity).

Finally, the vertical fiber of the spike is putatively composed by VP2 (515 aa), although the high flexibility and the C5 symmetry applied for the reconstruction makes impossible the protein docking. This protein, despite retaining a certain homology and being present in all viruses of the genus, is the spike complex protein with greater variability in its length and morphology (Figure 1.6). Nevertheless, the fiber structure displayed in HHIV-2 remembers that of other PRD1-like viruses, such as PRD1 itself or the adenovirus[73].

R

Chapter 5:

RESULTS IN BVDV RECONSTRUCTION

In reference to the results regarding BVDV reconstructed tomograms, this work is still in progress, presenting here only the concluded results. These projects under way are intended to complete an overall picture of the virus structure, especially regarding to the glycoproteins binding and organization.

The subtomogram averaging performed on the BVDV envelope glycoprotein Erns bound to mAbs is being carried out also for E2 glycoprotein, although more subtomograms need to be collected to obtain a reliable model of the union event.

Also the distribution of Erns glycoproteins on the surface of the envelope is being mapped with the purpose of finding an interpretable pattern of distances between the glycoproteins.

Finally, the RNA of BVDV has been aligned by subtomogram avering, obtaining a 3D map, whose secondary structure needs to be predicted.

5.1 Identification and selection of the different sub-boxes

Once the tomograms were reconstructed, the next step was to identify in the 3D volumes the different regions of interest to be sub-boxed. A total of 76 reconstructed tomograms of BVDV + Erns Ab were analyzed, observing different morphologies of the virus. These were the different subtomograms sub-boxed using the 3D browser dtmslice from dynamo software (Figures 5.1 and 5.2) [171] :

1) Subtomograms of the Standard BVDV virions: These are the BVDV particles presenting a compact morphology filled up by the RNA. The subtomogram dimensions are a box of 256 pixel (709 Å). The number of subtomograms extracted are 331.

2)Subtomograms of the inflated BVDV virions: These are BVDV particles whose size is variable but they always present the outer envelope detached from the inner RNA, which still remain inside. Box dimensions are 336 pixel (930 Å) and the number of particles 81.

3) Subtomograms of the empty particles: BVDV particles showing an absence of internal RNA but conserving the outer envelope. The box dimensions are 256 pixel (709 Å) and the extracted particles 33.

4) Union events between anti-Erns monoclonal Antibodies (WB210) and the viral glycoprotein Erns: The links of putative antibodies to the surface of the viral envelope was observable in all de viruses described before. The box dimensions were 80 pixel (222 Å) in all cases, counting 2854 subtomograms coming from standard virions, 579

subtomograms coming from inflated particles and 129 spikes coming from empty particles.

5) BVDV RNA genome: condensed RNA structures and hairpin loops observed in the tomograms. They are found free in the sample, although the condensed RNA inside the bigger inflated vesicles was also sub-boxed. The box dimensions are 140 pixel (388 Å), being extracted 192 subvolumes.

6) Tethering membranes: The envelope of two different viruses are in close contact, fusing part of their envelopes. In this case 52 sub-boxes of the connected envelopes were extracted with a box dimension of 140 pixel (388 Å). Also the edge regions of the fused envelopes were extracted, obtaining 98 subvolumes with box dimensions of 80 pixel (222 Å).

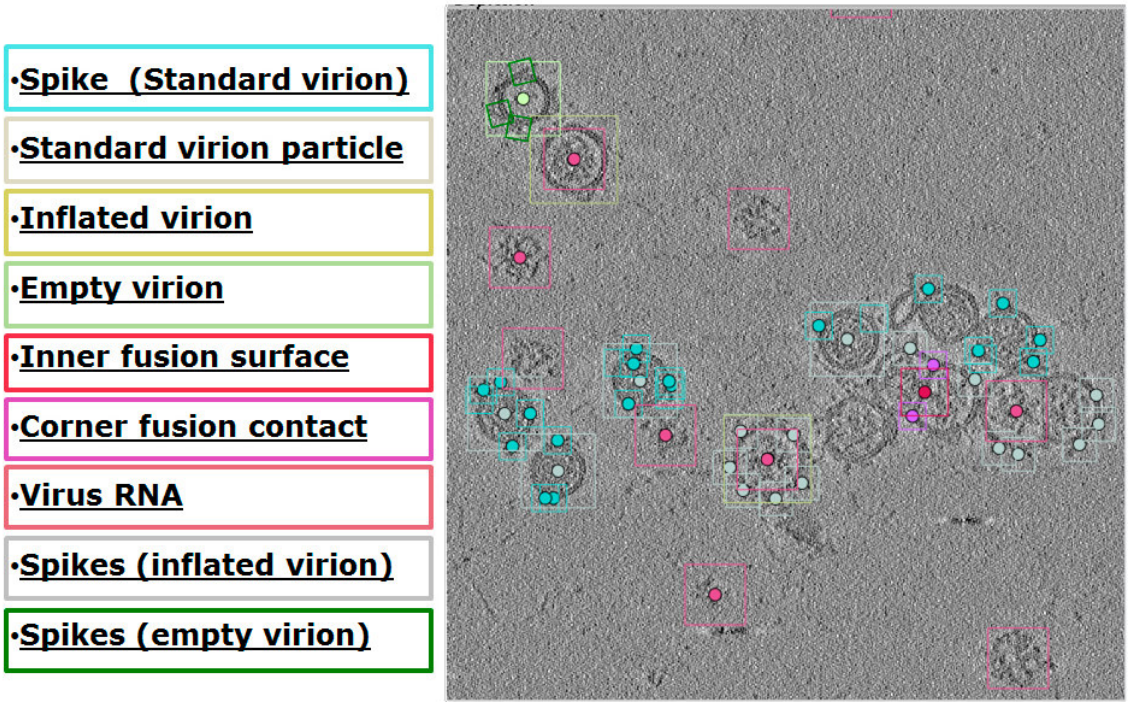


Figure 5. 1. Identification and selection of the different sub-boxes. Central section of a reconstructed tomogram of BVDV with monoclonal antibodies against Erns glycoprotein. The central point of the sub-boxes is marked by a colored sphere (with the same color of the referred sub-box). (Left) Name of the different morphologies and structures framed with the same color code at the tomogram.

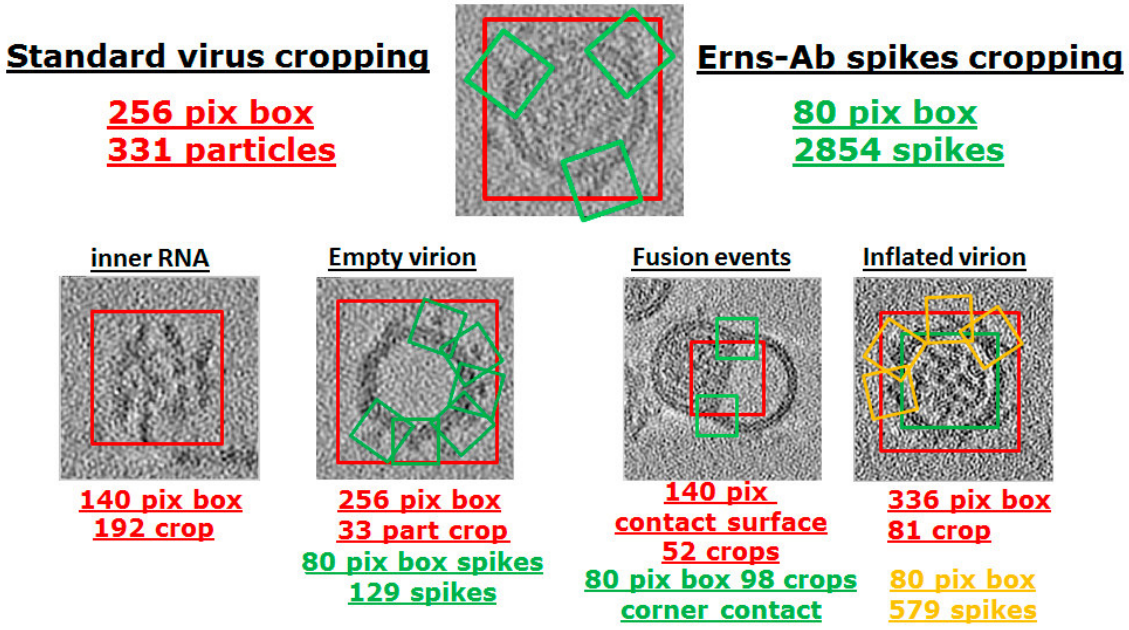


Figure 5. 2. Number of sub-boxes cropped for each morphology identified (as in Fig. 5.1). The box size (in pixel [3]) and number of crops are indicated for each case (marked with the same color code. The inner RNA of some inflated particles were also cropped (green square) and add up to those RNA found free in the tomogram.

On average each BVDV particle was decorated by 7 antibodies, but this was really variable, existing particles with just one spike identified and other viruses with around 20 spikes. Whether this represents the saturation of the binding sites available on the virion we can not state as also sterical hindrance caused by neighbouring mAbs and or binding efficiency of the mAbs cannot be estimated, furthermore some of the spikes could be blurred or difficult to identify, especially if they are mainly located in the missing wedge area.

5.2. Subtomogram averaging of BVDV envelope glycoprotein Erns bound to mAbs

To structurally characterize the regions where the specific monoclonal antibody (IgG) WB210 binds to the viral glycoprotein Erns, the 2,854 subvolumes extracted from the standard BVDV/mAb complexes were averaged together in dynamo [171]. The sub-boxes set was aligned and classified to obtain a three-dimensional model with a morphology closer to the real antigen + antibody junction. Thus, the subtomogram averaging of the subvolumes is necessary to achieve a final higher resolution of this model that could attests the likelihood of the data.

In the first alignment, a synthetic reference was used to orient all the membranes of the sub-boxes to the same Z plane, obtaining a preliminary model of the Ab-Erns union. In the next level of the processing, different alignment and classification projects, each of them which variable parameters, references and masks, were applied. The final model obtained was a bilobular spike attached to the viral membrane (Figure 5.3, B) (See methods).

