Structural characterization of the Human Endogenous Retrovirus and SARS-CoV-2

Anna-Sophia Krebs

Lincoln College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy in *Cellular Structural Biology*
Hilary term 2023

Abstract

Retroviruses, a type of RNA virus, have a unique life cycle wherein they reverse transcribe their viral RNA genome and integrate as a provirus into the host's DNA. When retroviruses integrate into the germ line of humans, they can form human endogenous retroviruses (HERVs), which make up approximately 8% of the human genome. HERV-K (HML-2) is the most recently acquired HERV but mutations have rendered it mostly silent except in many cancers, neurological diseases, and during the aging process.

To better understand HERV-K, I expressed its main structural protein Gag and determined its structure in the immature CA lattice of virus-like particles (VLPs). HERV-K VLPs have a greater distance between the membrane and the immature capsid (CA) lattice than other retroviruses, which correlates with the presence of small proteins SP1 and p15 between CA and MA. Using cryo-electron tomography (cryoET) and subtomogram averaging (STA), we resolved the immature lattice of HERV-K CA to 3.2 Å. The resulting map allowed me to build a hexameric atomic model with most side chains well resolved. I confirmed the trimeric and dimeric interactions between hexamers using site-directed mutagenesis and molecular dynamics simulations. A large conformational change in the lattice must occur during maturation, which is driven by the linker between the N-terminal and C-terminal domain, similar to HIV-1. A structure-based phylogenetic analysis with corresponding CA structures of other retroviruses revealed highly conserved assembly patterns across genera and evolutionary time.

SARS-CoV-2, another RNA virus, caused a worldwide pandemic in 2020. We analyzed the structure of the spike (S) protein derived from the ChAdOx1 nCoV-19 vaccine using subtomogram averaging. We found S to be in its prefusion conformation, which can elicit an immune response for protection. We also used a multimodal, multiscalar approach of frozen, infected cells to study assembly and egress of SARS-CoV-2. Our approach can now be applied to a wide range of viruses to understand different viral life cycles.

Acknowledgements

It is not possible to finish a DPhil without some scientific and emotional help. I would thus like to make the space here to thank all the people that were supporting me during this time.

This work would not have come to be without the help of my primary supervisor Peijun Zhang. Her input, expertise and access to microscope time made it possible for me to learn so much during my years in Oxford. I would also like to thank my secondary supervisor Thomas Bowden for being so kind and helpful to me and Juan, Margaret and Weixian for keeping the lab running. A big thank-you also goes to all the current and former members of the Zhang lab. You guys not only taught me most of what I know about tomography, but also helped to keep my sanity with non-lab focused fun.

I would have not been able to finish my DPhil without the help of my family. They were always up for a Skype call after a stressful week and especially during various pandemic-induced lockdowns, my parents gave me a place to stay to get a change of scenery for a little bit. Sebastian helped raise my spirits during holidays on the ski slopes and my aunties and grandmothers were always happy to chat or go for a walk.

A massive thank-you also goes to Niall, without whom I would have given up at least twice a week. He was there for me during the difficult times when nothing wanted to work and I came home dejected. But he (and the gang) also celebrated

the good times with me and kept my spirits high.

I also want to thank all my friends here in Oxford, back home, and wherever in the world they have ended up, for being there for me, getting my mind off things and having so much fun together. All of you made my time here so much more enjoyable.

Contents

List of Figures

List of Tables

List of Abbreviations

aa	Amino acid
ACE2	Angiotensin-converting enzyme 2
AIDS	Acquired immune deficiency syndrome
ALS	Amyotrophic lateral sclerosis
BVM	Bevirimat
CA	Capsid
CCD	Catalytic core domain
Covid-19	Coronavirus disease 2019
cryoEM	Cryo electron microscopy
cryoET	Cryo electron tomography
cryoFIB/SEM	Cryo focused ion-beam scanning electron microscopy
CSPT	Constrained single-particle tomography
CTD	C-terminal domain
CTF	Contrast transfer function
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DPBS	Dulbecco's phosphate-buffered saline
DMV	Double membrane vesicle
dNTP	Deoxyribonucleotide triphosphate

dTTP	Deoxythymidine triphosphate
dUMP	Deoxyuridine monophosphate
dUTP	Deoxyuridine Triphosphate
E	Envelope
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EER	Electron-event representation
EIAV	Equine infectious anemia virus
EM	Electron microscope
EMDB	Electron Microscopy Data Bank
ER	Endoplasmic reticulum
ERGIC	ER-Golgi intermediate compartment
ERV	Endogenous retrovirus
ESCRT	Endosomal sorting complex required for transport
FCS	Fetal calf serum
FIV	Feline Immunodeficiency Virus
FP	Fusion peptide
FSC	Fourier shell correlation
6HB	6 helix bundle
HBR	Highly basic region
HEK	Human Embryonic Kidney
HERV	Human endogenous retrovirus
HIV	Human Immunodeficiency Virus
HML	Human mouse mammary tumor virus-like

HRP	Horseradish peroxidase
HTLV	Human T-lymphotropic virus
IAV	Influenza A virus
ICTV	International Committee on Taxonomy of Viruses
IN	Integrase
IP6	Inositol hexaphosphate
LB	Luria Broth
LINE	Long interspersed nuclear element
LTR	Long terminal repeat
M	Membrane
MA	Matrix
MERV	Mouse endogenous retrovirus
MES	2-(N-morpholino)ethanesulfonic acid
MD	Molecular dynamic
MHR	Major homology region
MLV	Murine leukemia virus
MMTV	Mouse mammary tumor virus
mRNA	Messenger RNA
MoMLV	Moloney murine leukemia virus
M-PMV	Mason–Pfizer monkey virus
N	Nucleocapsid
NC	Nucleocapsid
NEAA	Non-essential amino acid
NES	Nuclear export signal

NLS	Nuclear location signal
NMR	Nuclear magnetic resonance
nsp	Non-structural protein
NTD	N-terminal domain
ORF	Open reading frame
PBS	Primer-binding site
PBS	Phosphate buffered saline
PBST	Phosphate buffered saline with Tween20
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PERV	Porcine endogenous retrovirus
PF96	PF46396
PIC	Preintegration complex
PLZF	Promyelocytic zinc finger
PM	Plasma membrane
PPT	Polypurine tract
PR	Protease
PTN	Pleiotrophin
RBD	Receptor-binding domain
RCSB	Research Collaboratory for Structural Bioinformatics
RMSF	Root mean square fluctuation
RNA	Ribonucleic acid
RNP	ribonucleoprotein complex
RSV	Respiratory Syncytial Virus

RT	Reverse transcriptase
S	Spike
SA	Splice acceptor
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SINE	Short interspersed nuclear elements
SBPA	Structure-based phylogenetic analysis
SD	Splice donor
SDS	Sodium dodecyl-sulfate
SDS-PAGE	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
SIV	Simian immunodeficiency virus
smRNA-FISH	Single-molecule RNA fluorescence in situ hybridization
SMV	Single membrane vesicle
SNR	Signal-to-noise ratio
SOC	Super Optimal broth with Catabolite repression
SPA	Single particle analysis
STA	Subtomogram averaging
SU	Surface protein
TM	Transmembrane domain
UTR	Untranslated region
vDNA	Viral DNA
VLP	Virus-like particle
VP	Vesicle Packet
vRNA	Viral RNA

Chapter 1

Retroviruses

Retroviruses are positive-sense RNA viruses which reverse transcribe their RNA and integrate into the host genome, thus their name. The most famous retrovirus is HIV-1 (Human Immunodeficiency Virus 1), the exogenous virus which causes AIDS (Acquired Immune Deficiency Syndrome). These exogenous viruses are much easier to study than endogenous retroviruses, more extensively studied and most of what we know today about endogenous viruses was first found in their exogenous counterparts. Thus, this chapter summarizes this complicated class of viruses and will mostly focus on exogenous retroviruses. As my main project during my thesis was based on an endogenous virus called HERV-K, I will try to emphasize when data on endogenous viruses is present but will otherwise direct the reader to the second chapter of this thesis, which is solely focused on my work done on human endogenous retroviruses (HERVs).

1.1 Disclaimer

This chapter provides an overview of the *Retroviridae* family. Luiza Mendonça and I wrote a review about the structural aspect of assembly and maturation of retroviruses (*1*). We both contributed equally to the writing and figures of the manuscript under the supervision of Peijun Zhang. The figures taken from our review are cited and marked accordingly.

1.2 *Retroviridae* family

The *Retroviridae* family is divided into the two subfamilies and eleven genera: *Alpharetroviruses*, *Betaretroviruses*, *Deltaretroviruses*, *Epsilonretroviruses*, *Gammaretroviruses*, *Lentiviruses* in the *Orthoretrovirinae* subfamily and *Simiispumaviruses*, *Prosimiispumaviruses*, *Felispumaviruses*, *Bovispumaviruses*, *Equispumaviruses* in the *Spumaretrovirinae* family (*2*) (Figure 1).

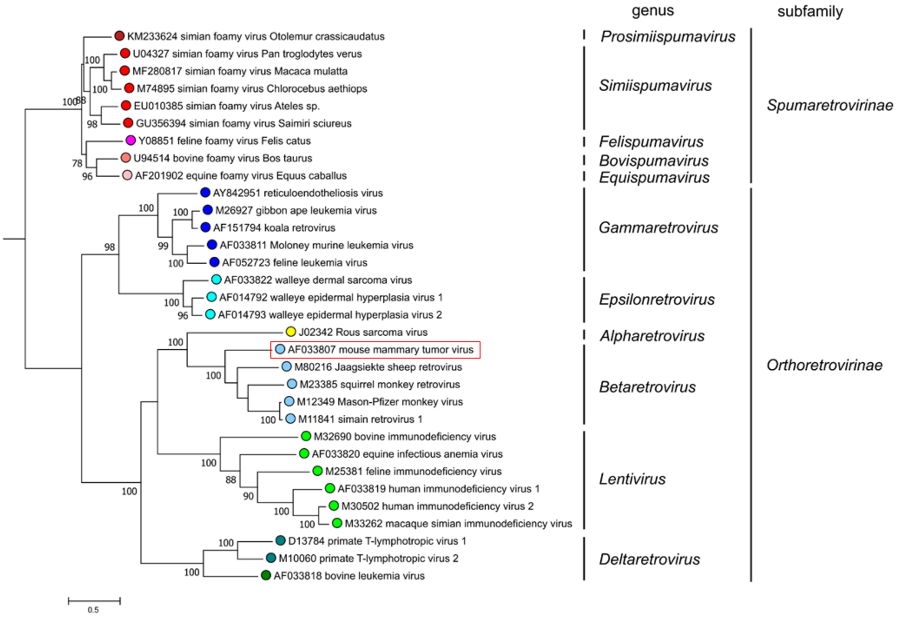


Figure 1 Phylogenetic tree of the retrovirus family The tree is based on an amino acid alignment of the reverse transcriptase and integrase genes. Not all known retroviruses are shown. This figure is taken from `https://ictv.global/report/chapter/retroviridae/retroviridae` where it is licensed under a Creative Commons Attribution-ShareAlike License (CC BY-SA 4.0).

Retroviruses share many structural traits such as a 5'LTR (long terminal repeat), PBS (primer binding site), proteins (Gag, Pro, Pol, Env), PPT (polypurine tract) and 3'LTR (*3*). Taxonomy of retroviruses was originally done based on phenotypes (*4*, *5*). However, this has proven unreliable and difficult to use. Now, classification is mostly done based on amino acid conservation of the reverse transriptase *rt* gene in *pol*, along with the presence of subdomains in the Gag protein as well as the timing of reverse-transcription in the life cycle (*3*, *6*). Narrow phylogenies in specific genera can also classified using the *Env* genes, the presence of accessory proteins and host associations (*2*, *7*).

1.3 Genome structure and retroviral proteins

Retroviruses have a minimum of three genes: Group-specific antigen (Gag), polymerase (Pol) and envelope (Env). HERV-K has an additional reading frame for protease (Pro), as have other betaretroviruses. Many retroviruses have additional proteins which are interspersed into their genome structure. They are needed to either regulate or complete part of their life cycle. The coding regions are flanked by an LTR on either side. The schematic in Figure 2 provides an overview of the HERV-K provirus.

1.3.1 Gag

Gag is the main structural protein. It is produced upon translation of the viral RNA as a long polyprotein. This polyprotein is sufficient to form non-infectious virus-like particles (VLPs) and a good target for structural studies when looking at the overall shape of a virus. During maturation Gag gets cleaved by the viral

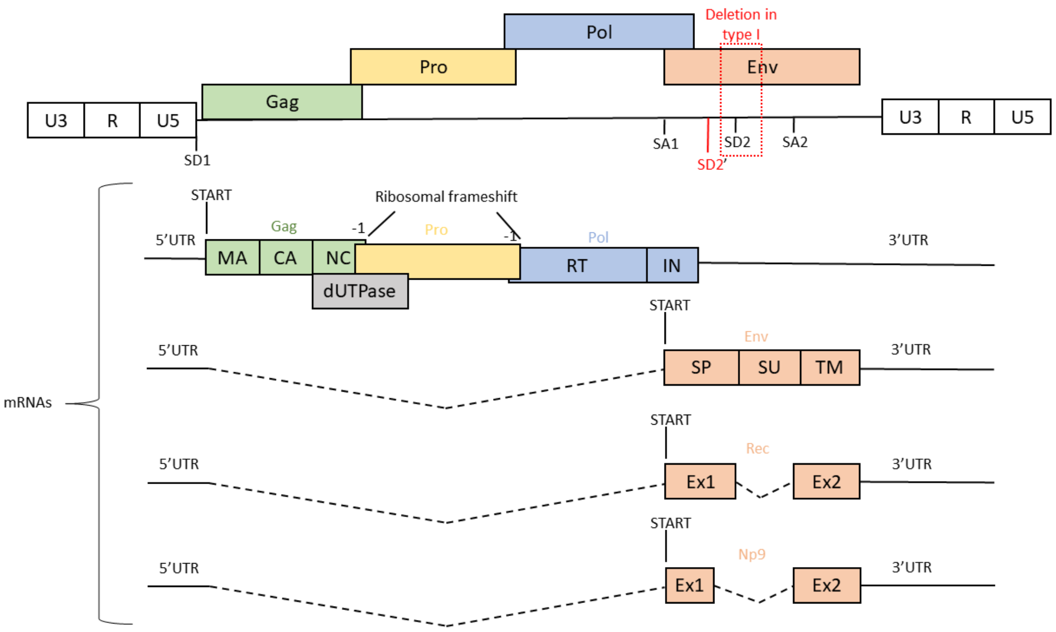


Figure 2 Proviral organization of HERV-K (HML-2) Schematic representation of an intact HERV-K (HML-2) locus with the deletion in type I proviruses marked. The four ORFs are shown in coloured boxes and overlap. Splice donor (SD) and splice acceptor (SA) sites are marked as ticks on the gene. The LTRs are shown as white boxes with their U3, R and U5 regions. The transcripts below the provirus are aligned and introns that are spliced out are marked as dashed lines.

protease into at least 3 smaller proteins: CA (capsid), MA (matrix) and NC (nucleocapsid). CA is needed to multimerize and form immature and mature virions. MA is necessary to target Gag to the assembly sites. NC binds the genomic RNA (gRNA) to load it into the virus. In some retroviruses, the cleavage of Gag by the protease additionally yields spacer peptides; in HERV-K (HML-2) these are called p15, SP1, QP1 and QP2 (*8*, *9*).

1.3.1.1 Immature Gag lattice

Immature particles of all retroviruses, except spumaviruses, have a similar overall morphology. They are made up of a dense protein shell consisting of Gag hexamers and a less dense middle holding the RNA (*10*) (Figure 3).

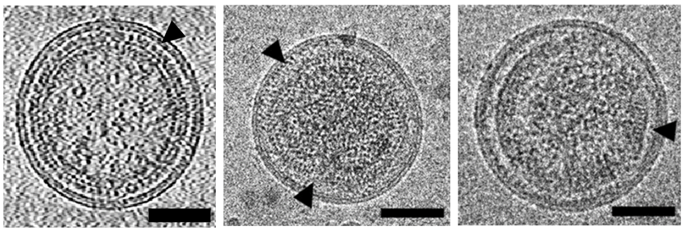


Figure 3 Retroviral immature VLPs Tomographic slice of HIV-1 (left), and cryoEM images of RSV (center) (*10*) and HTLV-1 (right) (*10*) virus-like particles, black arrowheads point to the CA lattice. Scale bars, 50 nm. This figure is taken from my paper (*1*) where it is licensed under a Creative Commons Attribution license (CC BY 4.0).

The Gag hexamers form a uniform lattice with a spacing of approximately 8 nm between hexamers (*15–18*). The lattice is held together by interactions of neighboring CA domains as well as SP1 or equivalent domains. Perfect hexamer lattices however would lead to a flat sheet. Thus, immature virions show holes and defects in their lattice to fit into a sphere (*19*). Due to these holes, resolving the structure of CA in the lattice proved very difficult to achieve. Initially, only some CA domains of different retroviruses were solved by X-ray crystallography and NMR (HIV-1: (*20–22*), RSV: (*23*, *24*)). Then the structure of HIV-1 CA C-terminal

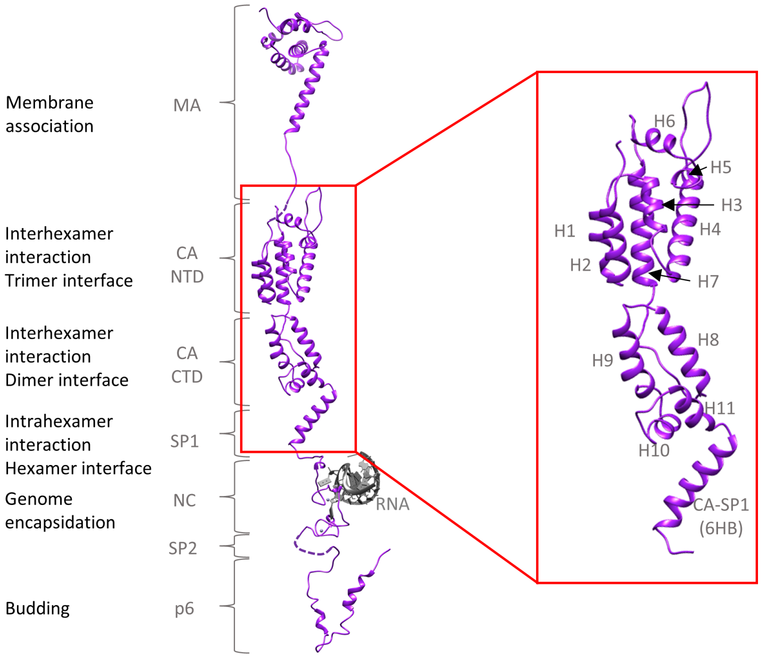


Figure 4 HIV-1 immature Gag domains HIV-1 Gag polyprotein domain organization with all CA helices labeled in the red box. Original PDB accession codes: 1UPH (MA) (*11*), 7ASH (CA and SP1) (*12*), 1F6U (NC and SL2 RNA loop) (*13*), 2C55(p6) (*14*). This figure is adapted from my paper (*1*) where it is licensed under a Creative Commons Attribution license (CC BY 4.0).

domain and its spacer peptide SP1 could be solved by X-ray crystallography (*25*). However, CA in its native environment could only be studied with the help of cryoET and subtomogram averaging (STA) and at last a resolution of 3 Å was achieved using this technique (*12*, *18*, *26–28*). All these studies showed that CA is made up of the two subdomains NTD (N-terminal domain) and CTD (C-terminal domain) with a linker joining them together. The NTD is generally formed by seven alpha helices and the CTD by 4 helices (*29*) (Figure 4). Interestingly, this holds true for almost all retroviruses even though the sequence conservation between the different retroviruses is low (*30*, *31*) (see also Figure 26).

Six CA monomers join together to form a hexamer, which is the main building block of the immature lattice. The side view of a hexamer in a lattice is shaped like a wine glass with the cup walls formed by the NTDs, the cup bottom by the CTD, the stem by a six-helix bundle and the base by the NC/vRNA layer below (*32*) (Figure 5).

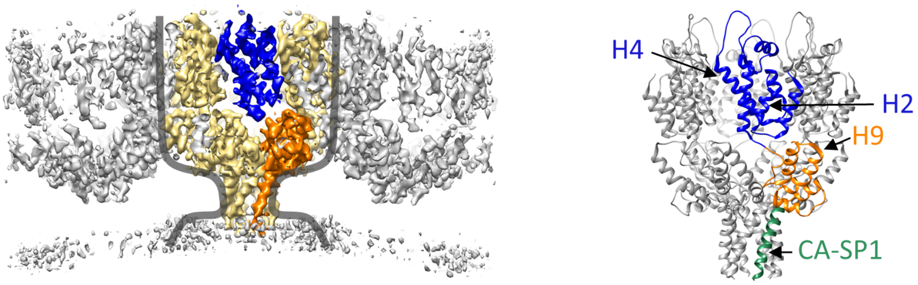


Figure 5 HIV-1 immature CA hexamer HIV-1 immature Gag structure, shown as cryoET subtomoram averaging map (left) and atomic model of the hexamer (right). One hexamer is highlighted in yellow in cryoET STA map within which one monomer is colored in blue (NTD) and orange (CTD and SP1). Grey outline shows the wine glass profile of the Gag hexamer (*12*). This figure is taken from my paper (*1*) where it is licensed under a Creative Commons Attribution license (CC BY 4.0).

The six-helix bundle (6HB) is formed by the C-terminus of CA and also by the spacer peptides downstream of CA in the viruses they exist in. The 6HB is needed to form intra-hexameric interactions to hold the hexamer together. The dimerization and trimerization of CA hexamers then gives rise to the complete CA lattice (Figure 6). These interhexameric interactions, unlike the interactions holding a hexamer together, vary between retroviruses. In HIV-1 for example, helix 2 is involved at the trimeric interface, and helix 4 interacts with helices 5 and 6 in neighbouring hexamers, helix 9 forms the dimeric interface and the major homology region (MHR) is used in intrahexameric interactions (*17*, *33*). Section 2.4.3.4 goes into further detail on how the interactions are formed, how retroviruses differ on that aspect and how HERV-K compares to them.

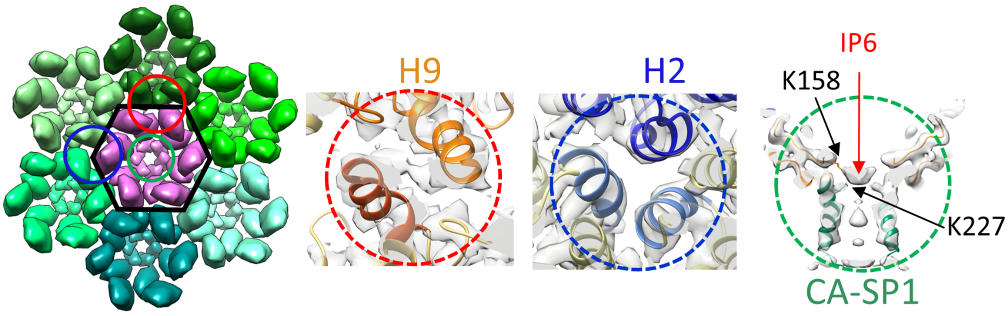


Figure 6 **HIV-1 immature CA lattice** HIV-1 immature lattice organization in top view with one hexamer marked by a black line. Details of the dimeric, trimeric and hexameric interfaces are shown in red, blue and green dashes, respectively. This figure is adapted from my paper (*1*) where it is licensed under a Creative Commons Attribution license (CC BY 4.0).

The 6HB on the underside of the lattice is a very important feature conserved in many retroviruses, having been found in HIV-1, MLV, EIAV and RSV (*12*, *34*, *35*). The MHR and the loop between helices 9 and 10 help to stabilize the 6HB. It could be shown by NMR and molecular dynamic simulations that the 6HB is important for maturation and infectivity of the virus by being in a dynamic helix-coil equilibrium (*36*). The positively charged region at the top of the 6HB binds inositol hexaphosphate (IP6), which is picked up in the host cell and acts as an assembly co-factor for the virus (*12*, *37*). The 6HB is also the site where the maturation inhibitors Bevirimat (BVM) and PF46396 (PF96) bind to treat HIV-1 infection (*38*, *39*).

While the CA part of Gag has been widely studied in immature virions and its structure resolved to high resolution *in situ*, the MA domain has remained an elusive part for a long time. Only in 2021 could this mystery finally be solved *in situ* in HIV-1 (*40*) and be compared to the trimeric crystal structure solved in 1996 (*41*). It showed that MA also forms a hexameric lattice separate from the CA lattice and its positions are not fixed relative to CA. The MA lattice covers only small patches of the inner leaflet and is poorly ordered. The lattice is formed by MA trimers with a spacing of 9.8 nm between hexamers and a large hole at the sixfold axis. These holes are lined by the inward facing highly basic region (HBR). The trimers are held together by a dimeric interface formed by interactions between the N-terminus and the top of helix 1 with themselves and the 3_{10} helix, also involving the residues needed during the myristoyl switch. MA in immature particles is in the exposed, membrane-inserted conformation (see also Maturation section 1.4.6). Helices 1 and 2 form a charged and hydrophobic surface parallel to the membrane. Helix 5 extends toward the CA lattice below.

1.3.1.2 Mature CA lattice

The mature retroviral capsid is formed upon cleavage of Gag by the viral protease (Maturation step in Figure 10) and a subset of CA monomers assemble to form the mature core (Figure 7a). This core helps to avoid detection of the genome by the host's immune system as the capsid enters into the host cell. The core is recognized by cellular proteins inside the cell to transport it through the cytoplasm towards the nucleus (*42–44*). It also provides a safe enclosure for reverse transcription and holds all the important enzymes and factors together that are needed in this important step of the viral life cycle. However, to complete the viral life cycle a virus must integrate its recently reverse transcribed vDNA into the host genome and the core must therefore easily open up when needed. To allow this event to happen, the core is a metastable structure.

The mature capsid core is mostly formed by hexamers like the immature core (Figure 7b). However, it incorporates 12 pentamers instead of holes to allow for curvature and closure of the core. Retroviral cores follow fullerene geometry models and their shape depends on how the pentamers are distributed. Lentiviruses form cones with 5 pentamers at the narrow end and seven pentamers on the broad end. Alpha-, beta-, gamma- and deltaretroviruses form cylindrical or polyhedral shapes by an even 6-6 distribution of pentamers or a random distribution respectively. There are very few retroviruses that do not follow the icosahedral geometry and incorporate more than 12 pentamers, such as MLV. MLV incorporates almost all its CA molecules with up to 24 pentamers and can have a less tightly packed core that has multiple layers or sometimes forms more than one core (*16*). HTLV-1 is an exception too as its mature cores are often incomplete, which can explain its low infectivity (*47*).

There are a few distinct differences between the immature and mature lattice. In its overall appearance HIV-1 forms a thinner mature lattice compared to the immature lattice (4 nm vs 14 nm) and the hexamers are also spaced further apart (10 nm vs 8 nm). At first glance CA monomers could almost look the same as their immature

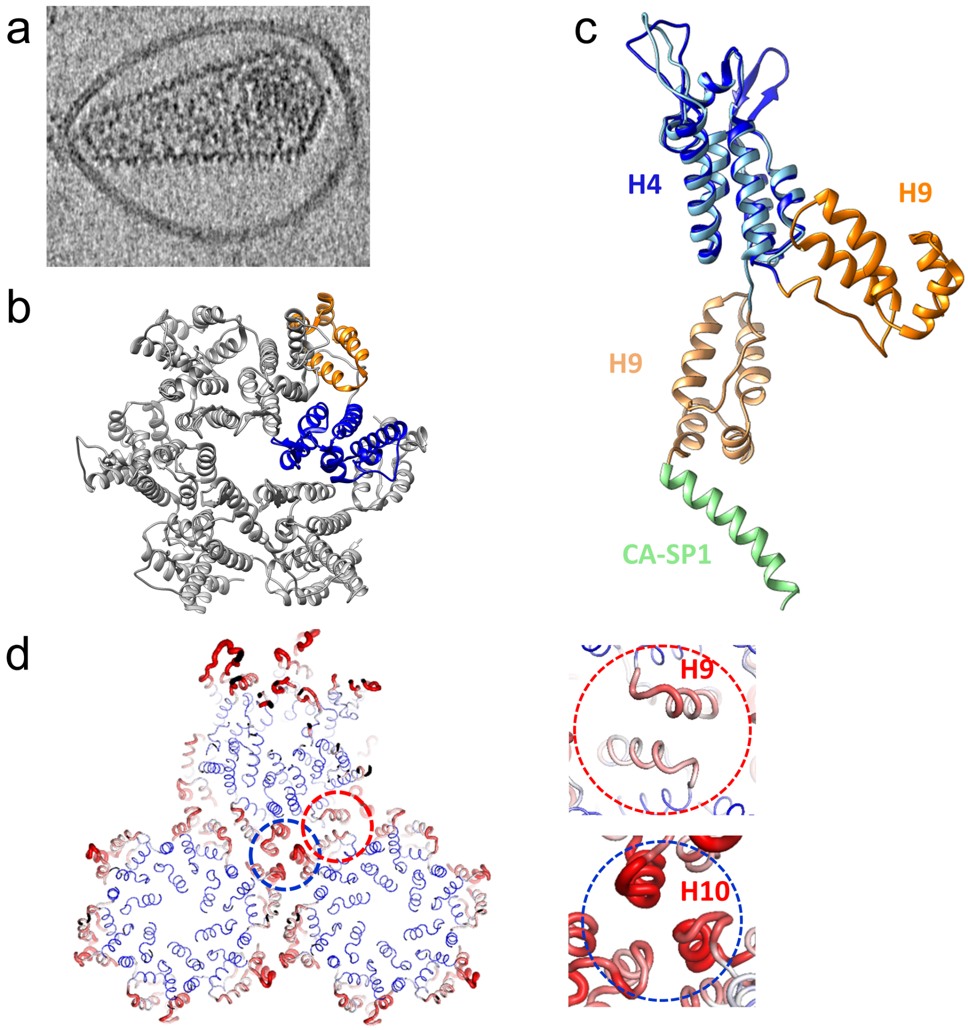


Figure 7 HIV-1 mature capsid organization (a) Tomographic slice of a mature HIV-1 particle with a capsid cone (*45*). (b) Atomic model of an HIV-1 mature CA hexamer derived from tubular assemblies (PDB 6SKK (*46*)). One CA monomer is coloured in blue (NTD) and orange (CTD). (c) Comparison of immature CA-SP1 (lighter shade) and mature CA (darker shade) structure, aligned at the NTD (blue). (d) HIV-1 mature lattice organization in top view. The width and color of the sausage are directly proportional to the B-factor, from blue (-30) to red (-100). Details of the dimeric and trimeric interfaces are shown in red and blue dashes, respectively. This figure is adapted from my paper (*1*) where it is licensed under a Creative Commons Attribution license (CC BY 4.0).

counterparts as it also forms an NTD and CTD with 7/4 helices respectively. However, the hinge between the two domains is angled differently and the last helix which contributes to the 6HB in the immature hexamer is flexible (Figure 7c). These small changes not only lead to drastically different interactions within the hexamer but also between hexamers. The mature lattice forms intermolecular NTD-CTD contacts that stabilize the hexamer and are completely absent in the immature form. The CTD is the main stabilizing domain between hexamers. In HIV-1, the dimer interface in the mature capsid is formed by helix 9 and the trimer interface by helices 10 and 11 (*48*, *49*). Plotting the B factors of neighboring hexamers shows that the dimer interface is more stable (lower B values) than the trimer interface (higher B values) (Figure 7d). In the center of the mature hexamer is an interface formed by positively charged side chains of the NTDs. It is an important feature of the mature lattice as an IP6 molecule has been reported to bind there to neutralize the positive charges, similarly to how an IP6 binds the top of the 6HB in the immature lattice (*37*, *50*). HIV-1 has a beta-hairpin at the top of the NTD which can adopt different conformations depending on pH but it is not present in all retroviruses. It might regulate capsid core permeability for nucleotides needed for reverse transcription in the interior of the core by changing accessibilities to the central CA hexamer channel (*51*).

Recently, the pentamer structure could also be solved in native HIV-1 capsids by cryoET and subsequent subtomogram averaging (*45*). Surprisingly, it is different to the non-native structure solved previously by crystallography of cross-linked pentamers. The pentamer NTD is rotated by approximately 19° compared to its hexameric counterpart. This rotation forms a 10-helix bundle with the neighbour-

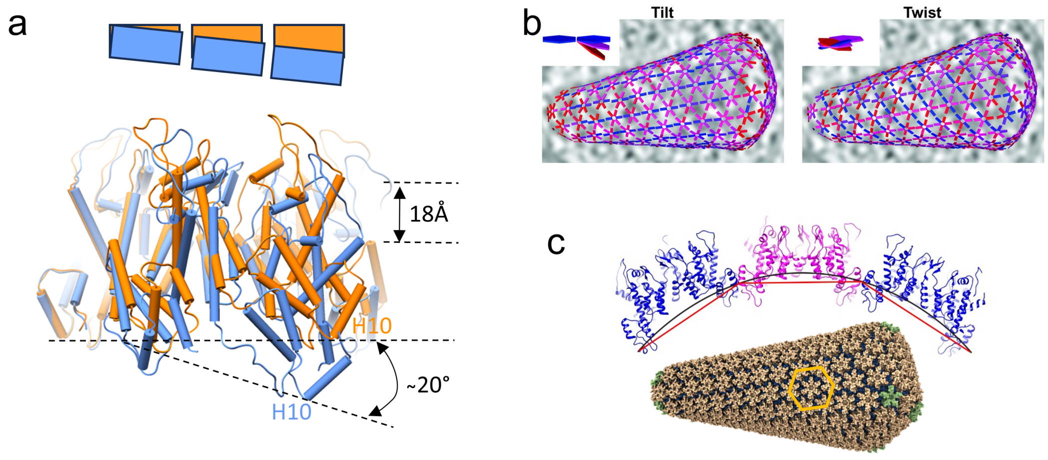


Figure 8 HIV-1 mature capsid curvature (a) Intrinsic curvature of the mature CA hexamer by superposition of a planar CA hexamer (orange, PDB 4XFX (*53*)) and a highly curved hexamer (blue, PDB 6SKK (*46*)). (b) Tilt (left) and twist (right) angles between hexamers within a single core. Insets show a schematic illustration of tilt and twist angles (*45*). The tilt/twist angle is indicated by the colour of the connecting lines between hexamer positions, from blue (less tilt along the long axis of the core) to red (more tilt along the circumference). (c) All-atom atomic model of the HIV-1 conical core (bottom) and a cross-section of three hexamers along the curved direction (top). The black line illustrates continuous curvature of the lattice given by both intra- and inter-hexamer curvature, while the red line illustrates discrete curvature given by inter-hexamer curvature alone (*46*). The lattice unit is marked with orange hexagon; pentamers shown in green (*48*). This figure is adapted from my paper (*1*) where it is licensed under a Creative Commons Attribution license (CC BY 4.0).

ing hexamer and excludes helix 3 from the interface. Also, the binding site at the NTD-CTD interface for host factors and small molecules such as PF74 opens up (*45*). The transition between hexamer and pentamer conformations has been proposed to be regulated by the arginine residues (R18) at the center of the HIV-1 CA pentamer by balancing its electrostatic destabilization with stabilizing the lattice (*35*, *52*).

As with the immature lattice, the structural conservation of the mature lattice between retroviruses is remarkable (*29*). The intrinsic curvature of the hexamer caused by the flexible linker between NTD and CTD on the one hand (Figure 8a) and the different tilt and twist orientations between CA hexamers on the other hand (Figure 8b/c) allows retroviruses to accommodate all the different morphologies (*45*, *46*, *54*) . This has not only been shown in HIV-1 with a cone shaped core but also in mature RSV which has a highly variable core (*35*, *46*).

1.3.2 Reverse Transcriptase

The reverse transcriptase transcribes the RNA into viral DNA (vDNA). The reverse transcriptase (RT) is expressed as part of the Gag-Pro-Pol polyprotein. The *rt* gene sequence is being used to classify retroviruses, including HERV-K proviruses (*55*), and it is thus particularly important to understand what this protein does. RT in retroviruses consists of two elements necessary for reverse transcription: the DNA polymerase that transcribes the RNA and the RNase H which degrades the RNA after transcription (*56*). HIV-1 RT is a heterodimer, where the larger subunit p66 has the active site for both the polymerase and RNase H. The smaller subunit p51 contains the same polymerase domains without the RNase H but in a different arrangement than the larger subunit and is only being used as structural support (*56*).

So far, only HIV-1 RT inhibitors are being assessed in clinical trials to inhibit the HERV-K RT such as Lamivudine to treat lymphoma (trial number NCT01528865) or Tenofovir alafenamide to treat ALS (trial number NCT02437110). The structure of the HERV-K RT has been solved in 2022 by Baldwin et al. (*57*) to develop specific RT inhibitors to treat cancer and neurological diseases. They showed that

Figure 9 Reverse Transcriptase (RT) RT overlay of HERV-K with HIV-1. HIV-1 is colored in gray. HERV-K has one molecule colored per usual conventions (fingers (blue), palm (red), thumb (green), connection (yellow), RNase H (orange)), the nucleotides in pink and the second molecule colored in black. PDB accession codes: 7SR6 for HERV-K (*57*), 5TXL for HIV-1 (*58*)

the HERV-K RT is an asymmetric homodimer instead of a heterodimer, with the RNase H not removed in the second subunit but instead disordered. The HERV-K RT forms the three subdomains fingers, palm and thumb which act as the polymerase and an additional domain at the C-terminus with the RNase H active site. This is the same fold as seen in HIV-1 (see overlay in Figure 9). The catalytical active sites of both the polymerase and RNase H are similar as well, showing the same RT mechanism. Surprisingly, the betaretrovirus MMTV, which has higher sequence identity to HERV-K in its RT, has been shown to act as a monomer (*59*). Comparisons to other viruses showed that the connection domain is very similar to the corresponding part in MoMLV (Moloney murine leukemia virus, gammaretrovirus) RT and the helical interactions between the thumb and RNase H are similar to the yeast retrotransposon Ty3 (*57*).

1.3.3 Integrase

The retroviral integrase (IN) is needed to insert the viral DNA into the host genome. IN is also produced as part of the Gag-Pro-Pol polyprotein and cleaved off during maturation. So far, only eight intact *int* genes in HERV-K sequences have been found (*60*) and it has thus not been widely studied. However, IN might be involved in various illnesses as IN can leave a genome unstable if it is too active or if the cell's DNA repair mechanisms upon integration are faulty (*60*).

The integrase has three (or four) domains: an N-terminal domain needed to multimerize, a C-terminal domain used to form the intasome, a catalytic core domain (CCD) in between with the DDE motif in the active site. Some retroviruses have an additional small domain on the N-terminal end which interacts with the vDNA (*60*, *61*). The structure of HERV-K IN has not been solved to date but it is expected to fold similarly as the MMTV integrase (*60*, *62*).

IN is not only responsible for the integration of the vDNA into the host genome but also helps in reverse transcription, organization of the RNA in mature virions and dimerization of the Gag-Pro-Pol precursor (*63*, *64*).

1.3.4 Protease

The protease of retroviruses is responsible to cleave Gag, Gag-Pro and Gag-Pro-Pol polyproteins during maturation to form a fully infectious virion. Retroviral proteases (PR) are generally encoded in a Gag-Pol polyprotein. Unlike in other retroviruses, the protease in HERV-K is encoded by its own reading frame (*65*) (see also Figure 2). Retroviral proteases act as homodimers and are included in the family of aspartic proteases. Even with low sequence identity, they all share the conserved regions in the active site, glycine-rich flaps and a C-terminal triplet (*65*, *66*). The protease monomer consists of four elements: two hairpin loops, an alpha helix and a loop with the active aspartic acid (*67*). Gag-Pro-Pol is present as an inactive monomer in the unmatured virion after budding. Only upon dimerization and autocatalytic cleaveage does it become fully active (*68*, *69*).

1.3.5 dUTPase

The deoxyuridine 5'-triphosphate nucleotidohydrolase (dUTPase) is encoded by a proteolytic cleavage of the Gag-Pro polyprotein in betaretroviruses with the N-terminus being the same as the NC part of Gag (*70*). All HERV-K families but one encode a dUTPase with different mutations and amino acid (aa) deletions (*71*). The consensus dUTPase sequence has been shown to be active but the endogenous proteins have not yet been confirmed to still be enzymatically functional in healthy cells (*72*, *73*). Primate lentiviruses, including HIV-1, alpharetroviruses and gammaretroviruses do not have this protein and it is thus the enzyme in retroviruses that has been studied the least(*70*).

dUTPases hydrolyze dUTP into dUMP and PPi which keeps the dUTP concentrations low in the cell. This helps to avoid the incorporation of dUTP instead of dTTP into DNA as most polymerases cannot distinguish between the two (*74*). This function is critical for retroviruses infecting non-dividing cells with low concentrations of cellular dUTPases to avoid mutations during the retroviral life cycle (*70*). The exception to this are primate lentiviruses which can also infect non-dividing cells by recruiting UNG, which is involved in DNA repair, via the viral protein Vpr to keep the mutation rates low enough (*75*). As betaretroviral dUTPases contain the NC part of Gag, they can additionally function similarly to NC and bind nucleic acids. This binding may locate the enzyme to places where reverse transcription takes place and the dUTP hydrolyzing activity is needed (*70*).

Human and retroviral dUTPases act as homotrimers with three active sites with high specificity to uracil (*74*). They contain 5 conserved domains and many dUTPase structures have been solved by crystallography (M-PMV (3TRL) (*76*), FIV (1F7D) (*77*) and EIAV (1DUN) (*78*)).

dUTPases can also serve functions independent of dUTP hydrolysis and HERV dUTPases act on cellular functions in disease. Psoriasis, an immune disease causing inflammation of the skin, has two single nucleotide polymorphisms in the dUTPase gene PSORS1 of a HERV-K (*79*). It could be shown that this dUTPase causes an inflammatory response and a higher antibody and T cell response in psoriasis patients compared to healthy individuals (*80*, *81*). However, as the enzyme does not need to be catalytically active to cause inflammation, the dUTPase might just be involved in cellular signalling (*70*, *80*).

1.3.6 Envelope

Envelope is translated as a glycosylated polyprotein precursor, cleaved by cellular proteases into the surface (SU or gp120 in HIV-1) and transmembrane domain (TM or gp41 in HIV-1) proteins. Env is active as a trimer on the viral membrane in which the SU and TM proteins of each monomer are non-covalently bound. Its heavily glycosylated surface, also called glycan shield, helps to protect the virion from being detected by the host immune system and helps to bind to the cellular surface. Several structures of viral envelope proteins have been solved to date (*82–85*). The HIV-1 SU protein gp120 fold is divided into the inner and outer domain with a bridging beta sheet in between them. Several disulfide bonds help the protein to keep its tertiary structure (*82*).