The resulting averaged electron density at about 16 Å resolution estimated as from the gold-standard Fourier Shell Correlation (FSC) at 0.143 (Figure 5.3, A), displays a distinct morphology with a small turret density protruding from the outer membrane with on top a large boat-shaped density, overall suggesting the presence of a two-fold symmetrical structure (Figure 5.3, B). This density was readily interpreted in terms of homodimers of Erns as composing the small turret while the two extending density lobes as two Fab domains of two distinct IgG molecules (Figure 5.4). The remaining Fab and Fc domains of each IgG are not visible due to the expected conformational flexibility of these elements when not bound and/or restricted. The crystal structures of the truncated Erns (residues 1-165; PDB ID 4DV5) and of an IgG Fab domain (from PDB ID 1IGY) were rigid-body fitted into density using Chimera software (Figure 5.4, insets). The Erns molecule possesses an amphipathic helix in the last twenty-two residues (residues 205-227) that anchor the dimers onto the envelope membrane [142], which is not included in the fitting. The rigid-body fitting allows to enquire about the Erns region recognized by the mAb WB210. Erns dimers are arranged vertically on the surface of the lipid bilayer at its C-terminal end, putatively binding the mAbs region by the outer α-helix (residues 90-105) that spans the most outer part of the glycoprotein.

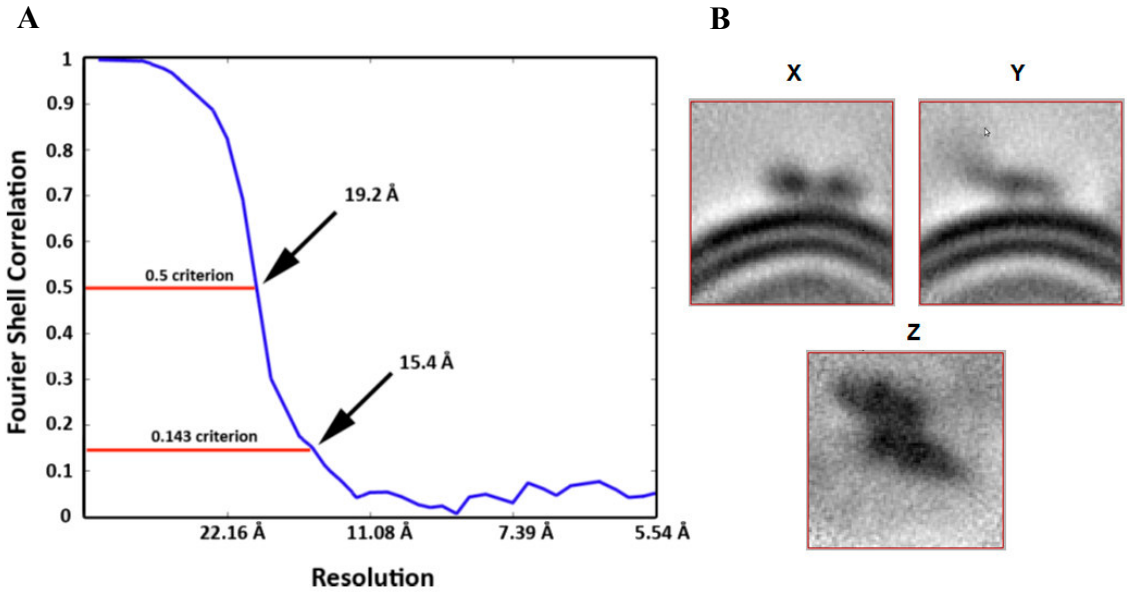


Figure 5. 3. Erns-mAb subtomogram averaging. **(A)** Fourier Shell Correlation (FSC) of the alignment obtained by "golden standard" methodology of the Erns-mAb union of 2,854 subvolumes, which reported a resolution of 15.4 Å at the 0.143 criterion. **(B)** Erns-mAb union raw central section, obtained by subtomogram averaging in dynamo [171], visualized from X, Y and Z planes. The FSC in (A) was measured from this volume.

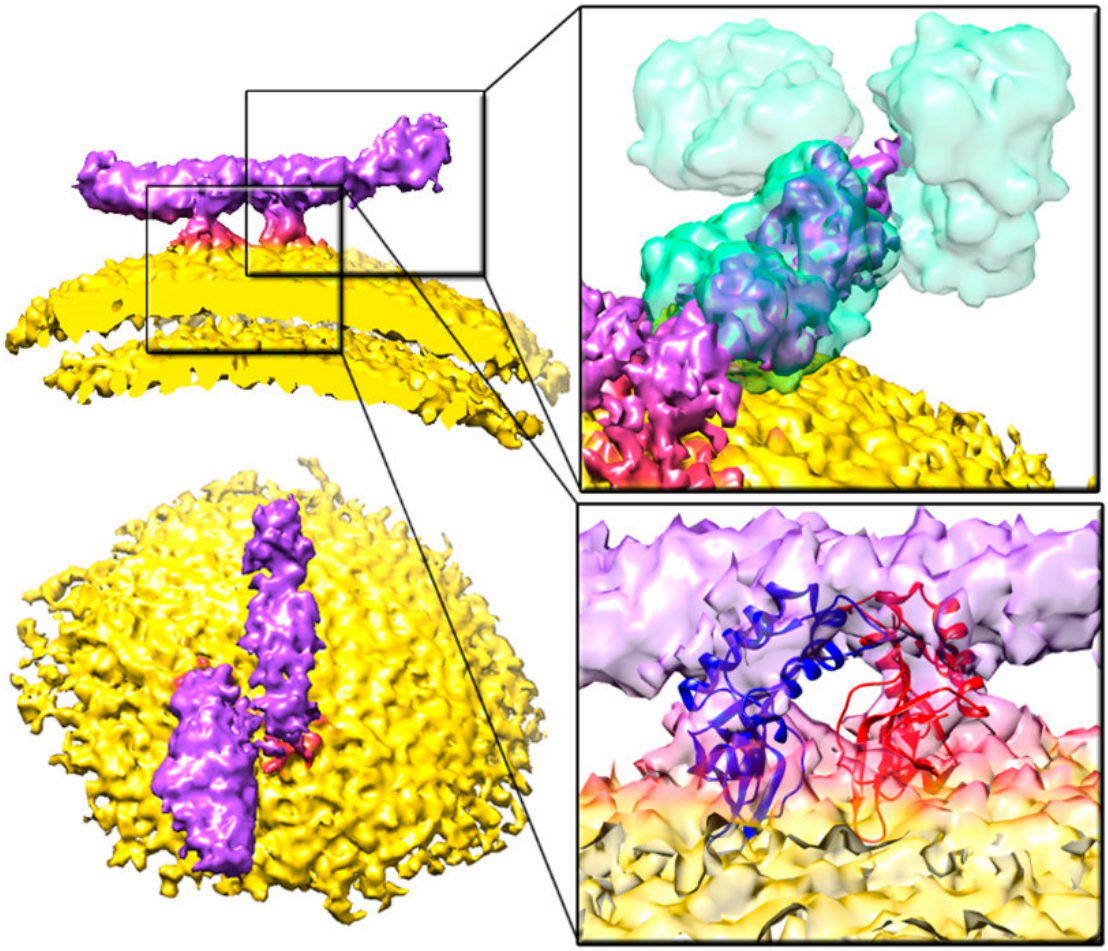


Figure 5. 4. Erns-mAb union model obtained from subtomogram averaging fitted into the density of the mAbs (top right) and the Erns dimer (down right) (PDB: 1IGY and 4DVK respectively). Lipid envelope (yellow), Erns glycoprotein (magenta), linked mAbs to Erns glycoprotein (purple) and mAbs (light cyan). The other Fab region and the constant region of the mAb, which are not fitted into the volume density, are shown with 60% mesh transparency. (Up left) Erns-mAb union model lateral view, with the membrane vertically sectioned before the glycoprotein dimer. (Down left) Erns-mAb union model viewed from top.

5.3 Morphological analysis of BVDV by volume segmentation

Finally, the tomograms presenting the higher variability of BVDV morphologies were also created by SIRT reconstruction, as this algorithm provides a better contrast of the viral elements (Figure 5.6, A) (see methods). These tomograms were segmented to better define the regions of the different structures observed in an individual tomogram, assigning characteristic colour annotations for the pixels belonging to each region (Figure 5.5).

In this way, the respective BVDV morphologies observed earlier during sub-boxing were individualized and segmented by SuRVos software [183]. The different perceived features were clearly discernible, getting a better picture of how BVDV elements are organized.

Regarding the standard virion, the inner RNA remains highly condensed, probably due to the core protein, and appears to join the inner leaflet of the envelope in several points (Figure 54.5, B). Respect to the inflated vesicles, this time the condensed RNA is mostly unattached from the envelope, leaving a free space in-between (Figure 54, B , in yellow). Concerning the empty viral vesicles, they conserve the outer lipid bilayer and the same round shape that the rest of the vesicles, not observing any rupture of the envelope that could explain the delivery of the inner RNA (Figure 5.6, d). Therefore, it is probable that this vesicles present an inefficient packaging of the viral RNA. In regard to the tethering membranes viewed in same vesicles, the common surface of the fused viral particles is formed by one layer coming from each virus (Figure 5.5, A). The freely found RNA also maintains a condensed structure in defined domains, despite not being enclosed by the viral envelope.