The HIV-1 TM protein gp41 consists of a C-terminal domain, a transmembrane domain and an extracellular domain, which is responsible for fusion via a 6-helix bundle made from the HR1 and HR2 domains. The C-terminal domain of lentiviral TM subunits are much larger than other retroviruses at 150 aa, compared to other retroviruses with 20-40 aa. The C-terminus influences many properties of Env such as Env incorporation, fusion and changes the conformation of gp120 as well as the extracellular domain of gp41 (*82*).

The different conformational states of Env help protect it further from the immune response of the host. The drastic changes upon binding to the primary receptor helps shield the coreceptor surface until Env is in close contact with the target cell (see also Figure 11). Different conformations also help to distinguish Env bound to viral membranes from Env on infected cells. Env in its unbound form is very dynamic and can change between three states, allowing it to be recognized by receptors while still being in the non-reactive ground state most of the time to evade detection (*86*, *87*).

1.3.7 Proteins specific to HERV-K

HERV-K has several additional proteins but they have not been studied in depth and many details of how they function are still unknown. Type I HERV-K (HML-2) sequences contain a 292 bp deletion in the *pol* and *env* genes which causes an alternative splicing of the mRNA and results in the expression of Np9. Type II sequences without the deletion express Rec instead of Np9 (*88*).

Np9 Np9 has no known function but is expressed in healthy cells (*89*). It locates to nucleoli (*90*). Np9 is crucial for the viability and motility of tetracarcinoma cells (*88*) but the role of Np9 in cancer is still poorly understood. Only some progress has been made in identifying its interacting proteins. Np9 as well as Rec can suppress PLZF (promyelocytic zinc finger), a tumor suppressor (*91*). Np9 can also activate β-catenin, ERK, Akt and Notch1 and their signaling pathways, thus promoting leukemia cell growth (*92*).

Rec Rec is produced in healthy and diseased cells (*89*). Rec helps to transport unspliced and singly spliced viral RNA from the nucleus to the cytoplasm, the same function as Rev in HIV-1 (*93*). However, Rev is also involved in the translation of the transported mRNA but this function has not yet been investigated for Rec (*73*). Rec contains an arginine-rich nuclear location signal sequence (NLS) which interacts with importin-β and ensures its import into the nucleus (*73*). Rec also contains a leucine-rich nuclear export signal (NES) which targets the protein and its cargo to be exported from the nucleus (*94*). The export function of Rec is achieved by the binding of up to three Rec tetramers to Rec responsive elements

(RcREs) in the 3' LTR of transcripts and shuttling of the whole complex through nuclear pores (*95*).

1.4 Life cycle of retroviruses

A retrovirus must complete a full life cycle in a host cell in order to multiply. Figure 10 summarizes the different stages in HIV-1.

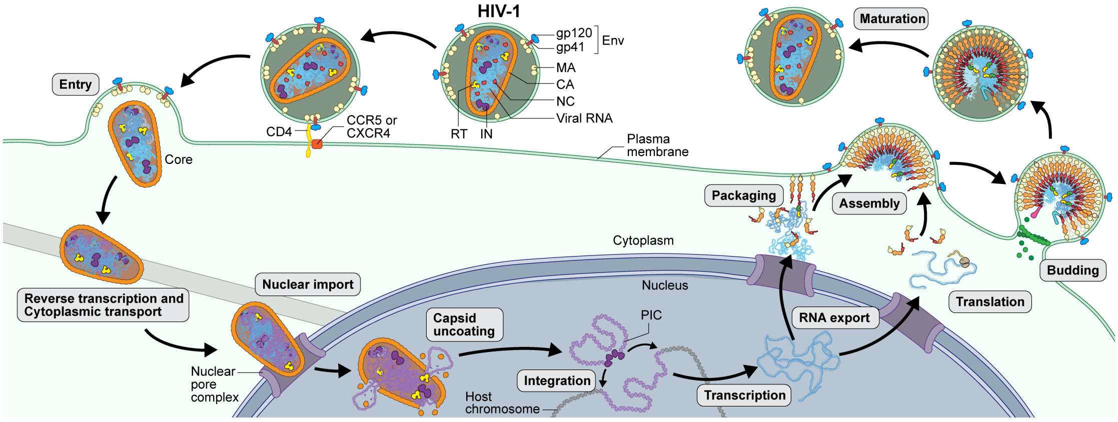


Figure 10 Viral life cycle Schematic of the HIV-1 viral life cycle. Clicking on each stage links directly to the corresponding chapter. This figure is adapted from `https://scienceofhiv.org/wp/life-cycle/` where it is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

1.4.1 Cell Entry

The first step in a retroviral life cycle is recognizing the correct host cell and fusion of the viral and cellular membranes to allow the viral core to enter into the cell. This step is mediated by Env, which is present in only a few copies on the HIV-1 viral surface. Env can recognize specific receptors on the target cell and

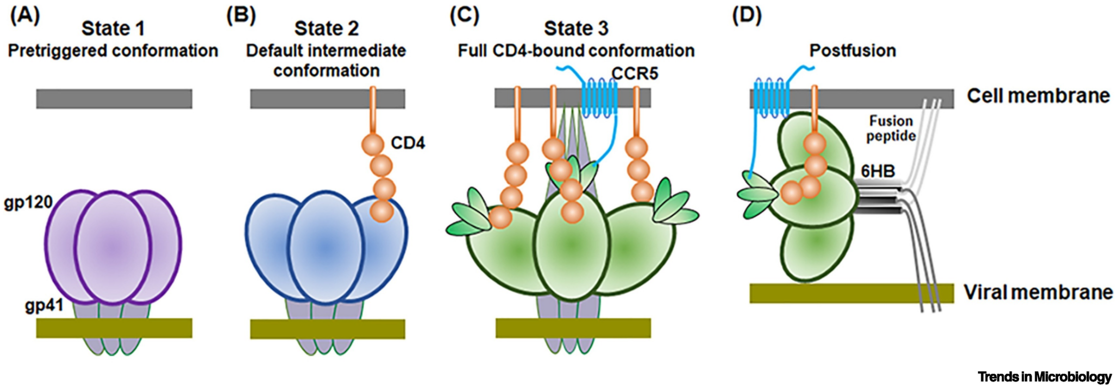


Figure 11 HIV-1 Env conformational states (A) Pretriggered conformation cannot be recognised by most antibodies. (B) Binding to one CD4 receptor leads to a partially open conformation. (C) State 3 exposes the coiled coil in gp41. (D) The fusion peptide of gp41 is inserted in the membrane and the fusion of the two membranes can succeed. This figure is taken from (*96*) where all rights are reserved by Elsevier. I acquired the permission to use this figure in my thesis under the license number 5598810697852.

mediates fusion of the viral membrane with the plasma membrane of the target cell (Figure 11). Not many receptors have been found yet for each retrovirus. The most widely studied interactions are between HIV-1 and its receptor CD4 on T cells. The HIV-1 Env is synthesized as an inactive precursor which gets cleaved into the active glycoprotein gp120 and the transmembrane protein gp41. Gp120/gp41 trimers on the viral surface recognize the CD4 receptors on the target cells. Spontaneous conformational changes from an inactive ground-state to a more active conformation which recognizes CD4 allows the recognition of target cells (*86*). The binding of gp120 to CD4 triggers a conformational change that forms a bridging sheet and exposes the binding site for the coreceptors CCR5 or CXCR4 (*87*). The binding of these coreceptors then triggers a further conformational change in the viral surface proteins which leads to the fusion of the two membranes (*82*).

1.4.2 Reverse Transcription, Uncoating & Nuclear Import

After cell entry, the virus must reverse transcribe its vRNA into vDNA for insertion into the host genome. Reverse transcription is catalyzed in a reverse transcription complex (RTC), containing the RT and several additional viral proteins. Where in the cell this process takes place, when it is imported into the nucleus and at what point it transitions into the preintegration complex (PIC) is under heavy debate. A study in 2002 showed active HIV-1 RTCs being transported by dynein and the microtubule network towards the nucleus (*97*). A different study showed that the HIV-1 capsid does not disintegrate immediately upon entry into the cell, but is being transported towards the nucleus as an intact cone (*98*, *99*). The intact cone cannot only interact with the microtubule network, but also provides a safe space for reverse transcription to occur. Early immunofluorescence studies showed that the HIV-1 capsid cone is shed at the nuclear pore during or just before entry (*100*, *101*). However, newer research showed that capsid enters the nucleus (*102–104*) but it has not yet been determined whether it happens as single CA molecules or as (partially) complete cones. The newest model summarized by Müller et al. (*99*) suggests a handover of proteins that bind the capsid in the cytosol such as CypA, to NPC proteins, to nuclear proteins such as CPSF6. It would thus not need any additional nuclear import factors for the translocation but the binding of FG repeats of Nups in the nuclear pore might be sufficient.

Reverse transcription is already initiated when the virus is assembled. However, the lack of dNTPs in the virus stops it from continuing before cell entry. Once the viral core has entered the cell, dNTPs from the cytosol can enter and transcription can continue (*105*). We do not yet know why reverse transcription needs the import into the nucleus and does not terminate in the cytoplasm (*99*).

1.4.3 Integration

Reverse transcription leads to a viral DNA with an U3-R-U5 end on both the 3' and 5' end. These LTRs are essential for integration into the host DNA (*56*). The intasome, a multimer of IN and the viral DNA (vDNA), is the functional unit during integration. Different stochiometries of IN have been observed in retroviruses but there are only ever two active sites per intasome. First, IN cleaves off the 3' hydroxyl groups of the vDNA to cut the chromosomoal DNA and attaching the 3' end of the vDNA to it. This reaction leaves gaps at the joining points which get repaired by cellular enzymes used to repair normal DNA damage. It leaves a stable provirus in the cell's genome, which is ready to be transcribed to RNA and will act as either mRNA or vRNA. Surprisingly, integration is done on chromatin and the nucleosomes present the DNA in the preferred form to the intasome (*61*, *106*).

Integration does not occur evenly across the genome but is guided by interactions of the integrase with host factors associated to the host chromatin (*61*). Therefore, integration hot spots exist in the genome.

1.4.4 Transcription & Translation

Transcription by the RNA Polymerase II leads to a non-spliced Gag-Pro-Pol mRNA with a 5'cap and a 3'poly(A)tail like any other cellular transcript. These primary transcripts give rise to Gag, Gag-Pro and Gag-Pro-Pol polyproteins and are also packaged into the virion as a dimer during assembly. A single splice event of the initial transcripts gives rise to the env mRNA and further splicing leads to mRNAs encoding accessory proteins like rec and np9 in the case of HERV-K. The different outcomes of a single primary mRNA need to be regulated to avoid complete splicing every time like in cellular pathways. Retroviruses achieve these different splicing events by having strong 5' splice site donors and weak 3' splice site acceptors and thus having incomplete or alternative splicing events. The binding of cellular splicing enhancers and silencers and secondary structure elements can further guide the different RNA outcomes (*107*).

The mRNA gets exported from the nucleus with the help of accessory proteins in complex retroviruses such as Rec in HERV-K, Rev in HIV-1, Rex in HTLV and Rem in MMTV. Simple retroviruses which do not have these additional proteins use *cis*-acting viral sequences to interact with export proteins (*107*). In HERV-K, Rec binds to the Rec responsive element (ReRE) on the 3' LTR of the mRNA via its arginine-rich nuclear import signal (*95*). The export from the nucleus is supported by the interaction of Rec with Staufen-1. Staufen-1 is known to be involved in RNA export and can help regulate translation and interact with Gag during assembly. However, it is still unclear how this is achieved (*94*).

Translation of the long Gag-Pro-Pol polyprotein always starts on the 5' end with

Gag. To translate the complete polyprotein, the stop codon of Gag either needs to be read through (i.e. MLV) or the reading frame needs to be shifted by -1 (i.e. HIV-1) (*108*). The reading frame of HERV-K is shifted before Pro and again before Pol to translate the complete Gag-Pro-Pol polyprotein. Frameshifting can happen due to the presence of a slippery sequence upstream and a structured pseudoknot sequence downstream of the termination codon. During frameshifting the two tRNAs in the P and A sites of the ribosome simultaneously slip into the -1 frame to adopt a two-of-three anticodon-codon pairing before translation continues in the new frame (*109*). Frameshifting is relatively rare, estimated to happen approximately 20-25% (*110*) in the betaretrovirus MMTV which leads to lower protein levels of Pro and Pol. This stochiometry is advantageous to the virion as more structural Gag proteins are needed compared to enzymatic proteins to form a fully functioning virus (*107*).

1.4.5 Assembly & Budding

After translation, Gag is transported to the viral assembly sites. This is done in an auto-inhibited conformation to prevent early assembly at the wrong place in the cell (*111*). Alpha-, gamma-, delta-, and lentiviruses assemble at the plasma membrane (PM) whereas beta- (except HERV-K) and spumaviruses do so at internal cellular membranes (*73*, *112*). Upon binding of MA to PI(4,5)P_2 (Phosphatidylinositol 4,5-biphosphate) on the PM, HIV-1 inserts its myristoyl group of MA into the inner leaflet of the PM (*11*, *113*). The myristoyl group is a semi-conserved feature of MA as it is present on many retroviruses (HIV, MLV, HTLV, BLV) but missing on others (RSV, EIAV, FV). Interestingly, MMTV and M-PMV do have this posttranslational modification even though they are part of the betraretrovirus group and thus do not assemble at the plasma membrane. Retroviruses have different requirements regarding their dimerization and oligomerization states before tracking to the PM can occur, which might be caused by different affinities of Gag to different lipids in cellular membranes (*114*, *115*).

Retroviruses are diploid; they package exactly two copies of their RNA into each virion. Gag can recognize the packaging signals in the 5'UTR of the gRNA which guarantees the highly specific packaging into the virus (*116*). The unspliced and untranslated HIV-1 RNA dimerizes at the PM, where it is recognized by the CCHC-type zinc finger in the NC domain of Gag, stabilized and recruited into assembly complexes (*117*, *118*). Even though NC is is the main RNA binding site of Gag, MA can also bind RNA via the HBR domain that also binds PI(4,5)P_2 at the membrane. Binding of cellular tRNAs in the cytosol to MA prevents early assembly and allows for targeting to the specific membranes where assembly occurs (*119*, *120*).

Env has a slightly different pathway than Gag to reach the assembly sites. It is glycosylated when travelling through the ER and Golgi, where it gets cleaved. In HIV-1, Env is then transported via the secretory pathway to the PM where some of it is internalized by the clathrin-mediated pathway again. This is necessary to keep levels low on the cell surface and on the virion to avoid detection by the immune system (*83*, *121*).

Once assembly has finished, nascent viral particles get pinched off from the membrane with the help of ESCRT (endosomal sorting complex required for transport). ESCRT is recruited by Gag via its late domains. The late domains in HIV-1 are located in its p6 domain and can bind the early ESCRT factors which can then recruit further factors to complete the scission of the membrane to release the virus. In HERV-K, the late domains are part of the p15 protein in Gag and provide the same function (*122*). Budding HIV-1 virions at the PM could be observed by cellular cryoET. The tomograms showed an abundance of hemispherical virions but only a few spherical ones attached to the PM via a membrane neck, indicating that the membrane fission to complete budding is a very fast process (*123*, *124*).

1.4.6 Maturation

A virus matures on three levels which occur simultaneously:

1. At a sequence level, the polyprotein is cleaved by the protease into all its single proteins.
2. At the structural level, a rearrangement occurs in MA, CA subdomains and NC, vRNA and IN.
3. At the architectural level, the immature lattice is converted into the mature core which contains hexamers as well as pentamers and completely encloses the ribonucleoprotein complex (RNP).

The viral maturation is a very fast process. This is evidenced as all viruses still attached to the cell membranes were found to be immature in morphology but most viruses near cells were already matured (*124*). The viral protease must thus be activated either at the same time as the completion of budding or shortly after. How this activation occurs is still unclear. Nevertheless, the sequence and kinetics of the cleavage must be tightly controlled and depend on the affinities between cleavage sites and the active site as well as how easily the cleavage sites can be accessed by the protease. Small alterations in the timing, rate or sequence of the cleavage events can alter virion infectivity (*125*). HIV-1 Gag is cleaved as followed: Between the spacer peptide 1 (SP1) and NC to release NC and vRNA which forms the RNP, then the cleavage between SP2-p6, then between MA and CA, between NC and SP2 and finally between CA and SP1. The last cleavage was found to be a maturation switch as it destabilizes the CA-SP1 6HB responsible to form the immature lattice (*126*). However, the cleavage rates and sequences have not yet been studied in HERV-K and a comparison to other retroviruses is difficult as HERV-K Gag has different small peptides which cannot be directly compared to SP1, SP2 and p6 in HIV-1.

The rearrangement during maturation is not only present in Gag but also in Env. The sparsely scattered Env trimers on the surface of the virion change from scattered and static to a state that is more clustered together and mobile (*127*).

To explain how all the three maturation levels can occur inside an enclosed virion, there are currently three models on how the maturation takes place in retroviruses:

1. Displacive transition: It postulates that the CA rearrangement happens at the same time as the maturation of the core. It was thought that the mature core is formed directly from the immature lattice by rolling away from the membrane and enclosing the NC and RNA (*128*, *129*). This hypothesis was set when *in vitro* studies showed that tubular CA-NC assemblies do not diffuse when the protease is added.

2. De novo assembly: This model means that the immature Gag lattice completely disassembles upon Gag cleavage and the mature lattice reassembles from a subset of CA monomers. It was initially proposed when structural studies showed that the CA NTD and CTD orientations and protein contacts change from immature to mature conformations in a way that cannot be contained within the space of the virion.

3. First displacive nucleation and then de novo assembly: This model is a combination of the previous two and the consensus reached by studies incorporating biochemical assays, cryoEM and computational modelling. First, some subunits in the immature lattice undergo a conformational change without disassembling and this sheet then rolls away from the membrane around the genome as described in the displacive model with a contribution from the RNP condensation. But the completion of the mature lattice is reached by the sheet growing in a de novo mechanism and closing the mature shell (*130*).

1.5 HERV-K

HERV-K is classified as a betaretrovirus-like endogenous retrovirus. The viruses that integrated into the human genome to form the HERV-K group possessed all the necessary proteins to successfully complete a full viral life cycle. However, transcription of proviruses is tightly regulated, and especially so in endogenous viruses like HERV-K. The enhancers and promoters in their LTRs, and thus transcription as a whole, are mostly silenced in healthy cells. However, it is not yet understood how and when this silencing of HERV-K is lifted and the RNA gets transcribed.

As a betaretrovirus, HERV-K proviruses have the additional reading frame for *Pro* and act like other viruses in the same genus. Surprisingly however, HERV-K assembles at the PM (*131*), whereas other betaretroviruses assemble at pericentriolar regions (*132*, *133*). As the life cycle of HERV-K has not been studied in depth, many of its finer details are still unclear.

The next chapter will go into greater detail into how endogenous viruses integrated into the human genome and I will present my work on HERV-K to help understand how HERV-K virions mature.

Chapter 2

HERV-K

2.1 Disclaimer

This chapter introduces endogenous retroviruses and shows all the structural work I have done on them. Some experiments that were performed by others are also included. The cryoET data was collected by Andrew Howe. My work on the tilt series reconstruction and subtomogram averaging is included, but the subtomogram averaging used to generate the final density map was done by Hsuan-Fu Liu and Ye Zhou under the supervision of Alberto Bartesaghi due to a long shutdown of our computers. The molecular dynamics simulations were done by Juan S. Rey, Lev Levintov, Juan R. Perilla. I got help from Geoff Sutton and Weng Ng to generate the structure-based phylogenetic trees. The data presented here is part of a manuscript which was accepted in Nature Communications at the time of writing (*134*).

2.2 Introduction

Endogenous retroviruses (ERVs) are prevalent in many species; PERVs (Porcine ERVs), MERV (mouse ERVs), HERVs (human ERVs). They fall into the class of transposable elements with long terminal repeats (LTRs), opposed to long/short retrotransposons lacking LTRs (LINEs/SINEs) (*135*).

ERVs arise when an exogenous retrovirus infects the germline of an organism. This virus can then get passed down to its offspring and survive generations, eventually getting fixed in the genome (*136*). ERVs can amplify in the genome through intracellular retrotransposition or by reinfecting the germline if its life cycle is still intact. As these endogenous viruses are also inherited vertically, they can acquire mutations and are subject to evolutionary pressures like any other gene.

2.2.1 Mutations in ERVs

The mutations acquired through the millions of years of an ERV being inserted into the genome rendered most unable to complete a full life cycle. Therefore, they cannot infect other individuals of their host species and became fixed in the genome.

Some ERVs are mutated to such a degree that they provide no known function to its host and are part of the DNA initially termed as junk. Others got repurposed in the course of evolution and the proteins provide a new function for the organism. The most famous example for such a repurposed protein is Syncytin-1. Syncytin-1 is the envelope glycoprotein of the human endogenous retrovirus HERV-W. Syncytin-1 is involved in the fusion of cytotrophoblast cells in the placenta (*137*, *138*).

Sometimes the host organism also uses part of ERVs to provide a regulatory function for the expression of other genes. LTRs can contain promoters, enhancers, hormone-responsive elements and polyadenylation signals (*139*). Some regulating examples are an element from a HERV-E insertion needed for the expression of pleiotrophin (PTN) or an element consisting of an Env with flanking LTRs mediating the expression of the insulin-like growth factor INSL4 (*140*).

2.2.2 Nomenclature HERVs

The nomenclature of ERVs is very often confusing and misleading. Many genes were found and named before scientists realised that they were part of an ERV. This lead to genes having multiple names and inconsistency across and within species. The International Committee on Taxonomy of Viruses (ICTV) has not yet classified endogenous retroviruses. Any naming scheme is therefore only an attempt to make sense of the many sequences that originated from exogenous retroviruses and a consensus has not yet been reached. I will try to give a brief summary of the naming strategy I will use within my thesis.

ERVs are named after the species they infect. This is reflected by the first letter in their names; HERVs are human (h) ERVs, PERVs are porcine (p, pigs) ERVs, MERVs are mouse (m) ERVs and so on. The HERVs are further classified by the tRNA binding to the RNA primer-binding site (PBS) for reverse transcription. The one letter code of this amino acid is added after a hyphen; HERV-Ws use a tryptophan, HERV-Hs use a histidine, HERV-Ks use a lysine etc. Some HERVs, especially if the class has many members, then get further classified into groups based on sequence similarity. HERV-K for example gets divided into the 11 classes HML-1 to HML-11 (*141*). Problems in this convention arise when an ERV integrated before the current species emerged and are thus present in several species or when ERV sequences are very similar but the tRNA used has been mutated or that part of the sequence is completely missing. Sometimes ERVs will therefore be named by their position in the genome such as 4q13.2 representing an *env* gene in humans (*141*) but I have decided not to use this strategy when other names exist as it gets unruly to read and use very quickly.

HERV-K (HML-2) is one of the most studied endogenous retroviruses, the virus I am focusing on in my thesis and a prime example of confusing nomenclature. HERV-K (HML-2) was first found by Callahan et al. in 1982 (*142*) and named after the name of the colony it was found in, HLM-2. Over the years it was termed many things, not always realising when studies talked about the same sequence. Only years later it was found that this initial clone was in fact a sequence of a HERV-K. In 1993, HERV-K sequences were divided into 6 subtypes named human mouse mammary tumor virus-like, or HML for short (*55*). The mouse mammary tumor viruses are betaretroviruses (see also Figure 1), the most similar retrovirus to HERV-K based on sequence and HERV-K has therefore been classed as betaretrovirus-like. Later on, the HML classification was refined to include HML1-10 and finally an 11th type was added when it was realised that not all HERV-K sequences fall into the first 10 subtypes (*141*). The first clone HLM-2 was found to be a member of the HML-2 subtype and thus renamed, further adding to the confusion as the names are so similar.

I will refer to this endogenous virus as HERV-K (HML-2) throughout my thesis as I think this terminology is the most clear but it also has the following names: HLM-2, HML-2, HERV-K10, HTDV/HERV-K, HERV-K, HERVK, ERVK or HERV-K(HML-2) (*73*).

2.2.3 HERV-K (HML-2)

The HERV-K (HML-2) proviruses are the most recently incorporated endogenous viruses in the human genome and mostly the result from different integration events (*141*). Comparisons of flanking LTR sequences can give an estimate of how old proviruses are. LTRs of a provirus are identical upon integration into a genome but can acquire mutations after integration. Comparisons of the number of acquired mutations in LTRs since integration provide a good base to calculate their age. However, estimations of the average elapsed time between mutations vary (*139*, *143*). HERV-K(HML-2) shows members to be present in both hominoids and Old World monkeys, giving them an age of over 35 million years (*144*). HERV-K113 stems from the most recent integration event and is probably the most widely studied member of the family. It is an example of a polymorphic provirus, not present in all humans and estimated to have integrated as recent as 200'000 - 1 million years ago (*145*, *146*).

Some HERV-K (HML-2) proviruses still possess complete open reading frames (ORFs) and some encode all structural and enzymatic proteins (schematic of a provirus in Figure 2 and section 1.3 explains the function of all proteins). Mutations and deletions however have rendered them incapable of completing a full life cycle. Its almost complete genome have made HERV-K (HML-2) a prime target to study HERVs. The Bieniasz lab and Heidmann lab almost simultaneously produced a consensus sequence for HERV-K (HML-2) based on the most complete proviruses in the genome (*131*, *147*). Lee & Bieniasz took 10 full length HERV-K (HML-2) sequences which all had at least one intact ORF to build a consensus sequence (Figure 12a). They show that their consensus virion is capable of completing a viral life cycle, including integration and reverse transcription. Surprisingly, they assemble at plasma membranes unlike other betaretroviruses which assemble in the cytoplasm (Figure 12b). Their consensus sequence can be seen as a good approximation of the ancestral sequence when the retrovirus integrated into the human genome. I am using their construct HERV-K_{con} throughout my work.

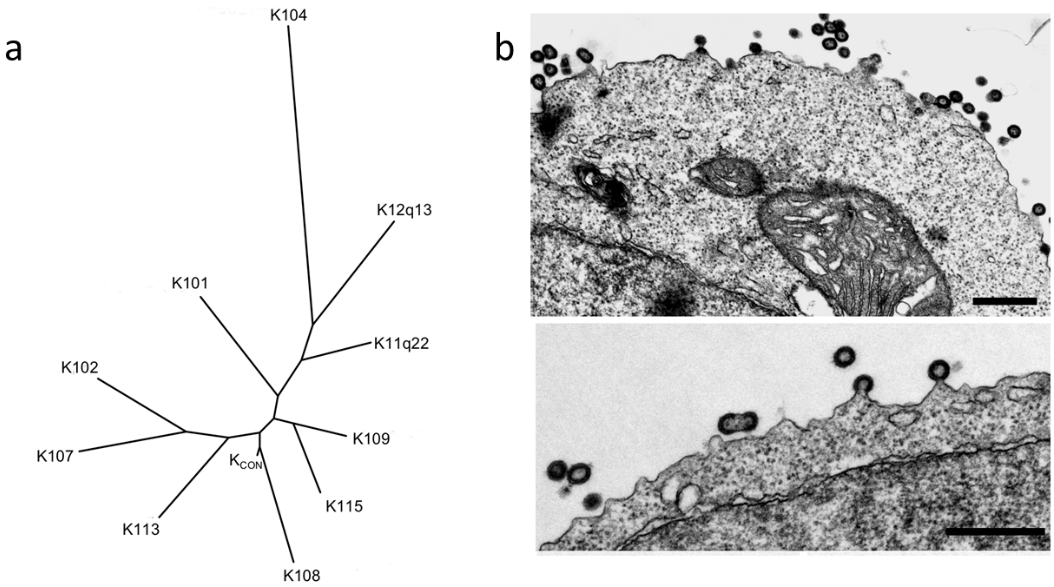


Figure 12 HERV-K_{con} (a) Phylogenetic analysis of HERV-K_{con} and the ten proviruses used to generate it. (b) Electron micrographs of 293T cells transfected with the plasmid. Scale bars are 500 nm. Both figures were taken from (*131*) where it is licensed under a of the Creative Commons Attribution license (CC BY 4.0).

2.2.3.1 HERV-K (HML-2) in illness

HERV-K (HML-2) has also gained a lot of attention in recent years as it is found to be active in several illnesses and hopes have soared that it could be used in diagnostics.

Cancer In healthy cells, the HERV-K genes are under tight epigenetic control and not expressed. However, during cancer this control can be lost and mRNA as well as proteins were found to be expressed in various cancers. Several transcription factors which are involved in cancers were shown to bind HERV-K LTRs and influence their activity (*148*). HERV-K (HML-2) could be associated with breast cancer, prostate cancer and many more. It is still under debate however if HERV-K plays a role in causing cancer or if it is merely a side event and expressed in inflammation reactions. HERVs can be involved in cancer in several different ways:

1. Integration into tumorsuppressor genes.
2. Acting as a promoter for oncogenes.
3. Env can act as an immunosuppressor.
4. Proteins can act directly as an oncogen.

Neurological disorders Several retroviruses have been shown to cause neurological diseases in mice and HIV-1 and HTLV can lead to ALS-like symptoms (*149*). The evidence that HERV-K (HML-2) contributes to amyotrophic lateral sclerosis (ALS, commonly known as Lou Gehrig's disease) is strong. All HERV-K (HML-2) transcripts were found in ALS patient tissues and its *env* gene contributes to the neurotoxicity in the disease. (*150*, *151*). However, no study has determined yet whether all provirus loci are contributing to the env toxicity and the causes for its activation are still unknown. Some therapy approaches for targeting HERV-K (HML-2) to treat ALS and cancer are summarised in (*149*).

Exogenous virus infection As HERVs stem originally from exogenous retroviruses, their proteins can interfere with a new infection once they are present in the genome by blocking cell receptors or interfering with intracellular trafficking (*152*). HERV-T env may have led to the extinction of its exogenous retrovirus (*153*) and HERV-K (HML-2) can interfere with HIV-1. Elevated HERV-K (HML-2) mRNA levels were found during HIV-1 infection (*154*–*157*). These transcripts may then be translated into proteins which can generate an immune response (*73*). Indeed, HIV-1 Gag can co-assemble with HERV-K (HML-2) Gag *in vitro* which can lead to lower HIV-1 virion levels and infectivity (*158*, *159*).

2.2.3.2 Expression of HERV-K (HML-2) in nondiseased adult tissues

HERV-K (HML-2) was thought to not be expressed in healthy cells for a long time, simply because it accumulated so many point mutations and deletions that it was unlikely to be transcribed. Epigenetic silencing further controls its expression

(*160*). So most research to date focused on diseased tissues. However, a targeted search for the Env protein of HERV-K in placenta using Western blots and immunohistochemistry, revealed that Env is indeed expressed in cytotrophoblast cells during placental development (*161*). A complete analysis of RNAseq data in healthy tissues further showed that many HERV-K (HML-2) proviruses can be detected, with high levels measured in the cerebellum, pituitary, testis and thyroid tissues (*152*). Surprisingly, some proviruses with intact ORFs were found to be expressed as well, but their significance remains unclear to date and no proviruses with known complete *gag-pro-pol* sequences were found (*152*). When looking at HERV-K (HML-2) in context of an illness, it is thus very important to take the normal expression of individual proviruses into account. When trying to use this virus as a marker or drug target, these offtargets need to be kept in mind. More research is required to understand how and why some proviruses are transcribed in nondiseased tissues. However, it opens interesting avenues to find more beneficial proviruses and to better understand the evolution of HERV-K.

Aging Surprisingly, endogenous retroviruses were recently found to be transcribed and producing VLPs during aging in organs and tissues of humans and mice (*162*, *163*). Aging can lead to increased levels of transposable element transcripts, successful retrotransposition of these and their decreased repression but only little research in the role of HERVs during aging has been done so far (*164*). So far, only single-molecule RNA fluorescence in situ hybridization (smRNA-FISH) experiments detected the presence of *env*, *pol* and a regulatory sequence in the LTR of HERV-K in senescent cells. This is driven by epigenetic derepression of these regions and even led to the production of VLPs in senescent cell lines. These VLPs not only triggered an innate immune response in the senescent cell itself but shed VLPs could also transmit aging information to young cells (*163*).

2.2.4 Project Aims

HERV-K has garnered increasing attention in the last few years and some drugs targeting this endogenous virus are currently in trial phase but overall it is still poorly understood. Not only is research needed to understand its role in various diseases but also much information is missing on how the retroviruses looked like when it integrated in the human genome. My aim is to elucidate the structure of the immature Gag lattice of this ancient virion and draw parallels to other, exogenous retroviruses which were more widely studied so far.

There are two main approaches in cryo-electron microscopy to reach such a goal; single particle analysis and cryo-electron tomography combined with subtomogram averaging. Both have advantages and disadvantages and need to be chosen carefully before starting to freeze samples.

Single particle analysis Single particle analysis (SPA) is often the easier approach and thus the preferred method for many projects. Recent technical and methodological advances even lead to proteins resolved at atomic resolutions (*165*). For single particle analysis, the sample gets imaged at a fixed angle in the electron microscope during data collection. Assuming that the protein sample is uniform and randomly orientated across images, one has thus taken images of almost every possible view of the protein. Thus, their 2D projections can then be aligned, classified, averaged and finally a 3D structure can be reconstructed (*166*).

While SPA is very often more straightforward to get started on, it can become problematic if the sample is not homogeneous or if all the particles show a preferred orientation on the support grid used for freezing. Non-homogeneous samples cause blurring when averaging different protein states together and preferred orientation leads to an under-representation of some orientations when generating a 3D map and causes streaking in one direction. Both problems limit the resolution one could achieve.

Cryo-electron tomography & subtomogram averaging Cryo-electron tomography can solve many problems of SPA but comes with its own challenges. During cryoET data collection, the sample gets imaged at different angles by tilting the stage in the microscope. The images of such a tilt series then get motion corrected and aligned, often using gold fiducials in the sample, to generate a tomogram. During subtomogram averaging, the proteins get extracted as smaller subtomograms from the initial tomogram. Then, these small 3D volumes can be aligned and averaged to form a final map.

CryoET can deal with difficult samples, including proteins in and on cells and heterogeneous proteins. However, spreading of the electron dose across a complete tilt series, instead of just taking one image like in SPA, leads to very low signal in the tomogram and makes it harder to analyse. Also, beam-induced motion during data collection as well as thick ice at high tilt angles limits the resolution achievable (*167*).

Constrained single-particle tomography can partially resolve some of these problems by combining the normal subtomogram averaging workflow with some advantages of SPA. It uses the same approach as normal subtomogram averaging methods to find the orientation of each particle in the tomograms. However, this information then gets used on the raw tilt series data which can now be treated very similarly as the single particle images to extract 2D projections and to reconstruct a final 3D map. This combination of the two methods leads to better orientation accuracy of the particles, reduces model bias and overrefinement and also allows the use of the simpler and more accurate CTF correction of 2D images instead of 3D volumes (*168*, *169*).

To resolve the structure of HERV-K, I purified VLPs from the consensus HERV-K construct HERV-K_{con} and characterized them. As the sample had great differences in VLP size and holes in the CA lattice, I used cryoET with subtomogram averaging to build an atomic model of the immature CA lattice. Comparisons of the final structure with its mature counterpart shed light on how the virion matures and parallels to the maturation of HIV-1 could be drawn. Furthermore, a structure-based phylogenetic analysis with other CA structures from immature retroviruses showed a surprising amount of similarities across retroviral classes and evolutionary time.

2.3 Methods

2.3.1 Molecular Biology

2.3.1.1 Site-directed mutagenesis by PCR

Site-directed mutagenesis can be achieved by doing a Polymerase Chain Reaction (PCR) with a forward and reverse primer, each carrying the desired mutations. The parent template is then removed using a methylation-dependent endonuclease such as DpnI. The final product is transformed into bacteria for further amplification, sequencing or long-term storage (Schematic in Figure 13).

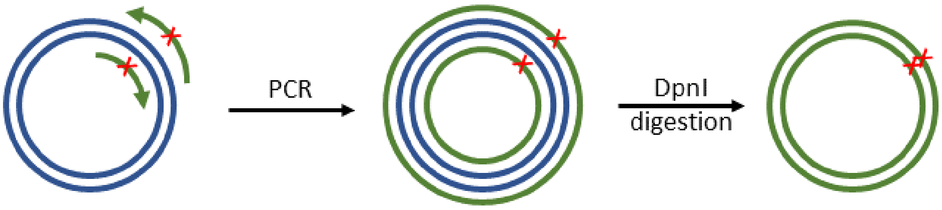


Figure 13 Site-directed mutagenesis Schematic to introduce a single mutations (red cross) into a plasmid (old: blue, new: green) with the help of a forward and reverse primers (green arrows).

The PCR was done using Pfu Turbo to incorporate single point mutations into the Gag gene. The QuikChange Primer Design tool (`https://www.agilent.com/store/primerDesignProgram.jsp`) was used to design the primers. They were ordered from Thermo Fisher at 25 nmol, desalted, with no special handling. The 50 µl reaction mixture in Table 1 with the program in Table 2 was used for the PCR.

After the PCR reaction finished, 1 µl DpnI was added to the mix and incubated at

DMSO	1.5 µl
dNTP	1 µl
10x Pfu Turbo buffer	5 µl
Plasmid (100 ng/µl	1 µl
Primer Fwd (10 µM)	1 µl
Primer Fwd (10 µM	1 µl
H_2O	38.5 µl

Table 1 Mutagenesis PCR reaction mixture

95 °C	3 min	
95 °C	30 s	repeat steps 2-4 18x
50 °C	30 s	
68 °C	11 min	
68 °C	5 min	

Table 2 PCR program

37 °C for 1 hour in the PCR thermocycler to degrade the template. Afterwards, 5 µl was used for heat shock transformation into DH5α *E. coli*.

2.3.1.2 Heat shock transformation into *E. coli*

All the transformations were done with competent DH5α *E. coli* (produced in-house, in the Zhang lab). The *E. coli* were thawed on ice. 1 µl DNA was added directly into the cells and mixed gently by flicking the tube. The mixture was incubated for 20 min on ice before heat shocking the bacteria in a 42 °C water bath for 45 s before immediately putting the tubes back on ice for 2 min. Then 200 µl of SOC media (produced in the Zhang lab) was added and the mixture put on the shaker for 45 min to allow for outgrowth and good resistance against the antibiotic. All the liquid was plated out onto LB agar plates supplemented with 100 µg/ml Ampicillin. The plates were incubated over night at 37 °C until the colonies were big enough to be picked.

2.3.1.3 Plasmid amplification

Mini preps A single colony was picked from LB agar plates with a sterile plastic loop and added to 4 ml LB, supplemented with 100 µg/ml Ampicillin. The bacteria were grown overnight at 37 °C with vigorous shaking. The next day any pellet was mixed back in, glycerol stocks were made if needed (see 2.3.1.5) and the bacteria then distributed into 3 Eppendorf tubes. The mini preps (QIAGEN) were done according to the manufacturer's instruction for each tube separately. LyseBlue was used to ensure proper mixing. The DNA was eluted in the last step with warm 50 µl H_2O. The DNA concentration was measured using a NanoDrop (Thermo Scientific). Any tubes with a concentration of less than 50 ng/µl were discarded.

I did mini preps for all constructs received or produced.

Mega preps A single colony was picked from LB agar plates with a sterile plastic loop and added to 4 ml LB, supplemented with 100 µg/ml Ampicillin. The bacteria were grown overnight at 37 °C with vigorous shaking. The next day 1 ml of cell suspension was transferred into 500 ml LB Amp and shaken again at 37 °C over night. Then, mega preps (QIAGEN) were done according to the manufacturer's instructions. The DNA was eluted with H_2O and the concentration measured using a NanoDrop (Thermo Scientific).

I used mega kits only to produce a greater quantity of the HERV-K_{con} Gag plasmid.

2.3.1.4 Sequencing

All plasmids were sequenced commercially after amplification by mini preps by Eurofins Genomics. All the constructs produced containing single amino acid mutations in Gag were also checked with the same supplier to ensure the incorporation of the correct mutations and to avoid any off-target mutations or deletions.
The following primers, which were ordered from Thermo Fisher in lyophilized form, were used:
Fwd1: TCAAGAAAGGAAGGAGATACTGAG
Fwd2: AAATCGTCTCTCTCACCCTCTC
Fwd3: AAGACCCAGGAAGTACCTGC
Fwd1 was always used. Fwd2 and Fwd3 were only used if the results of the first sequencing run didn't cover the complete Gag sequence.

2.3.1.5 Glycerol stocks

I converted all the constructs I made or received into glycerol stocks for future use. For this, 600 µl of bacterial suspension were mixed with 400 µl sterile glycerol, mixed well and frozen immediately in −80 °C for long-term storage.

2.3.1.6 SDS-PAGE gels

VLP samples from small scale transfections in HEK293T cells were used for SDS-PAGE gel electrophoresis (NuPAGE 4–12 % Bis-Tris, Novex). 12 µl sample was mixed with 4 µl loading dye per lane (NuPAGE SDS Sample Buffer 4x, Invitrogen). All the samples were heated on a heat block at 95 °C for 5 min prior to loading, except the ladder (NED protein ladder). The gels were run in MES buffer at 165 V for 35 min.

The gels were not stained with Coomassie Blue as Gag could not be detected this way. Instead, Western blots were performed to detect the proteins.

2.3.1.7 Western blots

After the SDS gel finished running, it was cut free and the wells cut off. Sponges, filter paper and nitrocellulose membrane were pre-soaked in transfer buffer before the transfer sandwich was assembled as followed: Sponge, filter paper, gel, membrane, filter paper, two sponges. A roller was run over each layer during assembly to avoid air bubbles being trapped. The transfer was run at 10 V for 60 min.

For protein detection, the membrane was first washed in PBST (PBS (Sigma) with 0.1% Tween20 (Sigma-Aldrich) (v/v)) before being blocked in blocking buffer (5% dry milk (Sigma-Aldrich) in PBST (w/v)) for 1 h at room temperature. Then, the membrane was rinsed in PBST and incubated over night at 4 °C with the primary anti-HERV-K Gag antibody (HERM-1841-5, 1:5000 in blocking buffer).

The next morning, the membrane was washed 3 times for 5 min each in PBST. Then, it was incubated with the secondary anti-mouse HRP antibody (A0168 Sigma Aldrich, 1:5000 in blocking buffer) for an hour at room temperature and washed again 3 times with PBST. The gel was imaged using the Clarify Western ECL substrate solutions (Bio-Rad) by mixing 0.5 ml solution A with 0.5 ml solution B before adding the mixture onto the membrane. Pictures were taken using the iBright imaging system (Thermo Fisher) with exposure times up to 15 min.