Looking at the linked antibodies against the viral glycoprotein Erns, in spite of the great variability of shapes and morphologies that the envelope presents, in all cases there are spikes attached to its surface. In fact, thanks to the segmentation of the particles, new spikes have been identified that had not been detected during the sub-boxing tomogram. (Figure 5.6, a-d))

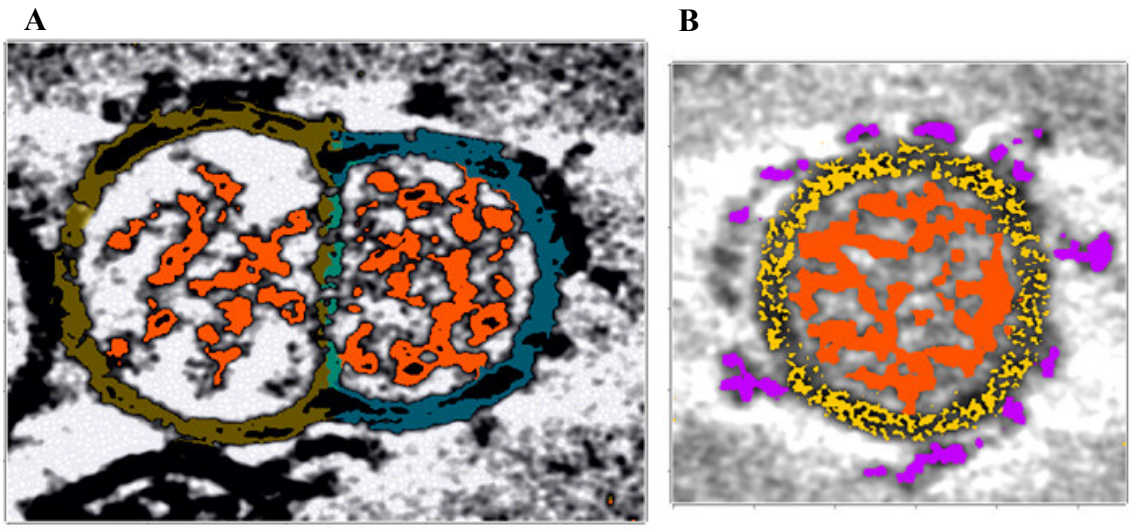


Figure 5. 5. Segmentation of BVDV morphologies. **(A)** Central section of the segmentation of BVDV fused particles. The volume has been processed with different filters in SuRVos [183]. The different structures observed were demarcated and colored (RNA in orange; envelope of each virus in dark yellow and blue). In this example we can discern how the common membrane of the fused viral particles is formed by one layer coming from each virus. **(B)** Central section of the segmentation of BVDV standard vesicle. The inner RNA (orange), envelope (yellow) and linked anti-Erns mAbs (purple) are also colored.

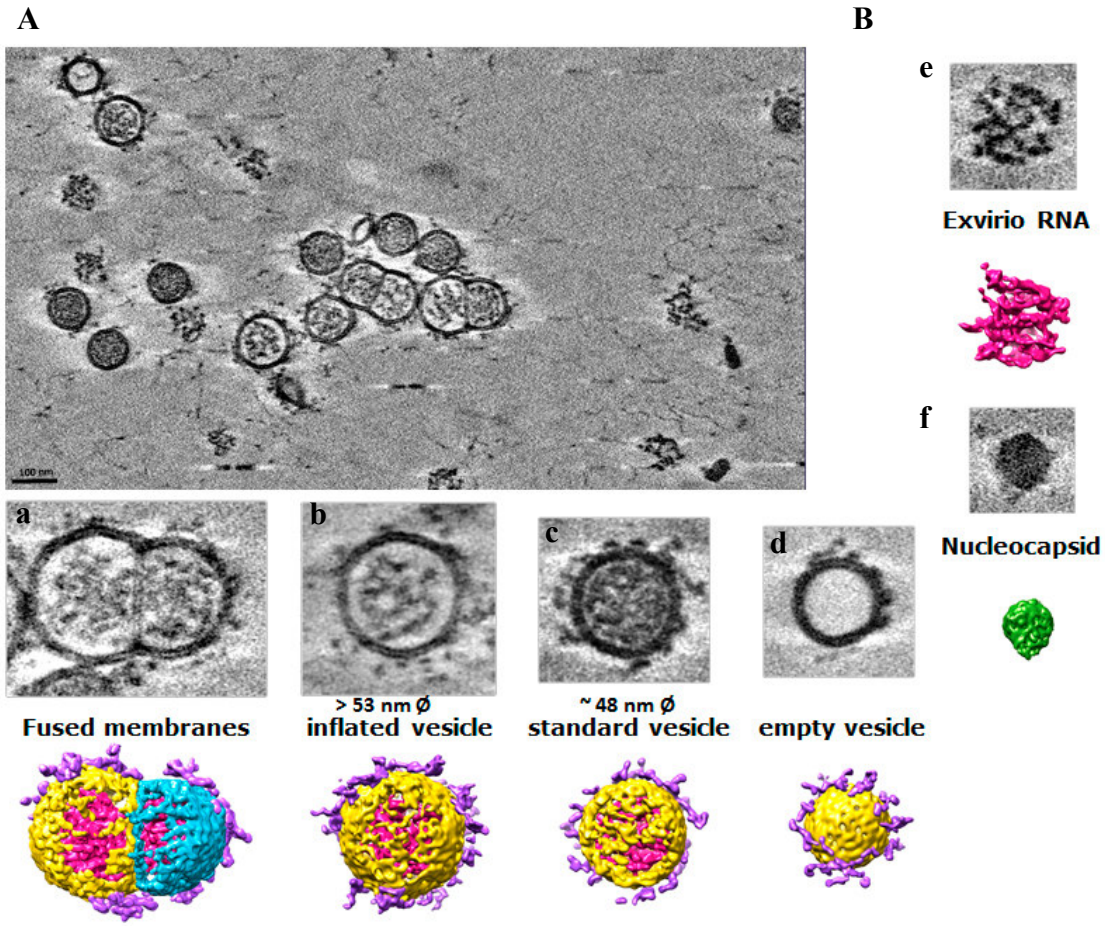


Figure 5. 6. Morphological analysis by segmentation. Reconstructed tomogram with representative BVDV sample composition. The sample was collected by Cryo-ET and the 3D reconstruction was performed in IMOD [164]**. (A)** Central section of the reconstructed tomogram. **(B)** Different morphologies of BVDV incubated with anti-Erns mAb found in the tomograms (a-f on top). Each of the structures were cropped independently from an individual tomogram and segmented by SurVos segmenting software [183] (a-f down). Lipid envelope (Yellow), viral RNA (magenta), linked mAbs to ERNS glycoprotein (purple), nucleocapsid (green).

Chapter 6:
DISCUSSION

6.1 Capsid assembly in icosahedral viruses with 2 major capsid proteins

In recent years, the structural knowledge obtained about membrane-containing icosahedral viruses belonging to PRD1-like lineage has meant a better understanding of their diversity and organization. In this sense, the assembly mechanism of the viral particles has provided interesting new insights.

Generally, the internal membrane of these viruses behaves as scaffold, onto which the capsid proteins are anchored. This generated protein-lipid system provides a mechanic stability in the mature virion. There is a common architecture in the assembly, based on the major capsid proteins (MCPs), which constitute the pseudo-hexameric capsomers of the capsid lattice, and the penton proteins, placed at the 5-fold vertices of the capsid. In addition to that, each virus type has developed its own strategies to improve the capsid packing. Although the MCPs of this lineage retain the same vertical β-barrel folding, the inter and intra-capsomer interactions and their membrane-associated connections vary considerably in each type of virus. Moreover, there are usually accessory structural proteins, responsible for several functions such as the correct registration of the capsomers and the cementing of the capsomers.

In this work we offer new insights about the assembly mechanisms in the particular case of viruses with two vertical single β-barrel as MCPs, displaying also capsomers with two distinct morphologies (e.g. two-tower and three-tower). Thus, by investigating the archaeal, halophilic, internal membrane-containing icosahedral viruses HCIV-1 and HHIV-2 at 3.7 and 3.8 Å resolution, respectively, using high-resolution (HR) Cryo-EM and proteomics analysis, we have finally deciphered the principles governing the organization of vertical single β-barrel viruses. This structural arrangement has been described at all levels, as detailed below.

In a first step, the specific interactions between VP7 and VP4 MCPs define the functional arrangement of the VP7–VP4 heterodimer. The specific interactions between these proteins consist of two main unions. On the one hand, the N-terminal end of VP7 embraces the base of the neighbouring VP4 jelly-roll placed clockwise. On the other hand, the protruding loops of both proteins fit in the notch of their respective pairs establishing hydrogen bond interactions. These junctions are reinforced by the larger hydrophobic interface area between both proteins. This building block is really stable and is probably formed prior to its incorporation into the capsid lattice.

Now, if we look at the MCPs of P23-77, which present homologous vertical single β-barrel proteins, the intra-capsomer interactions are different from those observed in HCIV-1 and HHIV-2. The MCPs of this virus are assembled in three different building blocks, from which the non-turreted MCP homodimer, VP16, is the main one. In this case the long N-terminal of VP16 is embracing the homodimer, while in HCIV-1 the N-terminal of VP7 is embracing the base of the VP4 jelly-roll. In contrast, the interactions between VP16-VP17 that form the heterotrimers and heterotetramers in P23-77 are different and much weaker than in HCIV-1 [74]. On the contrary, the recently published arrangement of SH1 heterodimer reflects the same type of unions described in this work, which is coherent with the highly conserved sequence homology of its MCPs and those of HCIV-1 and HHIV-2. Regarding the viruses presenting double β-barrel capsid proteins, they only need three MCPs for building up the capsomer, halving the number of necessary proteins for making the pseudo-hexamers and probably reducing the assembly error rate [191]. The intra-capsomer interactions in these viruses are also different. Each PRD1 bacteriophage subunit, for instance, has a trimerization loop (FG1) that interacts with the other two subunits [192].

In a second step, the consolidation of the capsomers is presumably stabilized by their outermost notches, which connect neighbouring capsomers within and across the IAU through interactions related to arginine residues. Moving to P23-77, the spanning densities connecting the capsomers are located at the same position as in HCIV-1, although the bridges are only formatted between VP16 proteins[74]. In HCIV-1 and SH1 viruses, instead, the intercapsomer interactions occur in all possible subunit combinations, that is, VP4-VP4, VP4-VP7 and VP7-VP7. In all cases the subunit pairs are packed by pseudo two-fold related motifs. Interestingly, despite the great similarity in the pattern of unions of HCIV-1 and SH1, in the first the arginine residues are conceivably involved in the H-bonds and salt-bridges across subunits, while in SH1 are the phenylalanine residues which dictate the capsomer assembly[185]. In double β-barrel viruses the inter-capsomer connections follow a different scheme. When 3 different capsomers interface with each other the 3 subunits involved interact in a cyclic fashion at the centre of the trimer, such as in PR772 bacteriophage [193].

In a third step, the VP7–VP4 heterodimers interact with membrane-proximal proteins GPS-II and GPS-III whose location, fold, and oligomerization state dictate the positioning and formation of the pseudo-hexagonal two-tower and three-tower capsomers onto the

membrane vesicle. The role of these GPS proteins establishing the correct registration of the two different capsomers, is a novel understanding of how the capsid blocks are addressed to conform the lattice.

Albeit in a different manner, the minor capsid protein VP11 of P23-77 fulfills a function similar to GPS-II and GPS-III. In contrast, the relative position in the genome gene sequence with respect to alphasphaerolipovirus is different. In addition, VP11 is a dynamically elongated dimer with a predominantly α-helical secondary structure and high thermal stability, this reminds of VP10 in HCIV-1, since both have an α-helical structure and a 2-fold symmetry. In addition, in both cases the protein between the capsid and the membrane is located in pseudo-hexamers presenting 2 turrets. In the case of P23-77 it has been shown that VP11 associates with lipids of the membrane and only with the MCP VP17 [88], although there are no structural data. In HCIV-1, VP10 is also associated to the membrane and has the axis of symmetry with the two VP4 towers as poles. Furthermore, the protein band intensity of VP11 in a Coomasie SDS-PAGE gel was measured, estimating a total of 147 proteins per virus [88]. This number is pretty close to either the number of copies of GPS-III (180) or GPS-II (120), suggesting that perhaps VP-11 is not present in all the capsomers. Therefore everything seems to indicate that VP11 in P23-77 and VP10 in HCIV-1 are structures with analogous function directing the scaffolding for the correct assembly of the MCPs, but their homologous origin is quite unlikely. A possible explanation may be due to the fact that the structural proteins which are not part of the IAU have evolved diversely in different genera of this family, adapting in each case a distinct organization of MCPs but always conserving the MCPs framework with single β-Barrel. These so-called GPS proteins are not limited to VP10 or VP12 and their counterparts in other viruses, but there are other genome proteins that have monitored other types of MCPs organization. In SH1, the GPS proteins proposed occupy the same densities that in HHIV-2 and HCIV-1, although the designed identity of the proteins is not homologous in some cases [185]. In the case of Adenovirus and PRD1, both double β-barrel, GPS are not necessary, since they organize equal MCPs and it is only required to organize and cement between capsomers. PRD1, for instance, disposes the minor capsid protein P30 along the 2-fold, embracing the base of the capsomers, and acting therefore as a cohesive scaffold [79]. In adenoviruses, this cementing has been further reinforced, displaying minor capsid proteins on the outer surface (protein IX), as well as inside the capsid (proteins IIIa, VI and VIII) [194]. Through a different but analogous system, also

PM2 bacteriophage (P3 and P6 proteins) or STIV virus (A55 and B130 proteins), exhibit glue proteins that stabilize the capsid shell [49]. Considering this, the protein system involved in the positioning and cohesion of the capsomers in double β-barrel-like viruses, is a different adaptation to the single β-barrel-like viruses. These viruses may have arisen in a cladogenic (divergent) or anagenetic way (having passed before by the e.g. state of HICV-1-like), although the latter one is hard to believe. Its fusion of MCPs probably made GPS useless, (e.g. in PRD1) but the membrane was still useful for anchoring the capsid, though in adenovirus the cemented capsid was reinforced and the membrane was no longer needed. This hypothesis is supported by the fact that there are no viruses with internal membrane and 2 MCPs displaying cementing proteins embedded between capsomers at both sides of the MCPs (internally and externally), as is the case with adenovirus. Thus, it is unlikely a common origin of the proteins involved in the positioning of the MCPs in PRD1-like lineage. Rather, it appears that each virus has followed its own strategies to consolidate the capsid, taking for that different minor structural proteins of the viral genome.

It would be interesting to know whether there is any virus presenting just one single vertical β-barrel as MCP and also an internal membrane. This assumption would be a previous evolutionary state regarding virus of PRD1-like lineage. In this instance there would be no differentiation in the positioning of the MCPs across the IAU. However, the large amount of proteins that would form the lattice could compromise the stability of the capsid. Therefore, also in this hypothetical viruses it is expected the presence of additional proteins for building up the capsomers, although their structural arrangement have yet to be revealed.

The complexity in the assembly of the capsid components is also reflected in the genomic organization of the virus. Normally the ORFs of functionally related genes are clustered under the same operon, facilitating the assembly process in separated successive steps. In HCIV-1 and HHIV-2 viruses the structural proteins related to the capsid and the vertex complex are transcripted together in just 2 contiguous operons [83]. In PRD1, despite containing a genome with less genes, the structural genes of the virion are mainly contained in the three late operons [195]. Finally, in Adenovirus the structural genes are organized in 5 transcription units regulated by the same promoter [196]. Therefore, evolutionarily the genomic organization of structural proteins seems to be more orderly

and systematic in adenovirus or PRD1 than in HHIV-2 or HCIV-1, possibly affecting the assembly eficiency. Nevertheless, this premise should be verified appropriately.

6.2 Structural adaptation of cell receptor binding complexes.

All known viruses utilize one or more viral proteins to successfully interact with the host receptors. This virus–receptor binding is usually highly specific, sometimes infecting a single species or cell type. The cell resistance to the infectivity can evolve in a host-virus arm race, provoking conformational changes that define the virus species [2]. In PRD1-like viruses, while the MCP proteins are really uniform throughout the species, the receptor binding complexes, also known as spike complexes, display a high variability of shapes. The location of the spike complexes is always located at the 5-fold vertices, where they connect with the penton, which acts as a "universal adaptor".

In the case of the HCIV-1 and HHIV-2 viruses, the morphology of their spike complex is totally different. In the case of HCIV-1 the spike is very similar to that described in SH1 virus, forming a horn-like shape with C2 symmetry. Despite this, there are slight differences between both spikes. First, each arm of HCIV-1 shows an architecture as a pearl necklace composed by six layers, while in SH1 there are only 5 globular layers. In the case of HCIV-1, the high flexibility of the spike has been verified by different 2D and 3D classifications, observing different aperture degrees of the horn. In SH1 its flexibility is not known with certainty, but it is probably somewhat smaller due to the absence of this last layer. Interestingly, the VP2 protein, which probably forms the scaffold of the horns, is shorter in HCIV-1 (593 residues) than in SH1 (824 residues). The sequence homology of both chains is quite high, except for the first 330 amino acids of VP2 in SH1

Regarding HHIV-2 virus, the ORFs encoding for the spikes do not have homology with the rest of the viruses belonging to the genus *Alphasphaerolipovirus*. HHIV-2 spike complex displays two charecteristic elements. The first one is a petal-like architecture with C5 symmetry, which vaguely resembles the spike shape pointed out in PM2 bacteriophage and STIV virus. The second one is a central vertical fiber, which also rembembers the spikes of other members of PRD1-like lineage. Therefore, it can be thought that the viruses of this lineage have very different morphologies in their spike complex, but these tend to converge in discrete models that present common elements, such as the fiber, the horns or the petal shape. This fact probably indicates that these spike

morphologies are probably the most efficient ones for the interaction with the host receptors in these type of viruses. In the case of HCIV-1 and HHIV-2, both are capable of infecting organisms of the same species, specifically the *Haloarcula* genus. It is therefore possible that each virus interacts with a different target receptor of the host or with different areas of the same receptor, thus using different infection strategies.

The morphological divergence of the spike complex shown in the HCIV-1 and HHIV-2 viruses has been described in other genera of the PRD1-like lineage, as is the case of the genus Corticoviridae. The bacteriophage PM2 and the prophage Cr39582, for instance, are syntenous along the whole genome sequence except for the region encoding for the spike proteins, even though they infect the same cell type [197].

Furthermore, there are viruses without any evolutionary relationship that nevertheless share analogous structures in their spike complex. This is the case of *Acidianus* filamentous virus 1 (AFV1), member of the *Lipothrixviridae* family, and HCIV-1 [198]. In both cases, cell receptor binding spikes present a claw-like morphology. This illustrates once again that some spike morphologies are favoured by evolutionary pressure.

Finally, the dissimilar vertex complexes of HCIV-1 and HHIV-2 have showed how the pentons work as adaptors for distinct spikes. In both viruses the spike is glued to the 5-fold vertex by strand–strand interactions, fastening both elements and solving the symmetry mismatch. In HCIV-1 the spike inserts five linear strands inwards the capsid, which are placed in parallel to each strand G of the VP9 proteins. In HHIV-2 the anchoring of the spike is somewhat different, since all the proteins in the complex are coaxially glued together to the central VP2 fiber. In this case, an accessory protein still to be identified is cementing the spike junction to the VP9 pentons, reinforcing the joint.
Inside the capsid, the 5-fold vertices of both viruses show a strong connection of the pentons with the inner membrane. Roughly the last 50 residues at the N-terminal of VP9 generate a five-helix coiled-coil structure that interacts with the membrane bilayer, which accommodates a membrane protein complex. This membrane complex is putatively formed by VP13 homopentamers, with the transmembrane helices forming a left-handed coiled-coil, acting as docking rafts for the VP9 subunits. Therefore, the five-folds could be considered as the nucleation site for the capsid assembly, extended later to the attachment of the pseudo-hexamers to complete the whole virion lattice.

Other viruses of PRD1-like lineage also follow this general scheme of assembly at the 5-fold vertex, although with significant differences. In PRD1-like Bacteriophage PR772, for instance, P5 and P2 spikes are attached to P31 homopentamer. Also, the protein P16 links the adjacent MCPs to this pentamer and at the same time is contact with the membrane, establishing the starting points for the virus assembly [193]. In Adenovirus, the fiber 3-fold N-terminal domain associates with the homopentamer of the 5-fold, solving also a symmetry mismatch, and the protein IIIa mediates the binding between peripentonal pseudo-hexamers and the pentons [199]. This time the 5-fold is not anchored to the membrane, since it is lacking in the virus.

In conclusion, the 5-fold vertex is the cornerstone for the starting of the assembly and infection processes in PRD1-like viruses. The organization of this axis and the possible mismatches between capsid and spike are solved with different binding strategies in each virus. In addition, the membrane provides an anchor support that stabilizes the 5-fold vertex and the underlying capsomers.

6.3 The challenge of subtomogram averaging in BVDV glycoproteins

Structurally, tackling pleomorphic viruses remains a challenging task. The advent of powerful microscopes, direct detection camera, phase plate, subtomogram averaging techniques and reliable software has allowed to progress in our understanding of viral pathogenesis and assembly. However, the identification and sub-boxing of small glycoproteins embedded at the envelope, as those presented in BVDV, still results difficult to discern by eye in a 3D viewer.

Regarding the BVDV envelope glycoprotein Erns bound to mAbs, it is likely that the high flexibility of the antibody is causing a high variability of the union event. While glycoprotein dimers are roughly arranged vertically on the surface of the lipid bilayer, the density lobes corresponding to the Fab domains of two distinct IgG molecules occupy an area considerably larger than the size of the Fab due to the wide range of unions As a consequence of this fact, it is possible that the C2 symmetry observed in the Erns dimmer crystals is not sufficient to also orient the Fabs in the same plane of symmetry. Therefore, the performed subtomogram averaging has given better results without applying any symmetry (C1) than with the C2 symmetry.