2.3.2 Mammalian Tissue Culture

2.3.2.1 Cultivation of mammalian cell lines

HeLa and U2OS cells were thawed from vapour phase nitrogen storage by holding the cryo vial under warm tap water. Once thawed, the vial was sterilised with ethanol, transferred into a laminar flow hood and added into a T75 flask with 10 ml DMEM media (Sigma), supplement with 10% FCS (Sigma), 1% non-essential amino acids (NEAA, Gibco), 1% L-glutamine (Gibco).

HEK293T cells were obtained by splitting them off from Strubi stock, which was initially purchased from ECACC and is maintained by Weixian Lu.

Mammalian cell lines were maintained in a sterile environment using a laminar flow hood when handling them. The cells were kept in T175/T75 flasks at 37 °C, 5% CO_2. They were passaged once they reached approx. 95% confluency. For this, the media was removed and the cells washed with Dulbecco's phosphate-buffered saline (DPBS, Gibco). Then 5/3 ml trypsin-EDTA (Gibco) was added and the cells put into the incubator for 5 min to detach them from the flask bottom. After this, 20/10 ml of supplemented DMEM media (10% FCS, 1% NEAA, 1% L-glutamine) was added to stop the reaction. Depending on how long the cells were kept before the next passage, 50-90% of the suspended cells were removed and the flask topped up to 25/12 ml with supplemented DMEM.

2.3.2.2 Transfection of HEK cells

All HERV-K_{con} Gag-Pro-Pol plasmids were kindly provided by Paul Bieniasz. Transfection was carried out with the HERV-K_{con} Gag and GeneJet as a vessel.

Large scale For cryoEM, 10 T75 flasks were seeded with 5×10^6 HEK293T cells each and grown in 12 ml supplemented DMEM media overnight. For each transfection 6.5 µg plasmid was then mixed with 250 µl non-supplemented, blank DMEM and 13 µl GeneJet was mixed with 250 µl non-supplemented, blank DMEM separately. The media on the cells was discarded and replaced with 7.5 ml supplemented media per flask. The GeneJet mixture was then poured into the plasmid mixture (not the other way around) and 500 µl added to the cells. The media was swirled gently to distribute it evenly across all cells and incubated at 37 °C, 5% CO_2 overnight. The next day, 3 ml of supplemented media was added to each flask before incubating for another 24 h to ensure sufficient nutrients for the cells.

Small scale Small scale transfections were done to detect the presence of Gag via Western blots. For this, only 1 T75 flask per construct was seeded. Otherwise, the same protocol was followed as in the large scale transfections. The positive control used the same plasmid HERV-K_{con} Gag as in the large scale transfection. The negative control was generated by mixing the GeneJet mixture with blank DMEM that had no plasmid added.

2.3.3 VLP Purification

Large scale The HERV-K_{con} Gag VLPs were harvested 48 hours post-transfection by spinning the supernatant at 500 g for 5 min and filtering the supernatant through a filter with 0.2 µm pore size to remove large cellular debris.

The filtrate was then transferred into big ultra-centrifuge tubes (Beckman Coulter, Ultra-Clear Centrifuge Tubes, 0.5x2 inch) with approximately 35 ml each. It was underlayed with 3 ml 8% OptiPrep (Sigma-Aldrich, 60% stock solution) in STE (100 mM NaCl, 10 mM Tris-Cl pH 8.0, 1 mM EDTA; Sigma-Aldrich). The tubes were balanced and spun at 100 000 g, 4 °C for 1 h 20 min. After the initial ultra-centrifugation, the liquid was sucked out with lint-free wipes very carefully and the pellets resuspended in the remaining liquid before being pooled together.

Next, one gradient was prepared in a small ultracentrifuge tube by layering 1 ml 30%, 1 ml 20%, 1 ml 10% OptiPrep over each other, starting with the most dense layer and being careful not to disturb the layer lines. The layer lines were marked with a pen before adding the resuspended pellet on top of the gradient. The tube was filled up with STE if needed before balancing. The tubes were spun in an ultracentrifuge at 120 000 g, 4 °C for 2 h 30 min.

After this, the bands were visualized in the dark room with a flash light and each band sucked into a syringe separately by punching a hole into the tube with the help of a needle. Only once all the bands were collected, were the needles pulled out of the centrifuge tubes to avoid the remaining liquid to spill out and disturb other bands. The harvested VLPs were then directly put into separate, small ultracentrifuge tubes. The tubes were filled up with STE and balanced before spinning for 1 h 20 min at 160 000 g, 4 °C to remove residual OptiPrep. The liquid was again sucked away with lint-free wipes and the pellet resuspended in the residual liquid. The VLPs were stored over night at 4 °C for plunge freezing the next day.

The VLP purification was always done in one day as this increased yield over splitting the protocol over two days. One prep yielded enough sample to freeze 4 grids.

Small scale The HERV-K_{con} Gag VLPs were harvested 48 hours post-transfection by spinning 3 ml of the supernatant at 500 g for 5 min and filtering the supernatant through a filter with 0.2 µm pore size to remove large cellular debris.

The filtrate was then transferred into small ultra-centrifuge tubes and underlayed with 1 ml 8% OptiPrep in STE. The tubes were balanced and spun at 150 000 g, 4 °C for 1 h. After this, the liquid was sucked out with lint-free wipes and the pellets resuspended in the remaining liquid

2.3.4 Structural Biology Methods

2.3.4.1 Plunge freezing

Vitrobot At first, a Vitrobot II (Thermo Fisher) was used to plunge freeze grids. Copper 2/2 grids (Quantifoil) grids were glow discharged on medium for 30 s. The VLPs were mixed with 6 nm gold fiducials before plunging. The Vitrobot was set to a humidity of 95%, 20 °C. 3 µl VLPs were applied per grid and blotted for 1–3 s with −15 blot force and 0 s wait time.

GP2 For the GP2, lacey carbon grids, 300 mesh, Cu (Agar Scientific) were used and also glow discharged for 30 s on medium. The VLPs were mixed with 6 nm Au fiducials and 3 µl sample applied onto the carbon side and 0.5 µl sample onto the non-carbon side to ensure proper recognition of the grid by the sensor when blotting. Blotting was always done from the non-carbon side with the settings set to 16 °C, 95% vapour, −185 °C ethane. Blotting times varied from 1 s to 4 s, while the best results were achieved at 3 s.

The following form was used to keep track of grids:

Sample		**Puck**	
Box ID	**Fiducials**	**Dilution**	
Grid type	**Grid batch**	**Glow discharge**	
Humidity	**Temp chamber**	**Temp ethane**	
Blot time	**Delay**	**Sensor sensing**	
Plunging comments:		**Clip box pos:**	**Scope pos:**
1.			
2.			
3.			
4.			

Table 3 **GP2 Plunging form**

2.3.4.2 Data collection

Grids were pre-screened on a Talos Arctica. The data cryoET data used for subtomogram averaging was collected on a Titan Krios equipped with a Selectris energy filter (Thermo Fisher Scientific). Pixel size was set to 1.5 Å, exposure at 1.17 s with a total dose of 127.5 $e^-/Å^2$ and a slit width of 5 eV. The tilt series were collected with a dose symmetric tilt scheme over a range of ±60° and a step size of 3°.

2.3.4.3 Subtomogram averaging

PEET & emClarity The tilt sereies were motion corrected with MotionCor2 and aligned using etomo. Bead tracking was fixed manually when needed to ensure perfect alignment. To generate an ab initio template, a SIRT-like corrected tomogram was generated and points set along the lattice by hand. A rough alignment was done which was used for a reference for a finer alignment using the parameters in Table 4. The template generated was then used for template search in emClarity, followed by cleaning to only include subtomograms inside VLPs. The subtomograms were then averaged and aligned at bin 6 and bin 5.

Constrained single-particle tomography (CSPT) The tilt series were re-aligned and reconstructed as I lost access to the processed data during this time. Astigmatic per-tilt CTF estimation was done using parameters from the alignment (*168*). Initial subtomogram positions were determined as follows: VLPs were manually marked in each tomogram and their radius estimated. The VLPs were then segmented along the CA lattice using semi-automatic geodesic surfaces (*170*, *171*). Subtomogram positions were then set uniformly across the segmentation surface with an inter-unit distance of 96 Å. This yielded 274 994 particles from 794 VLPs. Ab-initio STA was done using EMAN2 (*172*). Constrained single-particle tomography (CSPT) was then used for refinement. After several rounds of alignment and tilt-geometry refinement, duplicate particles as well as bad par-

Cycle number	1	2	3	4	5	6
Angular search range						
Phi Max	180°	45°	30°	7.5°	3.75°	1.875°
Phi Step	30°	15°	10°	7.5°	3.75°	1.875°
Theta Max	0°	45°	30°	7.5°	3.75°	1.875°
Theta Step	1°	15°	10°	2.5°	1.25°	0.625°
Psi Max	0°	45°	30°	7.5°	3.75°	1.875°
Psi Step	1°	15°	10°	2.5°	1.25°	0.625°
Search						
Distance	7	5	5	3	3	3
Hi Freq Filter						
Cutoff	0.15	0.2	0.2	0.25	0.25	0.25
Sigma	0.05	0.05	0.05	0.05	0.05	0.05
Reference						
Threshold	130	130	130	130	130	130

Table 4 PEET angular search parameters

ticles were excluded. Final reconstruction was done with 188 111 particles and yielded a final map with a resolution of 3.2 Å.

2.3.4.4 Structure-based phylogenetic trees

The following CA structures were used for structure-based phylogenetic analysis: 6SSM (HERV-K mature) (*54*), 7ASL (HIV-1 immature) (*12*), 6HWW (MLV immature) (*16*), 6HWI (M-PMV immature) (*16*), 5A9E (RSV immature) (*17*) and the immature HERV-K model we built. The monomers were pre-aligned in Chimera before analysis and multimers reduced to contain only one monomer. The structure-based phylogenetic tree was then calculated using SHP (*173*).

The amino acid sequence alignment was done using the Clustal Omega online tool with MView for colour annotations (*174*).

The pairwise amino acid sequence comparison to calculate the sequence identity and similarity of each retrovirus to HERV-K was done using the EMBOSS Needle online tool.

2.4 Results

2.4.1 HERV-K VLP Characterization

2.4.1.1 Purification of HERV-K VLPs

To be able to compare HERV-K to exogenous retroviruses, I expressed and purified HERV-K Gag VLPs from HEK293T cells as detailed in the large scale method sections 2.3.2.2 and 2.3.3. Ultracentrifugation in the density gradient yielded two bands as seen in Figure 14c and they could be extracted with only minimal disturbance. These bands were plunge frozen separately after a further round of high speed centrifugation to differentiate between them on grids.

Optimization of plunge freezing conditions was needed to yield good quality grids with enough VLPs per grid square to allow cryoET data collections. Switching from double-sided blotting done with the Vitrobot to one-sided blotting with the GP2 increased the amount of VLPs drastically (Figure 14a vs. Figure 14b). The reduction of cells transfected from 10 T175 to 10 T75 flasks and the use of cells at a lower passage could also decrease the amount of background protein showing up as small circles in Figure 14a. Further optimization of the freezing settings was needed to yield good quality ice. Once optimal conditions were found, repeating the complete workflow from transfection to freezing always yielded good grids.

Screening on the Talos Arctica of both bands showed why they were separated during purification. The upper band had mostly fused VLPs (seen in Figure 14d, panel iii) and only a few single VLPs, whereas the lower band consisted mostly of

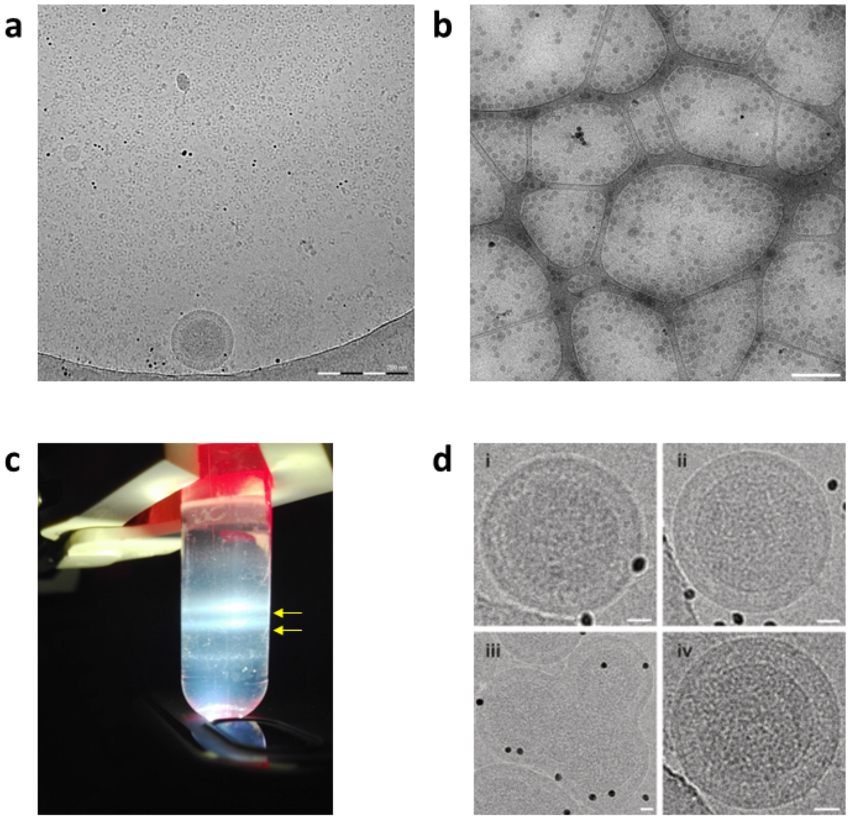


Figure 14 Grid optimization (a) Representative cryoEM image of purified immature HERV-K VLPs plunged with the Vitrobot at medium magnification. (b) Representative cryoEM image of purified immature HERV-K VLPs plunged with the GP2 at low magnification. Scale bar 1 µm (c) Density gradient after ultracentrifugation. Bands visualized with a bright light. (d) Examples of VLPs in the lower band (i and ii) and upper band (iii and iv), indicated by yellow arrows. Scale bars 20 nm.

spherical, non-fused spheres. Only the lower band was used for any further data collections and analysis.

2.4.1.2 Comparison with other retroviruses

HERV-K VLPs are roughly spherical and show a lipid bilayer on the outside and a dense CA lattice layer below (Figure 15a, b) as seen in many exogenous retroviruses. Martin et al. (*10*) characterized several immature retroviruses and measured their dimensions. I followed their protocol on how to measure distances to allow for a comparison as accurately as possible. The mean diameter of HERV-K VLPs is 173 nm, with a range from 120 nm to 210 nm (Figure 15c). HERV-K at 173 nm is thus bigger than other retroviruses, including but not limited to HIV-1, HIV-2, RSV and MLV.

When screening the VLPs, I also noticed that the distance between the capsid lattice and the lipid bilayer seemed bigger than in HIV-1, which was used for subtomogram averaging in our lab before. Measuring this distance gave a mean value of 21 nm with a range from 16 nm to 24 nm, which confirmed a greater distance to HIV-1 (10 nm (*10*)). HERV-K is only slightly larger than the gap in RSV at 18 nm.

The variable distance between the CA lattice layer and the membrane is most likely due to the presence or absence of additional domains between CA and MA. The HERV-K Gag polyprotein has two small proteins, SP1 and p15, located in that space. p15 is needed to recruit host proteins to release virions at the plasma membrane via is late (L) domain motif (*8*, *9*). HIV-1, HIV-2 as well as HTLV-1 Gag polyproteins do not contain any exra domains between MA and CA. MMTV contains the small proteins pp21, p3, p8 and n but they do not encode a late domain even though p8 and n are required for successful assembly (*175*). M-PMV

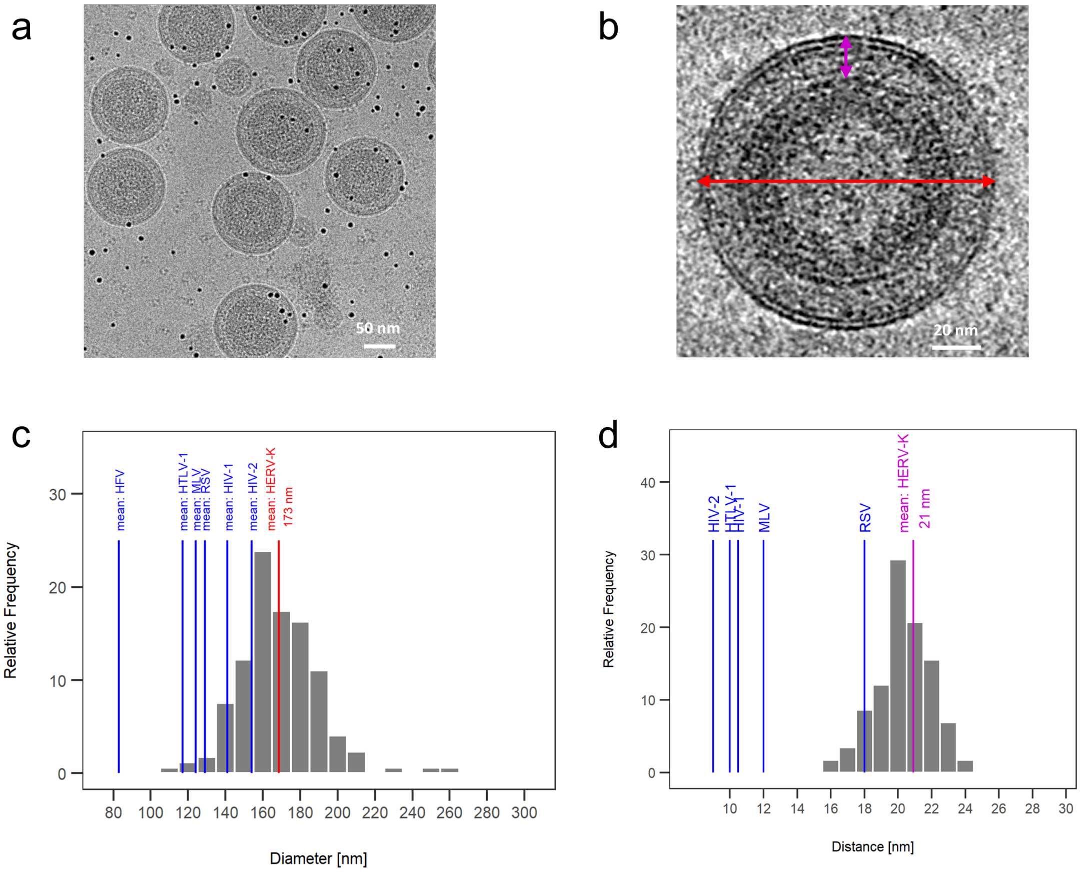


Figure 15 Characterization of HERV-K Gag VLPs Representative cryoEM images of purified immature HERV-K VLPs at medium (a) and high (b) magnifications. The diameter is marked in red and the distance between the lipid bilayer and capsid lattice is marked in purple. (c) Comparison of the mean (red line) and distribution (gray bars) of the diameter of HERV-K VLPs, compared with the mean diameters of other retroviruses (blue lines). (d) Comparison of the mean (purple line) and distribution (gray bars) of the distance between lipid bilayer and capsid lattice of HERV-K VLPs, compared with the mean distance of other retroviruses (blue lines).

encodes pp24 with an L domain and p12 (*176*, *177*). RSV encodes p2a, p2b and p10 with p2b as well as p10 containing L domains (*178*, *179*). MLV contains p12 which also has an L domain (*180*). Consequently, the large distance between membrane and CA lattice corresponds well with these additional domains between

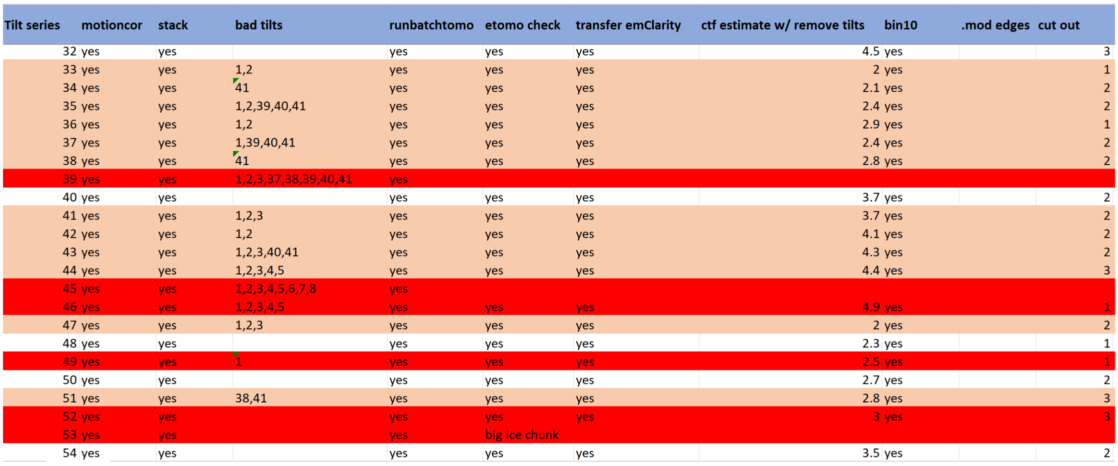
MA and CA.