Nevertheless, there are still options to improve the alignment and get a final model with increased resolution. In this sense, we are now in progress to accomplish this objective by completing the following tasks:

1) 3D reconstruction of new tomograms of BVDV Erns-mAb. By acquiring new tomograms it will be possible to obtain a larger number of subtomograms that will help to better align the final model.

2) Template extraction of Erns-mAb using the preliminary 3D model previously obtained: The model already obtained at medium resolution (15.4 Å resolution) can serve as a template to track new joints along the envelope of the already reconstructed tomograms. The identification of the binding events manually prevents in many cases detecting all the antibodies bound to the glycoprotein due to blurring caused in planes other than Z during the tomogram reconstruction. In addition, during the segmenting of some tomograms additional spikes have been observed that were not selected during the picking of the sub-boxes. Therefore, the use of a template closer to the real glycoprotein-Ab binding event could help increasing the number of selected spikes.

This same scheme is also intended to be carried out for the BVDV envelope glycoprotein E2 bound to its specific mAbs. In this case a glycoprotein homodimer should be also found at the envelope surface, although how the antibody is linked to glycoprotein have yet to be resolved.

Chapter 7:

CONCLUSIONS

Archaeal viruses HCIV-1 and HHIV-2

a)

For the first time in the field of electron microscopy, this work rigorously describes the structure of two icosahedral viruses belonging to the genus *Alphasphaerolipovirus* possessing two vertical single β-barrels.

Three-dimensional reconstruction from two-dimensional images of the HHIV-2 and HCIV-1 viruses has allowed the generation of a 3D model of the main viral proteins at high resolution, estimated at 3.8 Å and 3.7 Å for the capsid shell, respectively.

b)

The GPS-III and GPS II proteins of HHIV-2 and HCIV-1 are essential for the orientation of the MCPs in viruses with two different individual vertical β-barrels. Therefore, the presence of these proteins is essential for the assembly of the protein capsid in viruses of this type.

c)

We truly believe that the complexity of the organization of the assembly of individual vertical β barrels in viruses such as HHIV-2 and HCIV-1 has been resolved by the event of fusion of the two consecutive MCP genes in the virus genome, simplifying the capsid assembly and allowing the absence of internal membrane in some cases.

d)

Despite the high degree of habitat isolation of the different viruses belonging to the genus *Alphasphaerolipovirus*, the sequence of the MCPs and their folding are highly conserved, indicating that the selective pressure of the medium favours the maintenance of this type of assembly.

e)

HHIV-2 and HCIV-1 receptor recognition complexes do not retain a common protein sequence and folding. The selective pressure of these complexes has been different in each virus, evolving towards different structures adapted in each case to its specific cellular receptor.

Animal-infecting envelope virus BVDV

a)

The different BVDV virus morphologies have been described for the first time in 3D by Cryo-ET and subtomogram averaging techniques.

b)

The BVDV virus, in a virion state, has a pleomorphic three-dimensional morphology, externally showing a lipid bilayer decorated with glycoproteins without apparent symmetric ordering.

c)

The specific binding of Erns glycoprotein to the mAb WB210 indicates the presence of Erns dimers embedded in the lipid envelope.

d)

Such dimers are arranged vertically on the surface of the lipid bilayer at its C-terminal end and with antibody binding to the outer α-helices (residues 90-105) of ERNS.

8. BIBLIOGRAPHY

1. Häring, M. *et al.* Virology: independent virus development outside a host. *Nature* **436**, 1101–1102 (2005).

2. Coffin, J. M. Virions at the Gates: Receptors and the Host-Virus Arms Race. *PLoS Biol.* **11**, (2013).

3. Fan, Y., Sanyal, S. & Bruzzone, R. Breaking Bad: How Viruses Subvert the Cell Cycle. *Front. Cell. Infect. Microbiol.* **8**, 396 (2018).

4. Demogines, A., Abraham, J., Choe, H., Farzan, M. & Sawyer, S. L. Dual Host-Virus Arms Races Shape an Essential Housekeeping Protein. *PLoS Biol.* **11**, (2013).

5. Koonin, E. V. Viruses and mobile elements as drivers of evolutionary transitions. *Philos. Trans. R. Soc. B Biol. Sci.* **371**, (2016).

6. Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M. L. & Brüssow, H. Phage as agents of lateral gene transfer. *Curr. Opin. Microbiol.* **6**, 417–424 (2003).

7. Wen, A. M. & Steinmetz, N. F. Design of virus-based nanomaterials for medicine, biotechnology, and energy. *Chem. Soc. Rev.* **45**, 4074–4126 (2016).

8. Thangavelu, R. M., Ganapathy, R., Ramasamy, P. & Krishnan, K. Fabrication of virus metal hybrid nanomaterials: An ideal reference for bio semiconductor. *Arab. J. Chem.* (2018). doi:10.1016/j.arabjc.2018.07.006

9. Gill, J. J., Wang, B., Sestak, E., Young, R. & Chu, K. H. Characterization of a Novel Tectivirus Phage Toil and Its Potential as an Agent for Biolipid Extraction. *Sci. Rep.* **8**, 1–11 (2018).

10. Krupovic, M., Dolja, V. V. & Koonin, E. V. Origin of viruses: primordial replicators recruiting capsids from hosts. *Nat. Rev. Microbiol.* **17**, 449–458 (2019).

11. Krupovic, M. & Koonin, E. V. Multiple origins of viral capsid proteins from cellular ancestors. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E2401–E2410 (2017).

12. Legendre, M. *et al.* Diversity and evolution of the emerging Pandoraviridae family. *Nat. Commun.* **9**, (2018).

13. Legendre, M. *et al.* Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 4274–4279 (2014).

14. Karuppannan, A. K. & Opriessnig, T. Porcine circovirus type 2 (PCV2) vaccines in the context of current molecular epidemiology. *Viruses* **9**, 1–15 (2017).

15. Abrescia, N. G. A., Bamford, D. H., Grimes, J. M. & Stuart, D. I. Structure Unifies the Viral Universe. *Annu. Rev. Biochem.* **81**, 795–822 (2012).

16. Rodrigues, R. A. L. *et al.* An anthropocentric view of the virosphere-host relationship. *Front. Microbiol.* **8**, 1–11 (2017).

17. Martin, W., Weiss, M., Neukirchen, S., Nelson-Sathi, S. & Sousa, F. Physiology, phylogeny, and LUCA. *Microb. Cell* **3**, 582–587 (2016).

18. Bandea, C. The Origin and Evolution of Viruses as Molecular Organisms. *Nat. Preced.* 1–16 (2009). doi:10.1038/npre.2009.3886.1

19. Baltimore, D. Expression of animal virus genomes. *Bacteriol. Rev.* **35**, 235–241 (1971).

20. Lefkowitz, E. J. *et al.* Virus taxonomy: The database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res.* **46**, D708–D717 (2018).

21. Krupovic, M. & Bamford, D. H. Order to the Viral Universe. *J. Virol.* **84**, 12476–12479 (2010).

22. Sinclair, R. M., Ravantti, J. J. & Bamford, D. H. Nucleic and Amino Acid Sequences Support Structure-Based Viral Classification. *J. Virol.* **91**, (2017).

23. Rohwer, F. Global phage diversity. *Cell* **113**, 141 (2003).

24. Nasir, A. & Caetano-Anollés, G. Identification of capsid/coat related protein folds and their utility for virus classification. *Front. Microbiol.* **8**, 1–14 (2017).

25. Naitow, H., Tang, J., Canady, M., Wickner, R. B. & Johnson, J. E. L-a virus at 3.4 å resolution reveals particle architecture and mrna decapping mechanism. *Nat. Struct. Biol.* **9**, 725–728 (2002).

26. Helgstrand, C. *et al.* The Refined Structure of a Protein Catenane: The HK97 Bacteriophage Capsid at 3.44 Å Resolution. *J. Mol. Biol.* **334**, 885–899 (2003).

27. Logan, D. & Abu-Ghazaleh, Robin, Blakemore, Wendy, Curry, Stephen, Jackson, Terry, King, Andrew, Lea, Susan, Lewis, Richard, Newman, John, Parry, N. Structure of a major immunogenic site on

foot-and-mouth disease virus. *Nature* **362**, 3 (1993).

28. Abrescia, N. G. A. *et al.* Insights into Virus Evolution and Membrane Biogenesis from the Structure of the Marine Lipid-Containing Bacteriophage PM2. *Mol. Cell* **31**, 749–761 (2008).

29. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

30. Koonin, E. V. & Dolja, V. V. A virocentric perspective on the evolution of life. *Curr. Opin. Virol.* **3**, 546–557 (2013).

31. Krupovič, M. & Bamford, D. H. Virus evolution: How far does the double β-barrel viral lineage extend? *Nat. Rev. Microbiol.* **6**, 941–948 (2008).

32. Forterre, P. The two ages of the RNA world, and the transition to the DNA world: A story of viruses and cells. *Biochimie* **87**, 793–803 (2005).

33. Forterre, P. The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res.* **117**, 5–16 (2006).

34. Koonin, E. V. & Martin, W. On the origin of genomes and cells within inorganic compartments. *Trends Genet.* **21**, 647–654 (2005).

35. Koonin, E. V. & Dolja, V. V. Virus World as an Evolutionary Network of Viruses and Capsidless Selfish Elements. *Microbiol. Mol. Biol. Rev.* **78**, 278–303 (2014).

36. Gill, S. & Forterre, P. Origin of life: LUCA and extracellular membrane vesicles (EMVs). *Int. J. Astrobiol.* **15**, 7–15 (2016).

37. Ono, A. Viruses and lipids. *Viruses* **2**, 1236–1238 (2010).

38. Hoen, E. N., Cremer, T., Gallo, R. C. & Margolis, L. B. Extracellular vesicles and viruses: Are they close relatives? *Proc. Natl. Acad. Sci. U. S. A.* **113**, 9155–9161 (2016).

39. Jalasvuori, M. & Bamford, J. K. H. Structural co-evolution of viruses and cells in the primordial world. *Orig. Life Evol. Biosph.* **38**, 165–181 (2008).

40. Poranen, M. M., Bamford, D. H. & Oksanen, H. M. Membrane-Containing Bacteriophages. *eLS* 1–11 (2015). doi:10.1002/9780470015902.a0000779.pub3

41. Mäntynen, S., Sundberg, L. R., Oksanen, H. M. & Poranen, M. M. Half a century of research on membrane-containing bacteriophages: Bringing new concepts to modern virology. *Viruses* **11**, (2019).

42. Mateu, M. G. *Structure and Physics of Viruses*. (Springer Netherlands, 2013). doi:10.1007/978-94-007-6552-8

43. Andres, G., Charro, D., Matamoros, T., Dillard, R. S. & Abrescia, N. G. A. The cryo-EM structure of African swine fever virus unravels a unique architecture comprising two icosahedral protein capsids and two lipoprotein membranes. *J. Biol. Chem.* **295**, 1–12 (2019).

44. Hulo, C. *et al.* ViralZone: A knowledge resource to understand virus diversity. *Nucleic Acids Res.* **39**, 576–582 (2011).

45. Espejo, R. T. & Canelo, E. S. Origin of phospholipid in bacteriophage PM2. *J. Virol.* **2**, 1235–40 (1968).

46. Espejo, R. T. & Canelo, E. S. Properties of bacteriophage PM2: A lipid-containing bacterial virus. *Virology* **34**, 738–747 (1968).

47. Olsen, R. H., Siak, J. S. & Gray, R. H. Characteristics of PRD1, a plasmid-dependent broad host range DNA bacteriophage. *J. Virol.* **14**, 689–99 (1974).

48. Laurinavičius, S. *Phospholipids of lipid-containing bacteriophages and their transbilayer distribution*. (2008).

49. Khayat, R., Fu, C. -y., Ortmann, A. C., Young, M. J. & Johnson, J. E. The Architecture and Chemical Stability of the Archaeal Sulfolobus Turreted Icosahedral Virus. *J. Virol.* **84**, 9575–9583 (2010).

50. Santos-Pérez, I. *et al.* Membrane-assisted viral DNA ejection. *Biochim. Biophys. Acta - Gen. Subj.* **1861**, 664–672 (2017).

51. Azinas, S. *et al.* Membrane-containing virus particles exhibit the mechanics of a composite material for genome protection. *Nanoscale* **10**, 7769–7779 (2018).

52. G, S. Das myxomatogene Virus. Beitrag zum Stadium der Krankheit-serreger ausserhalbdes Sichtbaren. *Zbl. Bakt.* **23**, 865–873. (1898).

53. Stass, R., Ng, W. M., Kim, Y. C. & Huiskonen, J. T. *Structures of enveloped virions determined by cryogenic electron microscopy and tomography*. *Virus Structure and Function* (Elsevier Inc., 2019). doi:10.1016/bs.aivir.2019.07.009

54. Dubuisson, J., Helle, F. & Cocquerel, L. Early steps of the hepatitis C

virus life cycle. *Cell. Microbiol.* **10**, 821–827 (2008).

55. Stone, N. P., Demo, G., Agnello, E. & Kelch, B. A. Principles for enhancing virus capsid capacity and stability from a thermophilic virus capsid structure. *Nat. Commun.* **10**, 1–13 (2019).

56. Rochal, S. B., Konevtsova, O. V. & Lorman, V. L. Static and dynamic hidden symmetries of icosahedral viral capsids. *Nanoscale* **9**, 12449–12460 (2017).

57. Prasad, B.V. Venkataram ; Schmid, M. F. Principles of Virus Structural Organization. *Adv Exp Med Biol* **726**, 17-47. (2012).

58. Mattei, S., Schur, F. K. & Briggs, J. A. Retrovirus maturation - An extraordinary structural transformation. *Curr. Opin. Virol.* **18**, 27–35 (2016).

59. Baker, T. S., Olson, N. H. & Fuller, S. D. Adding the Third Dimension to Virus Life Cycles: Three-Dimensional Reconstruction of Icosahedral Viruses from Cryo-Electron Micrographs. *Microbiol. Mol. Biol. Rev.* **64**, 237–237 (2000).

60. Parent, K. N., Schrad, J. R. & Cingolani, G. Breaking symmetry in viral icosahedral capsids as seen through the lenses of X-ray crystallography and cryo-electron microscopy. *Viruses* **10**, 1–22 (2018).

61. Lucas, W. Viral Capsids and Envelopes: Structure and Function. *Encycl. Life Sci.* 1–7 (2010). doi:10.1002/9780470015902.a0001091.pub2

62. Fokine, A. & Rossmann, M. G. Molecular architecture of tailed double-stranded DNA phages. *Bacteriophage* **4**, e28281 (2014).

63. Baker, M. L., Jiang, W., Rixon, F. J. & Chiu, W. Common Ancestry of Herpesviruses and Tailed DNA Bacteriophages. *J. Virol.* **79**, 14967–14970 (2005).

64. Laanto, E. *et al.* Virus found in a boreal lake links ssDNA and dsDNA viruses. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 8378–8383 (2017).

65. Peralta, B. *et al.* Mechanism of Membranous Tunnelling Nanotube Formation in Viral Genome Delivery. *PLoS Biol.* **11**, (2013).

66. Wang, N. *et al.* Architecture of African swine fever virus and implications for viral assembly. **1439**, 1–10 (2019).

67. Mougari, S., Sahmi-Bounsiar, D., Levasseur, A., Colson, P. & La Scola, B. Virophages of Giant Viruses: An Update at Eleven. *Viruses* **11**, 733 (2019).

68. Anthony C. Woo, Morgan Gaia, Julien Guglielmini , Violette Da Cunha, P. F. Evolution of the PRD1-adenovirus lineage: a viral tree of life incongruent with the cellular 3 universal tree of life. *bioRxiv* 1–16 (2019).

69. San Martín, C. & Van Raaij, M. J. The so far farthest reaches of the double jelly roll capsid protein fold 06 Biological Sciences 0601 Biochemistry and Cell Biology. *Virol. J.* **15**, 1–6 (2018).

70. Demina, T. A., Pietilä, M. K. & Svirskaite, J. Conserved Elements in Icosahedral Membrane-Containing DNA. *MBio* **7**, 1–12 (2016).

71. Demina, T. A. *et al.* HCIV-1 and other tailless icosahedral internal membrane-containing viruses of the family Sphaerolipoviridae. *Viruses* **9**, (2017).

72. Jalasvuori, M., Pawlowski, A. & Bamford, J. K. H. A unique group of virus-related, genome-integrating elements found solely in the bacterial family Thermaceae and the archaeal family Halobacteriaceae. *J. Bacteriol.* **192**, 3231–3234 (2010).

73. Gil-carton, D. *et al.* Insight into the Assembly of Viruses with Vertical Single b -barrel Major Capsid Proteins Article Insight into the Assembly of Viruses with Vertical Single b -barrel Major Capsid Proteins. 1–12 (2015). doi:10.1016/j.str.2015.07.015

74. Rissanen, I. *et al.* Bacteriophage P23-77 capsid protein structures reveal the archetype of an ancient branch from a major virus lineage. *Structure* **21**, 718–726 (2013).

75. Santos-Pérez, I. *et al.* Structural basis for assembly of vertical single β-barrel viruses. *Nat. Commun.* **10**, (2019).

76. Aalto, A. P. *et al.* Snapshot of virus evolution in hypersaline environments from the characterization of a membrane-containing Salisaeta icosahedral phage 1. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 7079–7084 (2012).

77. Zhang, X. *et al.* Structure of Sputnik, a virophage, at 3.5-Å resolution. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 18431–18436 (2012).

78. Rux, J. J., Kuser, P. R. & Burnett, R. M. Structural and Phylogenetic Analysis of Adenovirus Hexons by Use of High-Resolution X-Ray

Crystallographic, Molecular Modeling, and Sequence-Based Methods. *J. Virol.* **77**, 9553–9566 (2003).

79. Abrescia, N. G. A. *et al.* Insights into assembly from structural analysis of bacteriophage PRD1. *Nature* **432**, 68–74 (2004).

80. Hyun, J. K. *et al.* Membrane remodeling by the double-barrel scaffolding protein of poxvirus. *PLoS Pathog.* **7**, (2011).

81. Krupovic, M. & Koonin, E. V. Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution. *Nat Rev Microbiol.* **13**, 105–115 (2015).

82. Atanasova, N. S., Roine, E., Oren, A., Bamford, D. H. & Oksanen, H. M. Global network of specific virus-host interactions in hypersaline environments. *Environ. Microbiol.* **14**, 426–440 (2012).

83. Jaakkola, S. T. *et al.* Closely Related Archaeal Haloarcula hispanica Icosahedral Viruses HHIV-2 and SH1 Have Nonhomologous Genes Encoding Host Recognition Functions. *J. Virol.* **86**, 4734–4742 (2012).

84. Atanasova, N. S., Demina, T. A., Buivydas, A., Bamford, D. H. & Oksanen, H. M. *Archaeal viruses multiply: Temporal screening in a solar saltern*. *Viruses* **7**, (2015).

85. Marrugal-Lorenzo, J. A., Serna-Gallego, A., Berastegui-Cabrera, J., Pachón, J. & Sánchez-Céspedes, J. Repositioning salicylanilide anthelmintic drugs to treat adenovirus infections. *Sci. Rep.* **9**, 1–10 (2019).

86. Lee, C. S. *et al.* Adenovirus-mediated gene delivery: Potential applications for gene and cell-based therapies in the new era of personalized medicine. *Genes Dis.* **4**, 43–63 (2017).

87. Ojala, V., Laitalainen, J. & Jalasvuori, M. Fight evolution with evolution: Plasmid-dependent phages with a wide host range prevent the spread of antibiotic resistance. *Evol. Appl.* **6**, 925–932 (2013).

88. Pawlowski, A. *et al.* The Minor Capsid Protein VP11 of Thermophilic Bacteriophage P23-77 Facilitates Virus Assembly by Using Lipid-Protein Interactions. *J. Virol.* **89**, 7593–7603 (2015).

89. Simmonds, P. *et al.* ICTV virus taxonomy profile: Flaviviridae. *J. Gen. Virol.* **98**, 2–3 (2017).

90. Shi, M. *et al.* Divergent Viruses Discovered in Arthropods and Vertebrates Revise the Evolutionary History of the Flaviviridae and

Related Viruses. *J. Virol.* **90**, 659–669 (2016).