2.4.2 Structure Determination of immature HERV-K Gag

2.4.2.1 CryoET with Subtomogram averaging

First cryoET data collection Screening of the immature VLPs showed them to be very heterogeneous in size and and morphology. We thus decided to use cryoET combined with subtomogram averaging (STA) to determine the structure of the immature CA lattice.

An initial data set containing 192 tilt series was collected on the Krios with a K2 detector at a pixel size of 1.39 Å/px. The tilt series were motion corrected and aligned using etomo with manual corrections during bead tracking. The stacks were then imported into emClarity, subregions were cut out and template matching performed. However, many tilt series had to have tilts removed due to bad tracking during data collection, as highlighted by the orange and red rows in the excerpt of my spreadsheet (Figure 16), where I kept track of the different processing steps. The removal of these tilts lead to extensive problems in the emClarity pipeline. emClarity was still under development and some steps were not yet tested and programmed correctly to handle the removal of frames. I found that removing frames in tilt series lead to wrong electron dose calculations, angles and bead removal across many steps such as tomogram generation, CTF estimation and template matching. I tried to correct the faulty files and lists but even after making every effort to solve the many problems, I did not manage to solve all the issues. I therefore decided to discard this data set completely.



Tilt series	motioncor	stack	bad tilts	runbatchtomo	etomo check	transfer emClarity	ctf estimate w/ remove tilts	bin10	.mod edges	cut out
32	yes	yes		yes	yes	yes	4.5	yes		3
33	yes	yes	1,2	yes	yes	yes	2	yes		1
34	yes	yes	41	yes	yes	yes	2.1	yes		2
35	yes	yes	1,2,39,40,41	yes	yes	yes	2.4	yes		2
36	yes	yes	1,2	yes	yes	yes	2.9	yes		1
37	yes	yes	1,39,40,41	yes	yes	yes	2.4	yes		2
38	yes	yes	41	yes	yes	yes	2.8	yes		2
39	yes	yes	1,2,3,37,38,39,40,41	yes						
40	yes	yes		yes	yes	yes	3.7	yes		2
41	yes	yes	1,2,3	yes	yes	yes	3.7	yes		2
42	yes	yes	1,2	yes	yes	yes	4.1	yes		2
43	yes	yes	1,2,3,40,41	yes	yes	yes	4.3	yes		2
44	yes	yes	1,2,3,4,5	yes	yes	yes	4.4	yes		3
45	yes	yes	1,2,3,4,5,6,7,8	yes						
46	yes	yes	1,2,3,4,5	yes	yes	yes	4.9	yes		1
47	yes	yes	1,2,3	yes	yes	yes	2	yes		2
48	yes	yes		yes	yes	yes	2.3	yes		1
49	yes	yes	1	yes	yes	yes	2.5	yes		1
50	yes	yes		yes	yes	yes	2.7	yes		2
51	yes	yes	38,41	yes	yes	yes	2.8	yes		3
52	yes	yes		yes	yes	yes	3	yes		3
53	yes	yes		yes	big ice chunk					
54	yes	yes		yes	yes	yes	3.5	yes		2

Figure 16 **First data set** Snapshot of the spreadsheet used to keep track during STA. Tilt series in red rows were discarded completely. Tilt series in orange rows had tilts removed but were still used for further processing steps.

Second cryoET data collection A second data set was collected, which was used for the structure determination of HERV-K immature CA. Data collection and parameters are summarized in Table 5. Its tracking across tilt series was greatly improved with bad tracking occuring in less than 1% vs. more than 80% in the first data set. It also showed greater contrast than the first data set as the data were collected using a Selectris energy filter with 5 eV slit width. Energy filters remove inelastically scattered electrons during data acquisition, which leads to a greater signal-to-noise ratio (SNR) (*181*). The Selectris filter (Thermo Fisher) is very stable and thus allowed the slit width to be set lower than 10 eV and thereby boosting signal as well as contrast.

Sample	HERV-K VLP
Data collection	
Magnification	105 000x
Voltage [kV]	300
Electron dose [e$^-$/Å]	127.5
Defocus range [µm]	-2 to -5
Detector	Falcon4
Energy-filter (slit)	5eV
Super-res mode	N/A
Acquisition scheme	-60°/60°, 3°, dose-symmetric
Frame number	EER format
Pixel size [Å]	1.503
No. of tilt-series/micrographs	124

Table 5 Data collection parameters

Initial subtomogram positioning I chose to avoid using a template of a previously solved retrovirus structure for particle picking to avoid reference bias as the sequence conservation to HIV-1 and other retroviruses is low (see also Figure 26b).

Initially, I tried to pick evenly across the lattice by hand for the initial average, which I then used as a template for the complete data set. However, this proved difficult as the lattice is easy to see in the middle of a VLP but gets increasingly more difficult when slicing up or down through the tomogram.

Still, the hand picked positions were then averaged using PEET to generate an initial template de novo (Figure 17a). This template was then used for particle picking using emClarity. However, the picking was only successful across the center of the tomogram and all of the top/bottom views were missing; hence not the complete sphere was sampled (Figure 17b bottom). Averaging and aligning these positions then also lead to severe streaking in z (Figure 17c).

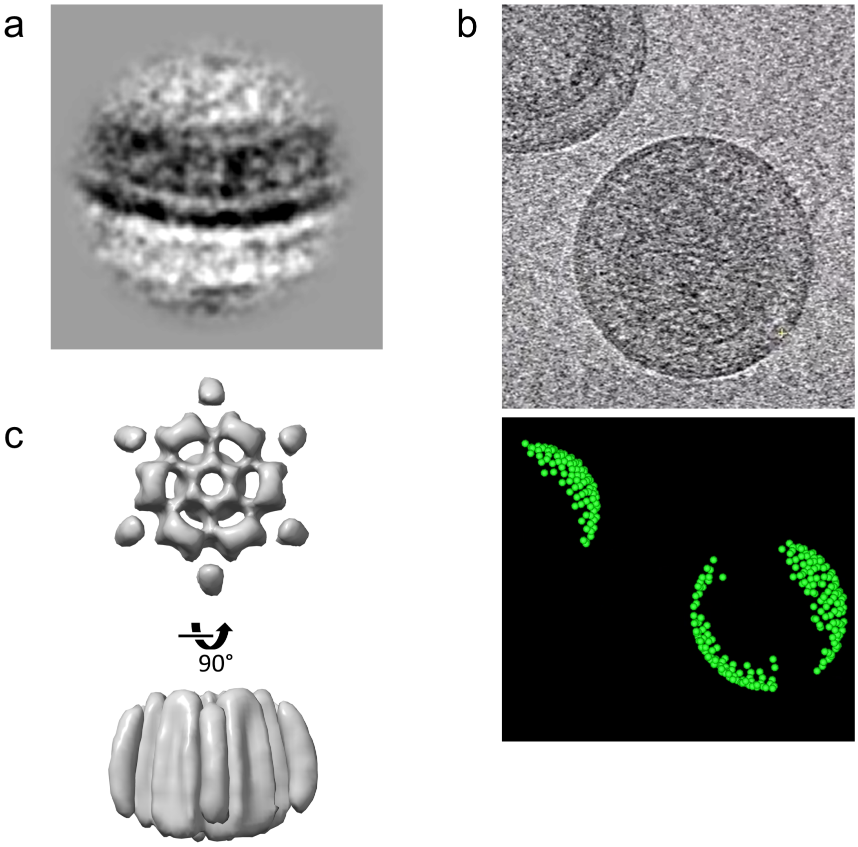


Figure 17 Template generation by picking particles by hand (a) Generated template by picking particles by hand along the CA lattice. (b) Tomogram slice above with the corresponding subtomogram positions below. (c) Bin5 average in top and side view, using the template from (a) for particle picking and subsequent subtomogram alignment and averaging.

Sadly, any further attempts to try to generate a better template and a high resolution map myself by sampling the complete sphere were dwarfed by a hacking incident in the Department of Structural Biology (STRUBI). Consequently, I was unable to access any computers for almost half a year and we reached out to Alberto Bartesaghi to help with this project. They were able to segment the tomograms along the CA lattice and uniformly distribute subtomogram positions along the surface of approximately 800 VLPs (methods in section 2.3.4.3).

Sample	HERV-K VLP
Data processing	
No. of tilt-series/micrographs	124
Symmetry imposed	C6
Final particle images	188 111 sub-volumes / 7 712 551 particle projections
Map resolution [Å]	3.2
FSC threshold	0.143

Table 6 Data processing parameters Data processing was done by Hsuan-Fu Liu and Ye Zhou.

Final map and model Consequent subtomogram averaging followed by high-resolution constrained alignment of 2D projections using C6 symmetry led to a map with a final resolution of 3.2 Å. Data processing paramters are summarized in Table 6. I deposited the map on the Electron Microscopy Data Bank (EMDB) under the accesssion code EMD-16511.

Subtomogram averaging could resolve the CA domain of Gag but no density was observed for MC, NC or any of the small, additional proteins. Sidechains in the map were clearly resolved which made it possible to build an atomic model into the density (Figure 18a/b).

I built and refined one monomer using Refmac and Coot. The mature HERV-K structure (*54*) was used as a starting point. It was cut apart into its two domains and they were fitted separately into the density. After a few rounds of real space refinement, I build the last helix at the C-terminus ab initio, as well as the linker between the two domains and the N-terminus. The monomer was then expanded to a hexamer before running a few more cycles of refinement.

The Perilla lab builds a new model using RosettaCM before they start with molecular dynamics simulations. There were only minor differences in side chain orientations between their model and the one I built.

I deposited the final structure and it can be accessed under the PDB code 8C9M. Table 7 reports the final MolProbity scores for validation.

Sample	HERV-K VLP	
All-Atoms Contacts		
Clashscore*, all atoms	1.49	
Protein Geometry		
Poor rotamers	0	100.0%
Favored rotamers	1152	100.0%
Ramachandran outliers	12	0.88%
Ramachandran favored	1308	96.46%
Rama distribution Z-score	-0.57 ± 0.20	
MolProbity score	1.12	
$C\beta$ deviations >0.25 Å	0	
Bad bonds	104/10956	0.95%
Bad angles	0/14814	0.0%
Peptide Omegas		
Cis Prolines	6/78	7.69%
Low-resolution Criteria		
CaBLAM outliers	42	3.1%
CA Geometry outliers	6	0.45%
Additional validations		
Chiral volumer outliers	0/1596	
Waters with clashes	0/0	0.0%

Table 7 Complete MolProbity validation report MolProbity validation results for the final model. *Clashscore is the number of serious steric overlaps (>0.4 Å) per 1000 atoms.

2.4.3 Structure analysis and comparison to exogenous retroviruses

2.4.3.1 HERV-K_{con} CA structure

The CA monomer consists of an N-terminal domain (NTD) and a C-terminal domain (CTD) with a linker in between and an additional helix 12 at the C-terminus. The NTD consists of 7 and the CTD of 4 helices (Figure 18c). The backbone, including the linker, is well resolved across almost the complete CA monomer. There are only 21 amino acids missing at the N-terminus.

Six CA monomers assemble together to form a hexamer, which is held together by the 6-helix bundle (6HB) formed by helix 12 (Figure 18d). At the top of the 6HB is a strong density (Figure 18e-g). This density is most likely an IP6 molecule; similar to HIV-1 where an IP6 molecule is also present in the same position and neutralizes the positive charges of two lysine rings. These lysine residues are also present in HERV-K (K158 and K227), thus probably providing the same function by interacting with the IP6 to stabilize the 6HB. Hydrophobic residues further downstream in the 6HB face into the inside of the hexamer and further stabilize the bundle.

Adjacent hexamers are held together by forming interactions at the trimer and dimer interfaces. In HERV-K, the trimer interface is formed by hydrophobic interactions involving helix 4 in the NTD and the dimer interface is held together by helix 9 in the CTD (Figure 18h-j).

There are no NTD-CTD interactions formed within the monomer or hexamer,

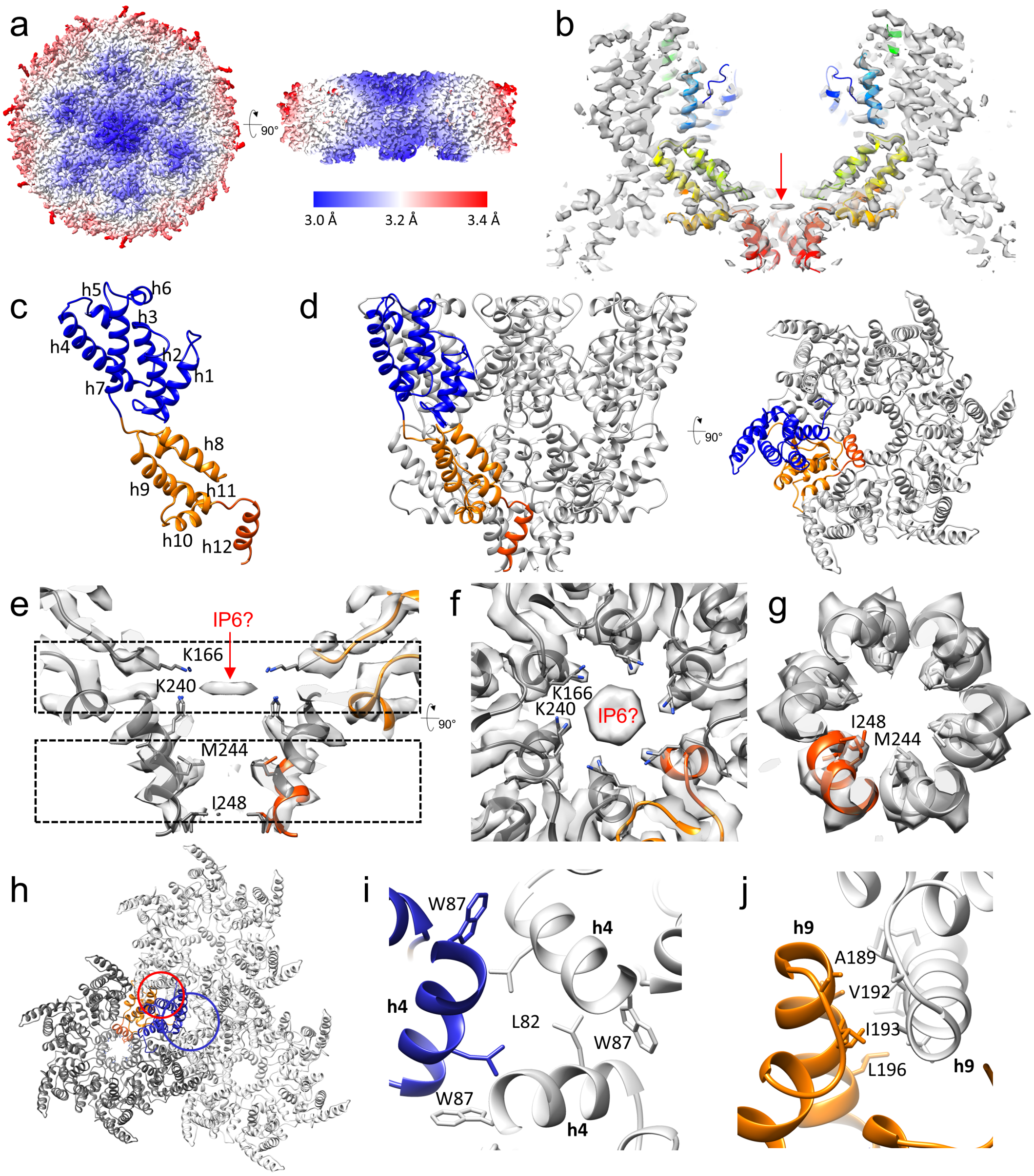


Figure 18 CryoET STA of immature HERV-K VLPs (a) CryoET STA map of immature HERV-K hexamer colored by local resolution (red to blue). (b) A slab of HERV-K hexamer map overlapped with the protein model (rainbow colors blue to red, N-terminus to C-terminus. (c) The protein model of the immature HERV-K monomer (NTD in blue, CTD in orange) with α-helices labeled. (d) The protein model of the immature HERV-K hexamer viewed from the side and top with one monomer colored in blue and orange. (e) A central slice of the HERV-K hexamer density map, showing IP6 density (red arrow) and two Lys rings (K166 and K240), in addition to the hydrophobic core of the 6HB (M244 and I248). (f-g) Density slabs at the top (f, top dashed box in e) and bottom (g, bottom dashed box in e) of the 6HB, with the protruding side chains shown as sticks. (h) The dimer (red circle) and trimer (blue circle) interfaces between adjacent immature HERV-K hexamers. (i) A detailed view of the trimer interface involving helix 4 with the contributing side chains shown as sticks. (j) A detailed view of the dimer interface involving helix 9 with the involved side chains shown as sticks.

similar to other retroviral immature CA lattices (*17*).

2.4.3.2 HERV-K immature Gag interface analysis

The interactions between hexamers were tested using single amino acid mutations followed by detection via Western blots, and molecular dynamics simulations.

Western blots The aim of incorporating single amino acid mutations into CA at specific loci and subsequent detection of Gag was trying to find which side chains are necessary to form VLPs. The mutations to alanine were incorporated using site-directed mutagenesis PCR and confirmed by sequencing the complete gene. The mutated Gag was then expressed in HEK293T cells at a small scale and the VLPs purified with a single ultracentrifugation step. The pellet was then run on an SDS-PAGE gel for subsequent detection by Western blot analysis using primary anti-HERV-K Gag antibodies. Many rounds of optimization were needed to yield a satisfying signal (see Figure 19 for all the reagents tested during optimization iterations).

Y66, E69, K85, D90, R97, R100, Q118, S121 and T122 were thought to be involved in forming the trimer interface and thus each amino acid was mutated to alanine separately. Western blots confirmed that mutations at positions Y66, E69, K85, D90 and R97 did indeed abolish VLP production and Gag could not be detected in the media. However, the other four sites did not reduce signals when mutated and are thus not essential to form interactions at the trimer interface (Figure 20a).

- Mutations
  - DMSO 🟢
  - dNTPs 🟢
  - Buffer 🟢
  - Pfu Turbo 🟢 🟢
  - Plasmid
  - Primers
  - H_2O
  - PCR program 🟢

  - DpnI 🔴
  - incubation time 🟢 🟢

  - DH5alpha E.coli
  - Transformation protocol

  - Sequencing primers

- Expression in cells
  - T75 flasks 🟣
  - HEK293T cells
  - DMEM 🟣
  - Glutamine
  - NEAA
  - FBS
  - Trypsin
  - Trypan Blue
  - GeneJetII 🔴
  - purified plasmid 🟣 🟣

- Western
  - RIPA lysis buffer 🟢 🟢 🟢 🟢 🟢
  - Protease inhibitor 🔴

  - OptiPrep in STE 🟢 🟢 🟢 🟢

  - Protein ladder 🔴
  - SDS gels 🟣

  - membrane
  - 1ary ab 🟢 🟢 🟢 🔴 🟢
  - 2ary ab 🟢
  - PBST
  - blocking solution

  - ECL solutions 🔴
  - Imaging 🟢 🟢

Figure 19 Western blot optimization Reagent list of the most important reagents used for mutation PCR, small scale expression and Western blot. Green dots mark reagents that were optimized, purple dots mark reagents with supply issues, red dots mark reagents that were found to be not functional and thus had to be replaced.

Molecular dynamics simulations To confirm the results from the mutational analysis and to test the stability of the interactions in the lattice, we decided to use molecular dynamics simulations and probe how much side chains move during a 200 ns simulation. So, after building a new model of the hexamer with RosettaCM (*182*), a hexamer of hexamers (one hexamer surrounded by 6 hexamers) was used to run two simulations. The results could then be analyzed to find more stable side chains and bring the other findings into a better context.