91. Sirohi, D. *et al.* The cryo-em structure of zika virus. *Am. J. Trop. Med. Hyg.* **95 (5 Supp**, 571 (2016).

92. Zhang, X. *et al.* Cryo-EM structure of the mature dengue virus at 3.5-Å resolution. *Nat. Struct. Mol. Biol.* **20**, 105–110 (2013).

93. Zhang, W., Kaufmann, B., Chipman, P. R., Kuhn, R. J. & Rossmann, M. G. Membrane curvature in flaviviruses. *J. Struct. Biol.* **183**, 86–94 (2013).

94. Buckwold, V. E., Beer, B. E. & Donis, R. O. Bovine viral diarrhea virus as a surrogate model of hepatitis C virus for the evaluation of antiviral agents. *Antiviral Res.* **60**, 1–15 (2003).

95. Wang, J.; Li, Y.; Modis, Y. Structural models of the membrane anchors of envelope glycoproteins E1 and E2 from pestiviruses. *Virology* **454–4555**, 93–101 (2014).

96. Tong, Y., Lavillette, D., Li, Q. & Zhong, J. Role of hepatitis C virus envelope glycoprotein E1 in virus entry and assembly. *Front. Immunol.* **9**, 1–7 (2018).

97. Kong, L. *et al.* NIH Public Access. **342**, 1090–1094 (2014).

98. Li, Y. & Modis, Y. A novel membrane fusion protein family in Flaviviridae? *Trends Microbiol.* **22**, 176–182 (2014).

99. Weiskircher, E., Aligo, J., Ning, G. & Konan, K. V. Bovine viral diarrhea virus NS4B protein is an integral membrane protein associated with Golgi markers and rearranged host membranes. *Virol. J.* **6**, 1–15 (2009).

100. Lai, V. C. H. *et al.* Generation and Characterization of a Hepatitis C Virus NS3 Protease-Dependent Bovine Viral Diarrhea Virus. *J. Virol.* **74**, 6339–6347 (2000).

101. Zhang, X. *et al.* Structures and functions of the envelope glycoprotein in flavivirus infections. *Viruses* **9**, 1–14 (2017).

102. Kuhn, R. J. *et al.* Structure of dengue virus: Implications for flavivirus organization, maturation, and fusion. *Cell* **108**, 717–725 (2002).

103. Twarock, R. & Luque, A. Structural puzzles in virology solved with an overarching icosahedral design principle. *Nat. Commun.* **10**, 1–9 (2019).

104. Callens, N. *et al.* Morphology and Molecular Composition of Purified Bovine Viral Diarrhea Virus Envelope. *PLoS Pathog.* **12**, (2016).

105. Reddy, T. & Sansom, M. S. P. The Role of the Membrane in the Structure and Biophysical Robustness of the Dengue Virion Envelope. *Structure* **24**, 375–382 (2016).

106. Childs, T. X Disease of Cattle - Saskatchewan. *Can. J. Comp. Med. Vet. Sci.* **10**, 316–9 (1946).

107. Olafson, P; MacCallum; Fox, F. H. An apparently new. *Cornell Vet.* **36**, 205–13 (1946).

108. Ramsey, F. K. et Chivers, W. H. Mucosal disease of cattle. *North Am. Vet.* **34**, 629–633 (1953).

109. Baker, J. C. Bovine viral diarrhea virus: A review. *Am. Vet. Med. Assoc.* **190**, 1449–58 (1987).

110. Collett, M. S., Larson, R., Belzer, S. K. & Retzel, E. Proteins encoded by bovine viral diarrhea virus: The genomic organization of a pestivirus. *Virology* **165**, 200–208 (1988).

111. Larson, R., Belzer, S. K. & Retzelt, E. Proteins Encoded by Bovine Viral Diarrhea Virus: The Genomic. *Organization* **208**, (1988).

112. C. M. Fauquet (auth.), Dr. R. I. B. Francki, Dr. C. M. Fauquet, Dr. D. L. Knudson, D. F. B. (eds. . Classification and Nomenclature of Viruses: Fifth Report of the International Committee on Taxonomy of Viruses. Virology Division of the International Union of Microbiological Societies. *Springer-Verlag Wien* (1991).

113. Charles Pellerin, Jan van Den Hurk, Jaqueline Lecomte, P. T. 1-s2.0-S0042682284714838-main.pdf. (1994).

114. Fulton, R. W. *et al.* Bovine viral diarrhoea virus (BVDV) subgenotypes in diagnostic laboratory accessions: Distribution of BVDV1a, 1b, and 2a subgenotypes. *Vet. Microbiol.* **111**, 35–40 (2005).

115. Khodakaram-Tafti, A. & Farjanikish, G. H. Persistent bovine viral diarrhea virus (BVDV) infection in cattle herds. *Iran. J. Vet. Res.* **18**, 154–163 (2017).

116. Patil, V., Taori, K. B., Rathi, V., Medsinge, A. & Singh, A. Antenatal diagnosis of pericardial teratoma. *Ultrasound International* **10**, 49–52 (2004).

117. Giangaspero, M. *et al.* Serological and antigenical findings indicating pestivirus in man. 53–62 (1993). doi:10.1007/978-3-7091-9300-6_5

118. Giangaspero, M. Zika Virus-Bovine Viral Diarrhea Virus Potential Interaction in the Etiopathology of Birth Defects in Brazil, a One Health Issue? *Community Med. Public Heal. Care* **6**, 1–4 (2019).

119. Uryvaev, L. V. *et al.* Contamination of cell cultures with bovine viral diarrhea virus (BVDV). *Bull. Exp. Biol. Med.* **153**, 77–81 (2012).

120. Richter, V. *et al.* Mapping the global prevalence of bovine viral diarrhoea virus infection and its associated mitigation programmes. *Vet. Rec.* **184**, 711 (2019).

121. Scharnböck, B. *et al.* A meta-analysis of bovine viral diarrhoea virus (BVDV) prevalences in the global cattle population. *Sci. Rep.* **8**, 1–15 (2018).

122. Ran, X. *et al.* A systematic review and meta-analysis of the epidemiology of bovine viral diarrhea virus (BVDV) infection in dairy cattle in China. *Acta Trop.* **190**, 296–303 (2019).

123. GILLESPIE JH, BAKER JA, M. K. A cytopathogenic strain of virus diarrhea virus. *Cornell Vet.* **50**, 73–79 (1960).

124. Birk, A. V., Dubovi, E. J., Cohen-Gould, L., Donis, R. & Szeto, H. H. Cytoplasmic vacuolization responses to cytopathic bovine viral diarrhoea virus. *Virus Res.* **132**, 76–85 (2008).

125. Peterhans, E., Bachofen, C., Stalder, H. & Schweizer, M. Cytopathic bovine viral diarrhea viruses (BVDV): Emerging pestiviruses doomed to extinction. *Vet. Res.* **41**, (2010).

126. Pestivirus virion morphogenesis in the absence of uncleaved nonstructural protein 2-3. *J. Virol.* **86**, 427–437 (2012).

127. Scruggs, D. W., Fleming, S. A., Maslin, W. R. & Wayne, G. A. Osteopetrosis, Anemia, Thrombocytopenia, and Marrow Necrosis in Beef Calves Naturally Infected with Bovine Virus Diarrhea Virus. *J. Vet. Diagnostic Investig.* **7**, 555–559 (1995).

128. GIANGASPERO, M., HARASAWA, R., WEBER, L. & BELLOLI, A. Genoepidemiological Evaluation of Bovine viral diarrhea virus 2 Species Based on Secondary Structures in the 5' Untranslated Region. *J. Vet. Med. Sci.* **70**, 571–580 (2008).

129. Bollini, M., Leal, E. S., Adler, N. S., Aucar, M. G. & Cavasotto, C. N. Discovery of Novel Bovine Viral Diarrhea Inhibitors Using

Structure-Based Virtual Screening on the Envelope Protein E2. **6**, 1–10 (2018).

130. Cai, D. *et al.* Enhanced immune responses to E2 protein and DNA formulated with ISA 61 VG administered as a DNA prime – protein boost regimen against bovine viral diarrhea virus. *Vaccine* (2018). doi:10.1016/j.vaccine.2018.07.054

131. Pascual, M. J. *et al.* Structure-based drug design for envelope protein E2 uncovers a new class of bovine viral diarrhea inhibitors that block virus entry. *Antiviral Res.* (2017). doi:10.1016/j.antiviral.2017.10.010

132. Lokhandwala, S. *et al.* Priming cross-protective bovine viral diarrhea virus-specific immunity using live-vectored mosaic antigens. *PLoS One* **12**, 1–23 (2017).

133. Hanon, J. B. *et al.* Evaluation of 16 commercial antibody ELISAs for the detection of bovine viral diarrhea virus–specific antibodies in serum and milk using well-characterized sample panels. *J. Vet. Diagnostic Investig.* **29**, 833–843 (2017).

134. Saged, H. Comparative Study between Serological and Molecular Methods for Diagnosis Bovine Viral Diarrhea Virus. *Int. J. Mol. Vet. Res.* 0–5 (2012). doi:10.5376/ijmvr.2012.02.0001

135. Moorthy, D., Mishra, N., Kalaiyarasu, S., Jhade, S. K. & Singh, V. P. Evaluation of currently available bovine viral diarrhoea virus (BVDV) and HoBi-like pestivirus (HoBiPeV) specific diagnostic tests in detection of highly divergent HoBiPeVs in cattle. *J. Virol. Methods* **272**, 113707 (2019).

136. Ahmad, A. Diagnostic Approaches for Detection of Bovine Viral Diarrhea Virus Persistent Infection. *J. Infect. Mol. Biol.* **2**, 53–60 (2014).

137. Stahl, K. & Alenius, S. BVDV control and eradication in Europe. *Jpn J Vet Res* **60**, 31–39 (2012).

138. Niskanen, R. & Lindberg, A. Transmission of bovine viral diarrhoea virus by unhygienic vaccination procedures, ambient air, and from contaminated pens. *Vet. J.* **165**, 125–130 (2003).

139. Wang, F. I., Deng, M. C., Huang, Y. L. & Chang, C. Y. Structures and functions of pestivirus glycoproteins: Not simply surface matters. *Viruses* **7**, 3506–3529 (2015).

140. Jordan, R. *et al.* Inhibition of host ER glucosidase activity prevents

Golgi processing of virion-associated bovine viral diarrhea virus E2 glycoproteins and reduces infectivity of secreted virions. *Virology* **295**, 10–19 (2002).

141. Murray, C. L., Marcotrigiano, J. & Rice, C. M. Bovine Viral Diarrhea Virus Core Is an Intrinsically Disordered Protein That Binds RNA. *J. Virol.* **82**, 1294–1304 (2008).

142. Krey, T. *et al.* Crystal structure of the pestivirus envelope glycoprotein Erns and mechanistic analysis of its ribonuclease activity. *Structure* **20**, 862–873 (2012).

143. Aberle, D. *et al.* Structure of the Membrane Anchor of Pestivirus Glycoprotein Erns, a Long Tilted Amphipathic Helix. *PLoS Pathog.* **10**, (2014).

144. Tautz, N., Tews, B. A. & Meyers, G. *The Molecular Biology of Pestiviruses*. *Advances in Virus Research* **93**, (Elsevier Inc., 2015).

145. El Omari, K., Iourin, O., Harlos, K., Grimes, J. M. & Stuart, D. I. Structure of a Pestivirus Envelope Glycoprotein E2 Clarifies Its Role in Cell Entry. *Cell Rep.* **3**, 30–35 (2013).

146. Schirrmeier, H., Strebelow, G., Depner, K., Hoffmann, B. & Beer, M. Genetic and antigenic characterization of an atypical pestivirus isolate, a putative member of a novel pestivirus species. *J. Gen. Virol.* **85**, 3647–3652 (2004).

147. Silveira, S. *et al.* HoBi-like is the most prevalent ruminant pestivirus in Northeastern Brazil. *Transbound. Emerg. Dis.* **65**, e113–e120 (2018).

148. Giammarioli, M. *et al.* Genetic detection and characterization of emerging HoBi-like viruses in archival foetal bovine serum batches. *Biologicals* **43**, 220–224 (2015).

149. Richter, V. *et al.* A systematic worldwide review of the direct monetary losses in cattle due to bovine viral diarrhoea virus infection. *Vet. J.* **220**, 80–87 (2017).

150. Thomann, B. *et al.* Economic evaluation of the eradication program for bovine viral diarrhea in the Swiss dairy sector. *Prev. Vet. Med.* **145**, 1–6 (2017).

151. Food and Agriculture Organization of the United Nations. Statistical yearbook of the Food and Agricultural Organization for the United Nations. *Http://Www.Fao.Org/Docrep/017/I3138E/I3138E07.Pdf*

(2014).

152. Marschik, T. *et al.* A cost-benefit analysis and the potential trade effects of the bovine viral diarrhoea eradication programme in Styria, Austria. *Vet. J.* **231**, 19–29 (2018).

153. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).

154. Richter, M., Reimann, I., Schirrmeier, H., Kirkland, P. D. & Beer, M. The viral envelope is not sufficient to transfer the unique broad cell tropism of Bungowannah virus to a related pestivirus. *J. Gen. Virol.* **95**, 2216–2222 (2014).

155. Clough, R. N., Moldovan, G. & Kirkland, A. I. Direct detectors for electron microscopy. *J. Phys. Conf. Ser.* **522**, (2014).

156. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 (2013).

157. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).

158. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

159. Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).

160. Scheres, S. H. W. Maximum-likelihood methods in cryo-EM. PartII: application to experimantal data. *Methods Enzym.* **6879**, 295–320 (2010).

161. Ilca, S. L. *et al.* Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. *Nat. Commun.* **6**, 1–8 (2015).

162. Heymann, J. B. & Belnap, D. M. Bsoft: Image processing and molecular modeling for electron microscopy. *J. Struct. Biol.* **157**, 3–18 (2007).

163. Zheng, S. Q. *et al.* MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat.*

Methods **14**, 331–332 (2017).

164. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer Visualization of Three-Dimensional Image Data Using IMOD. *J. Struct. Biol.* **116**, 71–6 (1996).

165. O'Toole, E., van der Heide, P., Richard McIntosh, J. & Mastronarde, D. Large-Scale Electron Tomography of Cells Using SerialEM and IMOD. 95–116 (2018). doi:10.1007/978-3-319-68997-5_4

166. Radermacher, M. Weighted Back-Projection Methods. 91–115 (1992).

167. Gilbert, P. F. The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. II. Direct methods. *Proc. R. Soc. London. Ser. B. Biol. Sci.* **182**, 89–102 (1972).

168. Turoňová, B., Schur, F. K. M., Wan, W. & Briggs, J. A. G. Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4 Å. *J. Struct. Biol.* **199**, 187–195 (2017).

169. Danev, R. & Baumeister, W. Expanding the boundaries of cryo-EM with phase plates. *Curr. Opin. Struct. Biol.* **46**, 87–94 (2017).

170. Danev, R., Buijsse, B., Khoshouei, M., Plitzko, J. M. & Baumeister, W. Volta potential phase plate for in-focus phase contrast transmission electron microscopy. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15635–15640 (2014).

171. Castaño-Díez, D., Kudryashev, M., Arheit, M. & Stahlberg, H. Dynamo: A flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments. *J. Struct. Biol.* **178**, 139–151 (2012).

172. Castaño-Díez, D., Kudryashev, M. & Stahlberg, H. Dynamo Catalogue: Geometrical tools and data management for particle picking in subtomogram averaging of cryo-electron tomograms. *J. Struct. Biol.* **197**, 135–144 (2017).

173. Wan, W. & Briggs, J. A. G. *Cryo-Electron Tomography and Subtomogram Averaging*. *Methods in Enzymology* **579**, (Elsevier Inc., 2016).

174. Leigh, K. E. *et al.* Subtomogram averaging from cryo-electron tomograms. *Methods Cell Biol.* **152**, 217–259 (2019).

175. Gaspari, M. & Cuda, G. Nano LC–MS/MS: A Robust Setup for

Proteomic Analysis. *Nanoproteomics* **790**, 115–126 (2011).

176. Fiala, G. J., Schamel, W. W. A. & Blumenthal, B. Blue native polyacrylamide gel electrophoresis (BN-PAGE) for analysis of multiprotein complexes from cellular lysates. *J. Vis. Exp.* 1–7 (2010). doi:10.3791/2164

177. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).

178. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

179. Adams, P. D. *et al.* The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94–106 (2011).

180. Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 531–544 (2018).

181. Wang, R. Y. R. *et al.* Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. *Elife* **5**, 1–22 (2016).

182. Yang, J. *et al.* The I-TASSER suite: Protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2014).

183. Luengo, I. *et al.* SuRVoS: Super-Region Volume Segmentation workbench. *J. Struct. Biol.* **198**, 43–53 (2017).

184. Ang, K. S. & Schaposnik, L. P. On the geometry of regular icosahedral capsids containing disymmetrons. *J. Struct. Biol.* **197**, 340–349 (2017).

185. Colibus, L. De *et al.* Assembly of complex viruses exemplified by a halophilic euryarchaeal virus. *Nat. Commun.* **10**, 1–9 (2019).

186. Jaatinen, S. T., Happonen, L. J., Laurinmäki, P., Butcher, S. J. & Bamford, D. H. Biochemical and structural characterisation of membrane-containing icosahedral dsDNA bacteriophages infecting thermophilic Thermus thermophilus. *Virology* **379**, 10–19 (2008).

187. Kivelä, H. M. *et al.* Quantitative dissociation of archaeal virus SH1 reveals distinct capsid proteins and a lipid core. *Virology* **356**, 4–11 (2006).

188. Veesler, D. *et al.* Atomic structure of the 75 MDa extremophile Sulfolobus turreted icosahedral virus determined by CryoEM and X-ray crystallography. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 5504–5509 (2013).

189. Paul D. Adams, Isaiah T. Arkin, D. M. E. & A. T. B. Computational searching and mutagenesis suggest a structure for the pentameric transmembrane domain of phospholamban. *Nat. Struct. Biol.* **2**, 154–162 (1995).

190. Vostrikov, V. V., Mote, K. R., Verardi, R. & Veglia, G. Structural dynamics and topology of phosphorylated phospholamban homopentamer reveal its role in the regulation of calcium transport. *Structure* **21**, 2119–2130 (2013).

191. Jaalinoja, H. T. *et al.* Structure and host-cell interaction of SH1, a membrane-containing, halophilic euryarchaeal virus. *Proc. Natl. Acad. Sci.* **105**, 8008–8013 (2008).

192. Benson, S. D., Bamford, J. K. H., Bamford, D. H. & Burnett, R. M. Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. *Cell* **98**, 825–833 (1999).

193. Reddy, H. K. N., Carroni, M., Hajdu, J. & Svenda, M. Electron CRYO-microscopy of bacteriophage PR772 reveals the elusive vertex complex and the CAPSID architecture. *Elife* **8**, 1–62 (2019).

194. Flatt, J. W. & Butcher, S. J. Adenovirus flow in host cell networks. *Open Biol.* **9**, (2019).

195. *Viral Molecular Machines*. *Advances in Experimental Medicine and Biology* (Springer-Verlag New York, 2012).

196. Crisostomo, L., Soriano, A. M., Mendez, M., Graves, D. & Pelka, P. Temporal dynamics of adenovirus 5 gene expression in normal human cells. *PLoS One* **14**, 1–18 (2019).

197. Leigh, B. A., Breitbart, M., Oksanen, H. M., Bamford, D. H. & Dishawb, L. J. crossm a Pseudoalteromonas sp . Isolated from the Gut of Ciona. *genome Announc.* **6**, (2018).

198. Goulet, A. *et al.* Acidianus filamentous virus 1 coat proteins display a helical fold spanning the filamentous archaeal viruses lineage. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 21155–21160 (2009).

199. Liu, H. *et al.* Atomic structure of human adenovirus by Cryo-EM reveals interactions among protein networks. *Science (80-.).* **329**,

1038–1043 (2010).

200. Prangishvili, D. *et al.* The enigmatic archaeal virosphere. *Nat. Rev. Microbiol.* **15**, 724–739 (2017).

201. Golomidova, A. K. *et al.* Branched lateral tail fiber organization in T5-like bacteriophages DT57C and DT571/2 is revealed by genetic and functional analysis. *Viruses* **8**, 1–21 (2016).

202. Harris, A. *et al.* Influenza virus pleiomorphy characterized by cryoelectron tomography. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 19123–19127 (2006).

203. Mei, K. *et al.* Cryo-EM structure of the exocyst complex. *Nat. Struct. Mol. Biol.* **25**, 139–146 (2018).