In the helix 4 trimer interface R100 and Q118 were found to be most flexible (Figure 20b); where flexible side chains tend to be less involved in stabilizing the interface. Stable interactions were instead formed by charged and hydrophobic residues in helix 4, such as E69, Y66, K85, D90, R97 and L82, W87 with low RMSF values (Figure 20c,d). Together with the Western blot results, we could thus prove that only the top of helix 4 is involved in forming the trimer interface, without the involvement of R100 and Q118 further downstream. The dimer interface was

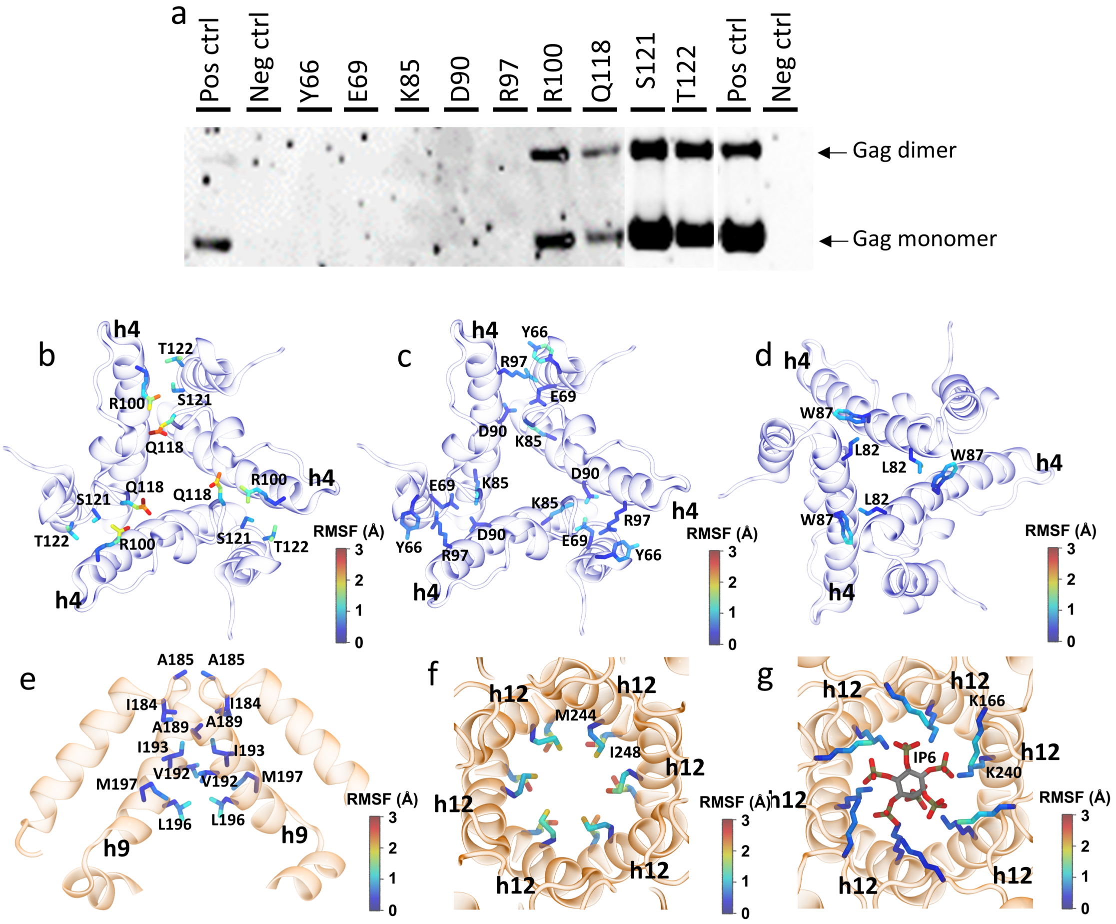


Figure 20 MD simulations and mutational validation of immature HERV-K intermolecular interactions (a) Western-blot of VLP assembly carrying trimer interface mutants of HERV-K Gag (b-d) Snapshots of the trimer interface involving helix 4 with the contributing side chains shown as sticks. Residues in (c) and (d) show stable interactions. (e) A snapshot of the dimer interface involving helix 9 with the contributing side chains. (f-g) Views of the 6HB with the contributing side chains of the hydrophobic core (M244 and I248) (e) and the K166 and K240 rings with the IP6 molecule (f). In each snapshot, protein is represented as cartoon (NTD in blue, CTD in orange). The amino acid side chains are colored based on the heavy-atom RMSF computed throughout the MD simulation. Subfigures b-g were prepared by Ye Zhou and Juan Rey.

also confirmed by low RMSF values and shown to be formed by hydrophobic interactions between helices 9 of neighboring CTDs (Figure 20e).

K166 and K240 lysine rings, which were thought to bind an IP6 molecule and stabilize the 6HB, were indeed more flexible when a second simulation without a docked IP6 molecule was run (Figure 20f, g). Also, helix packaging was less tight in the 6HB without IP6 and the 6-fold symmetry was lost, but the hexamer did not completely fall apart.

2.4.3.3 Comparison to mature HERV-K

The mature structure of HERV-K CA was solved by Acton et al. (PDB accession code 6SSM). A comparison of their mature structure with our new immature form reveals how the lattices differ (please note I count the helices differently than in their paper to comply with retroviral conventions and my immature structure). The RMSD is small when aligning the NTD and CTD separately, but high when done simultaneously (Figure 21a). The helices and their orientations in the separate domains do not change; with the exception of the additional helix 12 at the C-terminal end in the immature structure. The differing orientations of the NTD and CTD are caused by a different angle of the hinge in between the domains. Thus, neither the complete dimeric nor the trimeric interactions are shared between the lattices. Such conformational change has also been observed in HIV-1 (*34*).

The immature lattice is more tightly packed and more vertically elongated than the mature lattice. This can be achieved by the location of the NTD and CTD on top of each other but also causes the loss of any interactions between the domains (Figure 21b-e).

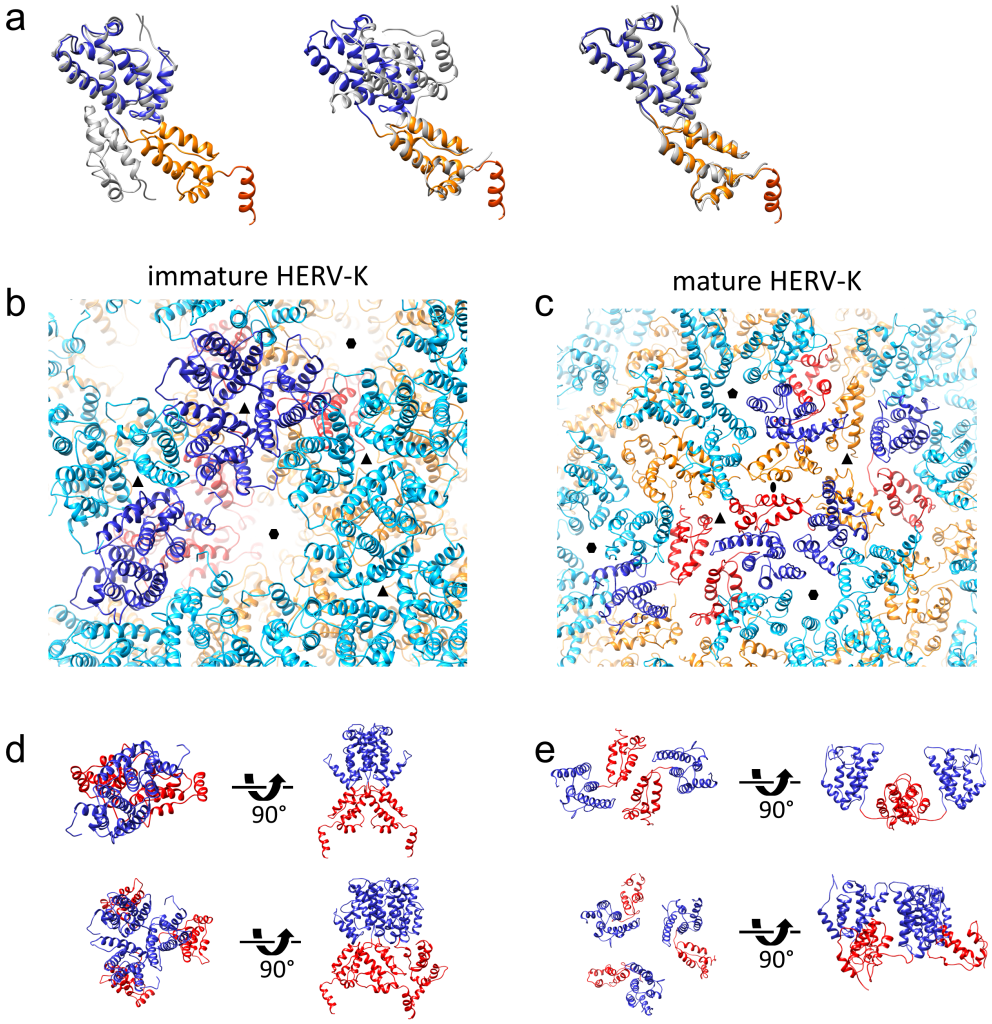


Figure 21 Comparison between immature and mature HERV-K. (a) Overlay of the immature HERV-K structure (NTD blue, CTD orange) with mature HERV-K (grey, PDB: 6SSM), aligned to the NTD (left), CTD (middle), and NTD/CTD separately (right). (b-c) Arrangement of HERV-K immature (b) and mature (c) lattices. NTD is colored in blue/cyan, CTD in orange/red. (d-e) Detailed views of the dimer (top) and trimer (bottom) interface in the immature (d) and mature (e) lattices.

The dimer interfaces in both the immature and mature lattice are formed by helix 9, similar to HIV-1. Some interacting side chains (A189, V192, I193) are conserved between the two forms (*54*) but the crossing angles of helix 9 differ (−50° immature vs 95° mature) which causes a different interface overall. The similarity between HERV-K and HIV-1 in both the immature and mature interfaces is striking and hints at a conserved structural maturation process (*1*, *130*).

2.4.3.4 Comparison to other immature retroviruses

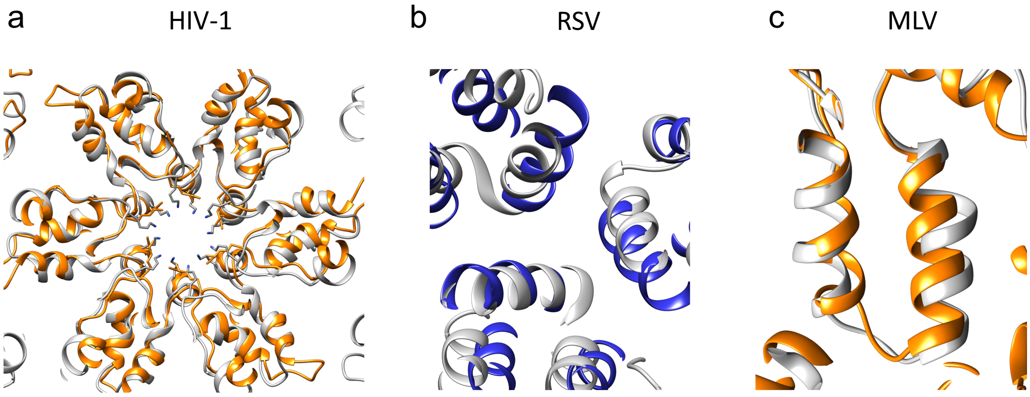


Figure 22 Conservation among the immature retroviruses Structural similarity of immature CA intermolecular interfaces, compared with HIV-1 (hexamer interface, a, PDB: 7ASL), RSV (trimer interface, b, PDB: 5A9E) and MLV (dimer interface, c, PDB: 6HWW). HERV-K structure is colored in blue and orange, other retroviruses in gray.

IP6 IP6 was found to be present in several immature exogenous retroviruses at the top of the 6HB, the same position as the extra density in HERV-K. Sequence comparisons between HERV-K and HIV-1 show a striking resemblance in their terminal helices and both bind IP6 by two lysine rings (Figure 22a). RSV also binds IP6 in a similar pocket but instead of using two lysine rings, it does so with one lysine and one arginine ring (*35*). Surprisingly, M-PMV, the only other immature structure of a betaretrovirus solved, does not have the 6HB at its C-terminus. Still, the CA-NC junction is essential for infectivity as well as assembly and maturation (*183*). It might be that the 6HB is a feature that was not needed for betaretrovirus replication in the cytoplasm and thus lost while the viruses evolved.

Lattice defects Mapping of the subtomogram positions back to the original tomograms shows a hexagonal lattice with defects to accommodate its curvature into a sphere. This type of immature CA lattice can be found in many modern retroviruses as explained in section 1.3.1.1. It thus seems to be a conserved feature in retroviruses that not only spans across retroviral genera but must also have evolved earlier than when HERV-K initially integrated into the human genome.

Lattice arrangements Comparisons between different immature CA structures shows HERV-K to be very similar to M-PMV and RSV but deviates from HIV-1 and MLV (Figure 23 and Figure 24). This difference in CA domain orientation of M-PMV and HIV-1 has also been previously reported (*34*).

Many retroviruses form dimeric interactions in the CTD via their helices 9 i.e. HERV-K, HIV-1, M-PMV as well as the least similar MLV (Figure 22c). MLV additionally uses its 3_{10} helix and the base of helix 9 to form more stable interactions (*16*). Interactions in the NTD differ however. MLV requires helix 7 and the loop between helix 4/5 (*16*). RSV utilizes helices 2 and 7 as well as the upstream domain p10 to form the dimeric interactions (*184*). HIV-1 uses helix 1 and helix 2 to form the dimeric and trimeric interfaces (*17*).

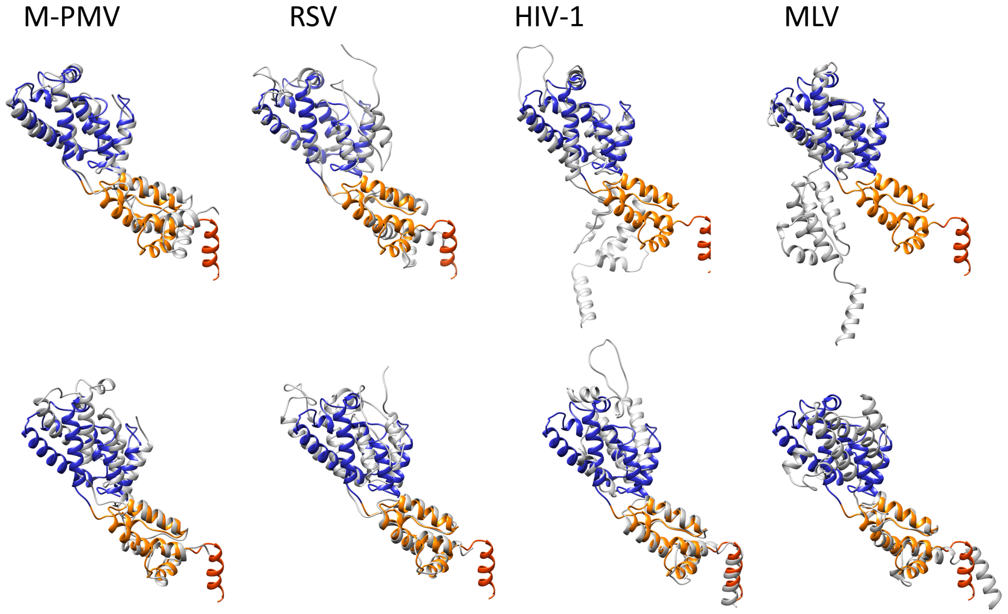


Figure 23 Conservation among immature CA monomers Overlay of the immature CA monomer of HERV-K (NTD blue, CTD orange) with those of other retroviruses (grey) aligned to the NTD (top) or NTD/CTD separately (bottom), shown are M-PMV (PDB: 6HWI, betaretrovirus), RSV (PDB: 5A9E, alpharetrovirus), HIV-1 (PDB: 7ASL, lentivirus), MLV (PDB: 6HWW, gammaretrovirus).

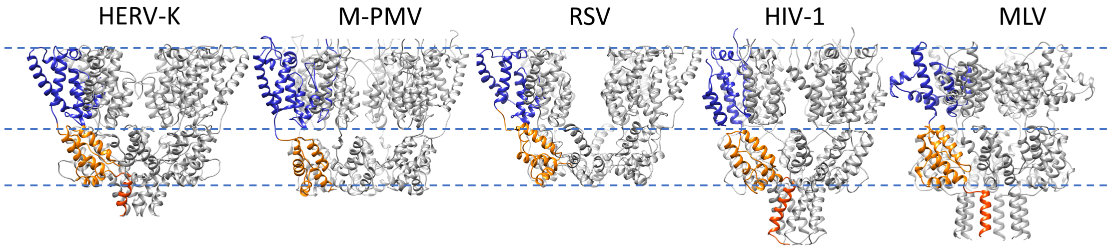


Figure 24 Conservation among immature CA hexamers Comparison of the immature CA hexamer structures with one monomer each colored with the NTD in blue and CTD orange, shown are HERV-K, M-PMV (PDB: 6HWI, betaretrovirus), RSV (PDB: 5A9E, alpharetrovirus), HIV-1 (PDB: 7ASL, lentivirus), MLV (PDB: 6HWW, gammaretrovirus).

2.4.3.5 Structure-based phylogenetic analysis (SBPA)

Structure-based phylogenetic trees can be used to better understand the similarities between structures without input from their amino acid sequences. It thus allows a comparison without bias from how closely related the viruses are and shows structural conservation patterns despite low sequence conservation (Figure 26). SBPA utilizes medium-to-high resolution structures for a pairwise comparison to generate structural distances. Hence, it can add information to sequence-based approaches and is especially helpful when analyzing evolutionary patterns on not closely related proteins with low sequence homology. The pairwise structural alignment done during SBPA calculates the equivalence between two superimposed structures. SHP additionally also uses the shape of the polypeptide chain and a rotation function to calculate the relative orientation of two structures (*173*, *185*). I used SHP (*173*) for calculations and plotted an unrooted tree with the branch length representing the distance between protein structures.

I performed a structure-based phylogenetic analysis on immature CA monomers from retroviruses in different families (Alpharetrovirus: RSV, Lentivirus: HIV-1, Gammaretrovirus: MLV, Betaretrovirus: M-PMV, Betaretrovirus-like: HERV-K) as well as the mature HERV-K CA structure for both the complete monomer and the NTD/CTD separately (Figure 25).

A structure-based phylogenetic analysis of just the CTD or NTD can be useful for a direct comparison of the isolated domains but can be problematic when looking at the overall picture. Especially when looking at the CTDs with just minute differences between retroviruses, even small changes can lead to different

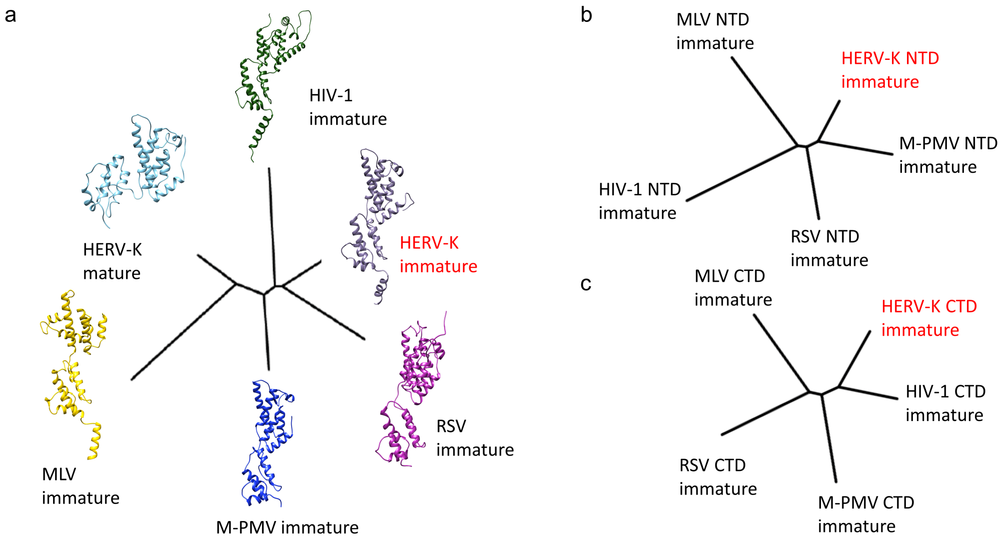


Figure 25 **Phylogenetic analysis** Structure-based phylogenetic tree, calculated using SHP (*173*) of (a) the complete CA monomer, (b) just the NTD or (c) just the CTD.

results. Technical problems when building the protein model (especially in the last helix that forms the 6HB and is missing in the RSV model) might thus skew the results when looking at the CTD in isolation. Only when analysing the complete monomer do we take into account the linker between the two domains. And with the importance of this linker for the orientation of the NTD/CTD relative to each other and subsequently for the interactions formed between monomers and hexamers, only then do we see the complete picture of how similar two viruses are in their CA lattice.

When looking at the SBPA done on the complete CA monomer, the mature HERV-K CA is located furthest away from the immature HERV-K CA. It obviously has exactly the same amino sequence as its immature form and thus reflects that the calculation is not based on sequences information at all. Its location can be explained by the details provided in the previous section, especially the different

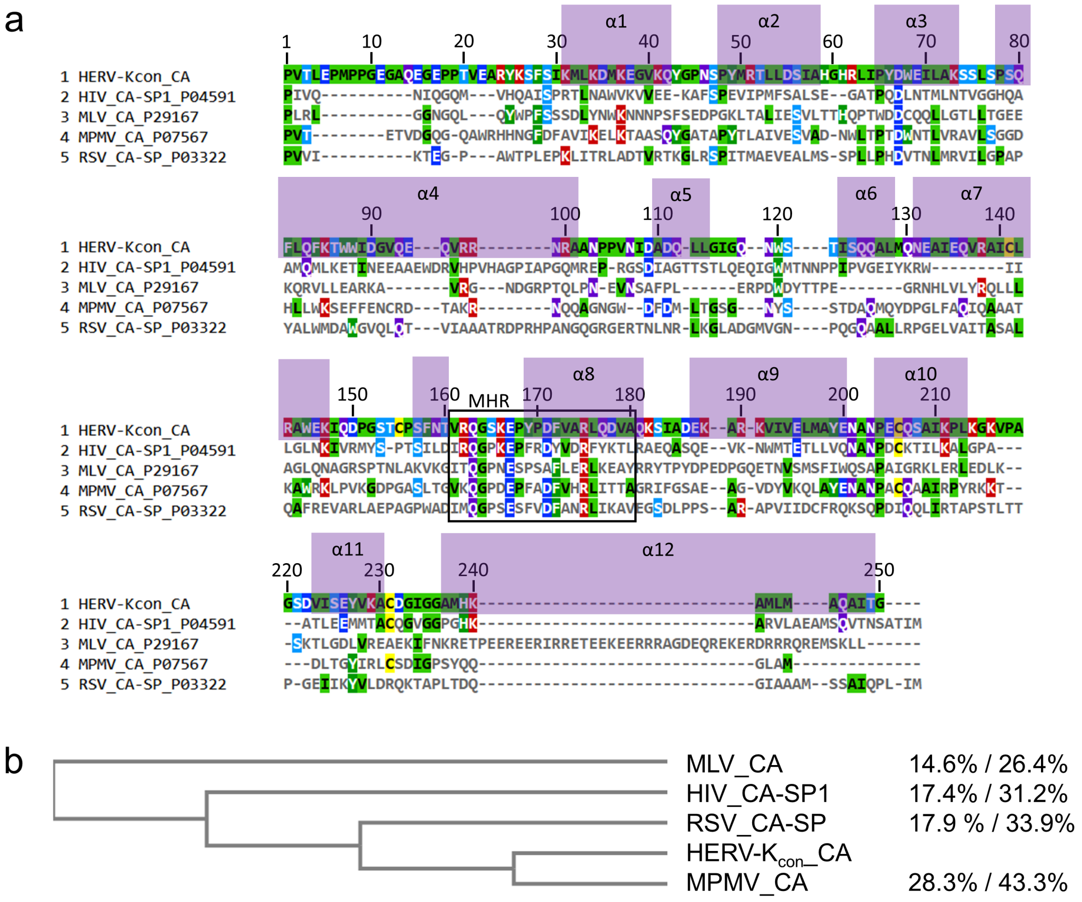


Figure 26 Sequence comparison of retroviruses (a) Protein sequence alignment of CA (and the short spacer protein after CA if it exists) of the same retroviruses as in Figure 25. The helices of the immature HERV-K$_{con}$ are labelled in purple boxes. (b) Phylogram only based on the sequences shown in (a). The sequence identity / similarity to HERV-K is shown on the right for each branch.

hinge angles connecting the NTD and CTD. Surprisingly though, M-PMV is also not located next to the immature HERV-K model. M-PMV is also in the betaretrovirus family and, as HERV-K was initially classified based on its sequence, most closely related to HERV-K (see also Figure 26b). M-PMV has a similiar NTD as HERV-K, reflected in its proximity to HERV-K on the SBPA of the NTDs, which results in a similar trimeric interface as well. However, its CTD helix assembly lacks the C-terminal helix normally used to form the 6HB. RSV is the structurally closest retrovirus to HERV-K. Indeed, it overlaps well and its trimeric interface is almost the same as well (Figure 22b). Even though these retroviruses are very similar at the monomeric level, the quaternary assemblies still differ. MLV is placed the farthest apart from HERV-K. This is caused by a poor superimposition of both the CTD and NTD (Figure 23) and is reflected in its distance even when taking the CTD and NTD into account separately (Figure 25b/c). This distance could also be observed when comparing mature forms (*54*).

2.5 Discussion

Studying an endogenous retroviruses, or more specifically its consensus sequence, is like studying a fossil. It helps us understand the architecture of extinct retroviruses and allows for a comparison to other retroviruses. We were successful in elucidating the structure of the immature lattice of HERV-K by using an approach using cryoET and STA. The atomic model shows a well defined NTD, CTD and an additional helix at the C-terminus. The CA monomers assemble together to form a hexameric lattice, whose formation was conserved during evolution. Interestingly, it also has a small molecule at the top of its 6HB, which we propose to be IP6. This molecule can help in the assembly of the immature virion, similarly to how it was already found in modern retroviruses. The structural rearrangements needed to form the mature HERV-K lattice are also similar to other retroviruses. Molecular dynamics simulations and the detection of mutant proteins allowed us to verify the dimeric and trimeric interactions needed to hold the lattice together. A structure-based phylogenetic analysis showed a striking structural conservation. It suggests a similar assembly and maturation process across retroviral genera and evolutionary time.

Chapter 3

SARS-CoV-2

3.1 Disclaimer

The work reported in this project is part of a few big collaborative projects. Two papers have already been published ((*186*) and (*187*)) and a third one is submitted. Due to the size of these projects, the following parts were done by others: Luiza Mendonça performed the infections and prepared the cryoEM samples. Andrew Howe collected the cryoET data with the help of Julika Radecke, Tao Ni, Yanan Zhu, Luiza Mendonça and Helen Duyvesteyn. 3D reconstruction and subtomogram averaging was done by Luiza Mendonça and me in (*186*), and by Andrew Howe, Luiza Mendonça, Julika Radecke, Yuewen Sheng, Vivian D. Li, Tao Ni, Dapeng Sun and me in (*187*), and Yanan Zhu, Yuewen Sheng, Pranav Shah, Tao Ni and me for the 19E/19E6 ChAdOx1 vaccine work with each person aligning and cleaning one data set. Tao Ni performed cryo-ET subtomogram classification. Elizabeth R. Allen, Luiza Mendonça, Alexandra Spencer, Susan Morris performed the FACS experiments. James B. Gilchrist did targeted cryoFIB milling of lamellas. Laura C. Zanetti-Domingues, Benji Bateman and Marisa L. Martin-Fernandez collected serial cryoFIB/SEM volume imaging data. Long Chen, Laura C. Zanetti-Domingues, Dapeng Sun and Marta Szynkiewicz did 3D segmentations of tomograms and cryoFIB/SEM volume data. Ilias Kounatidis, Mohamed A. Koronfel and Luiza Mendonça acquired and processed soft X-ray cryo-tomographic data.

3.2 Introduction

The coronavirus disease 2019 (Covid-19) is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). It caused a global pandemic, killing millions and also led the United Kingdom to go into lockdown in March 2020. This had caused my studies to be interrupted as laboratory access was not allowed for months. As my main project was stalled during this period, I joined collaborative projects to better understand this new virus including structural characterization of ChAdOx1 Covid-19 vaccines and multi-scale imaging of SARS-CoV-2 infection. This chapter reports the work I contributed to these projects and highlights how it integrates into the data collected by the rest of the team.

3.2.1 Viral proteins

SARS-CoV-2 is classed as a novel betacoronavirus in the subgenus *Sarbecovirus*. β-coronaviruses are enveloped viruses with a large positive-stranded RNA genome. They encode several viral proteins with the main structural protein being S, E, N and M. Next to those, several non-structural proteins are produced as well.

Spike (S) glycoprotein The spike protein is transcribed from the *S* gene. S needs to assemble into a homotrimer to function. Each monomer consists of an ectodomain, a transmembrane anchor and a short intracellular tail (*188*). SARS-CoV-2 spike proteins contain two protein cleavage sites. The furin cleavage site is located between the two subunits and processing occurs in the Golgi during maturation which yields an S1 subunit at the N-terminus and a S2 subunit at the C-terminus with a transmembrane region anchor (*189*). The second cleavage site in the second subunit S2 is needed for cell entry by exposing a hydrophobic fusion peptide upon cleavage by cell surface proteases such as TMPRSS2 (*190*). S1 consists of an N-terminal domain (NTD) and a receptor-binding domain (RBD) and two C-terminal domains. S1 binds the host receptors, whereas S2 acts like a spring needed for membrane fusion during cell entry.

The spike protein has two major states. The prefusion state can have the RBDs either in down or up conformations with ACE2 only being able to bind RBDs in their up conformation (*191*). The postfusion state however lacks S1. Thus, once S1 has been shed, S cannot revert back to the prefusion state and this transition must be tightly controlled. Both the prefusion (*191*, *192*) and postfusion (*192*, *193*) structures, as well as several spike proteins bound to antibodies or the receptor ACE2 (*191*, *194*) could be determined to a high resolution by cryoEM studies. However, most were solved from purified complexes and not studied in a cellular context.

As the pandemic progressed, the SARS-CoV-2 S protein has continued to evolve. The balance between stability and its ability to fuse with the PM have shifted and will continue to do so while the immunity in the human population increases. The viral infectivity depends on the stability of S, its proteolytic processing, its ability to sample the different RBD states, the binding strength with cell receptors as well as other factors (*190*). Most of the vaccines on the market as of 2023 are targeting the spike protein. The rapid evolution of S thus increases our need to adapt and evolve the vaccines as well.

Nucleocapsid protein (N) The genomic viral RNA (vRNA) binds to the nucleocapsid protein inside the virus to form ribonucleoprotein complexes. It is needed to encapsulate and protect the RNA.

Envelope protein (E) E is the smallest structural protein. It is only present in a few copies in the virion and can only be found intracellularly in the ER, Golgi and ER-Golgi intermediate compartment (ERGIC) during viral expression (*195*). Comparisons with the SARS-CoV-1 E protein suggest that it acts as an oligomeric ion channel. It might thus be important to regulate the ion balance across the membrane and is involved at many steps during the viral life cycle (*196*).

Membrane protein (M) M is a transmembrane protein, embedded into the viral envelope. It contains three transmembrane domains and the N-/C- termini are exposed inside/outside virion. It binds to N, E and S is thus essential to the formation of the viral core and promotes assembly (*197*).

3.2.2 Life cycle

Cell entry SARS-CoV-2 recognizes its target cells by binding a receptor called angio-tensin-converting enzyme 2 (ACE2) (*194*, *198*). The virus can use two different pathways to infect a cell, depending on the availability of proteases (*199*). If a protease is present at the PM, viruses can use an early pathway (*200*). For SARS-CoV-2, the protease used is TMPRSS2 (*201*, *202*). However, in the absence of a protease at the PM, SARS-CoV-2 can also use a late pathway and fuse instead at the endosomal membrane. This is achieved by using the clathrin- and non-clathrin-mediated endocytosis pathways (*200*, *202*). Upon cleavage, S inserts its fusion peptide into the membrane, which allows fusion with the envelope surrounding the virus (*192*, *193*, *203*). The virus then releases its contents into the cytoplasm and it can start synthesizing its transcripts and proteins.

Transcription & Translation The non-structural proteins 3, 4 and 6 (nsp3, nsp4 and nsp6) are part of the replicase complex and can induce the formation of double membrane vesicles (DMVs) (*204–206*). These interconnected compartments are the main sites where replication of the viral genome, RNA synthesis and its transcription take place (*207*, *208*).

Assembly & Egress The virions are assembled at intracellular membranes derived from the ER, Golgi and ERGIC. Viral release is thought to happen through exocytosis or lysosome pathways (*207*, *209–211*).

3.2.3 Aim

Most vaccines were developed to raise an immunity response against the spike glycoprotein. We sought to characterize the spike protein expressed from the adenoviral ChAdOx1 nCoV-19/AZD1222 vaccine. This vaccine is derived from a replication-defective adenovirus and can generate a generate humoral and cellular immune response to the encoded protein (*212*). Especially important was trying to determine if the proteins were being displayed in the pre- or postfusion confirmation as an immune reaction against the prefusion state is more likely to be effective. Furthermore, we looked at differences in the S structure in the E and E6 (hexaPro stabilized) vaccine construct against the Beta/South Africa SARS-CoV-2 variant.

Also, SARS-CoV-2 assembly and egress is a poorly understood part of the viral life cycle. It is hard to target and difficult to study with high resolution techniques as it is partially buried inside the cell. To target this we wanted to use a correlative multi-modal multi-scale cryo-imaging approach. This approach allows us to combine low resolution techniques (serial cryoFIB/SEM and X-ray tomography) to gain an overview of the cell with high resolution techniques (cryoET of cell edges and lamellas) to study individual spike proteins inside and outside the cell. Thus, it helps us understand egress in a near-native cellular context and reveals complete pathways of SARS-CoV-2 egress.

3.3 Methods

3.3.1 Sample preparation

ChAdOx1 nCov-19 derived SARS-CoV-2 EM grids (G300F1, R2/2, gold, Quantifoil) were glow-discharged and treated with fibronectin for 30 min before being washed with PBS aund UV-sterilized for 1 h. 1.6×10^5 U2OS or HeLa cells were seeded and grown for 24 h at 37 °C. ChAdOx1 nCoV-19 was then added at MOI 1 or MOI 10 and the cells grown for another 48 h.

SARS-Cov-2 England/02/2020 1.6×10^4 Vero Ccl-81 cells (ATCC) were seeded on the carbon-side of G300F1 R2/2 gold grids which were glow-discharged and pre-treated with fibronectin. Infection of SARS-CoV-2 England/02/2020 was done at different MOIs. The cells were incubated at 37 °C for 24 h after infection in DMEM (Merck) supplemented with 1% FCS, 10 units/ml penicillin (Gibco), 10 µg/ml streptomycin (Gibco), 2 mM L-glutamine (Gibco). The cells were fixed with 4% paraformaldehyde in PBS for 1 h at room temperature.

3.3.2 Plunge freezing

Grids were plunge frozen using a GP2 (Leica) with 1 µl of 10 nm gold fiducials (Electron Microscopy Sciences) applied to the back. Grids were blotted from the back before plunge freezing.

3.3.3 CryoET data acquisition

Tilt series were collected on a FEI Titan Krios G2 (Thermo Fisher Scientific) 300 kV electron microscope equipped with a Gatan BioQuantum energy filter and post-GIF K3 detector (Gatan). Tilt series were collected using SerialEM with zero-loss imaging with a slit width of 20 eV. Pixel size was set to 1.63 Å or 2.13 Å with defocus values ranging from −2 µm to −7 µm. Lamella were collected at 50 µm defocus and 7.58 Å pixel size. Tilt series were collected using a grouped dose-symmetric tilt scheme with a range of ±60° at 3° increments in groups of 3.

3.3.4 CryoET image processing

All the frames of each tilt series were motion corrected using MotionCor2. Tilt series were aligned using eTomo from the IMOD suite, using the gold fiducials to enhance tracking. Then, the tilt series stacks as well as the corresponding information were transferred to emClarity 1.5.3.03 for subtomogram averaging.

3.3.5 Subtomogram averaging

Subtomogram averaging was done using the workflow described in (*26*). CTF estimation was performed for each tilt. Initial subtomogram positions were picked by using the template matching function. EMDB-21452 was used as a reference which was low-pass filtered to 40 Å. Template matching results were cleaned manually by overlaying the tomogram with the points in IMOD and plotting the directions of each subtomogram position in Chimera. After subtomogram cleaning, the subtomograms were separated into two independent half-data sets before they were averaged and aligned with 3x, 2x and 1x binning using threefold symmetry. 3 iterations were performed at each binning step with decreasing search angles and translational shifts. The final reconstructions were done using cisTEM. Model fitting was done using Chimera.

3.4 Results

3.4.1 ChAdOx1 nCoV derived spike protein structure

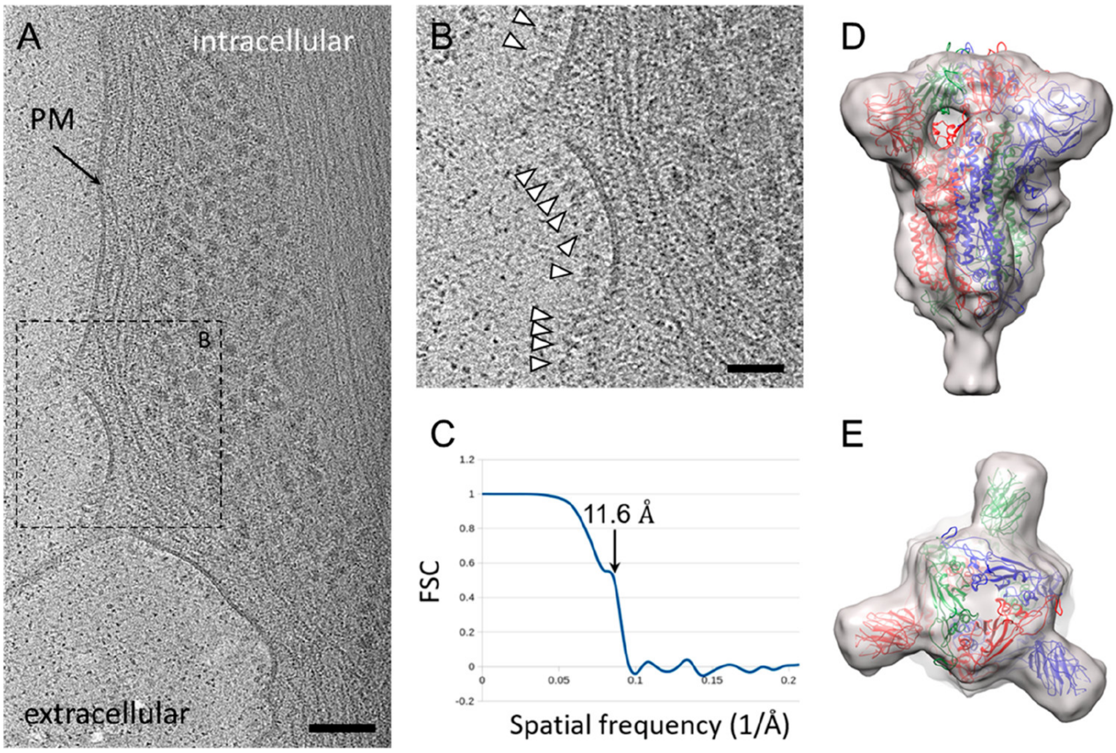


Figure 27 CryoET and subtomogram average of ChAdOx1 nCoV-19 derived spike (A) Tomographic slice of U2OS cell transduced with ChAdOx1 nCoV-19. The slice is 6.4 Å thick; PM = plasma membrane, scale bar = 100 nm. (B) Detailed view of the boxed area marked in (A). White arrowheads indicate spike proteins on the cell surface; scale bar = 50 nm. (C–E) Subtomogram average of ChAdOx1 nCoV-19 spikes at 11.6 Å resolution as indicated by Fourier-Shell correlation at 0.5 cutoff (C), shown from side view (D) and top view (E). SARS-CoV-2 atomic model (PDB 6ZB5) (*213*) is fitted for reference. This figure is taken from (*186*) where it is licensed under a License Creative Commons Attribution license (CC BY 4.0).

The vaccine ChAdOx1 nCoV-19 encodes a spike protein and we sought to understand how it is displayed in transfected cells. Trafficking of S to the PM could be confirmed using flow cytometry by correct binding to the receptor ACE2 and human mAbs. To elucidate its structure, we transfected U2OS and HeLa cells with the vaccine vector and plunge froze them on grids for native cyroET and subtomogram averaging analysis.

I motion-corrected the tilt series and aligned them using eTomo with much manual correction to ensure correct fiducial tracking. The tomograms showed many protruding proteins at the PM which were absent in the control cells (Figure 27A/B). The shape of these proteins is consistent with the shape of SARS-CoV-2 spike proteins in their prefusion conformation. To further confirm this, we used emClarity for subtomogram averaging of these proteins on the cell surface. We achieved a final resolution of 11.6 Å with 3-fold symmetry (Figure 27C). An atomic model of a previously determined SARS-CoV-19 spike protein in its prefusion confirmation fits nicely into the density (Figure 27D/E). Even though side chains could not be resolved, we could rule out a spike protein in its postfusion confirmation as the postfusion atomic model is very elongated and did not not fit into the density. Furthermore, a second round of template matching with pre-/postfusion spike templates could only find very little postfusion spikes.

Most neutralizing antibodies target spike in its prefusion state. We could confirm that ChAdOx1 nCoV-19 displays spike in its prefusion conformation and is thus a good candidate to elicit a strong antibody response after vaccination.

3.4.2 Influence of stabilizing mutations

While the initial vaccine characterized in the section before is based on the early Wuhan-Hu-1 strain, mutations in the S protein requires further development in the vaccine. We thus transfected U2OS cells with two constructs of a second generation vaccine against the Beta/South Africa SARS-CoV-2 variant. The 19E6 construct encoded for additional HexaPro stabilization mutations which was thought to stabilize the prefusion conformation, while the 19E construct had no additional stabilizing factors.

We once again used flow cytometry to check expression of the S protein. Cells were checked at 24, 48 and 72 hours post-infection with either construct. Analysis showed that expression did not increase after the first time point but that the 19E6 infection lead to a higher level of S on the cell surface in its prefusion state.

Then, we used cryoET and STA to check the conformations of both constructs on cell surfaces. A summary of data collection and refinement parameters can be found in Table 8.

To maximize efficiency, each data set was processed by one person in the team. I motion-corrected and aligned the tilt series collected on cells 48 h after infection with the ChAdOx 19E vaccine. I then transferred the stacks to emClarity, ran its CTF estimation and template matching functions before cleaning these initial subtomogram positions in Chimera. To ensure consistency, the same parameters were used across all the data sets. The data sets for the different time points were only pooled together after template cleaning for STA. The final maps reached a resolution of 9.6 Å for the *in situ* 19E S and 9.0 Å for the 19E6 spike protein using

Sample	ChAdOx 19E6	ChAdOx 19E
Data collection		
Magnification	42 000x	42 000x
Voltage [kV]	300	300
Electron dose [e^-/Å]	130	130
Defocus range [µm]	−2.5 to −6.5	−2.5 to −6.5
Camera	K3	K3
Energy-filter (slit)	20eV	20eV
Super-res mode	yes	yes
Acquisition scheme	-60°/60°, 3° group of 3, dose-symmetric	-60°/60°, 3° group of 3, dose-symmetric
Frame number	10	10
Pixel size [Å]	2.2	2.18
No. of tilt-series	68	143
Model refinement		
Final particle images	28896	29279
Symmetry imposed	C3/C1	C3/C1
Map resolution [Å]	9.0/9.7	9.6/11.9
FSC threshold	0.143	0.143
B-factor sharpening	−200	−200
Accession codes	EMD-16403/EMD-16404 EMPIAR-11456	EMD-16405/EMD16406 EMPIAR-11457

Table 8 Data collection parameters

C3 symmetry (Figure 28a/b). The central helix bundle is clearly resolved and some N-linked glycans can also be seen (Figure 28c/d). However, coloring the map by the local resolution shows that the RBD domains are more flexible in 19E than 19E6 (Figure 28a). STA without symmetry applied and subsequent classification showed that most S proteins are present in the one-RBD-up conformation (85% in 19E6 and 58% in 19E), which is consistent with soluble protein analysis (*189*).

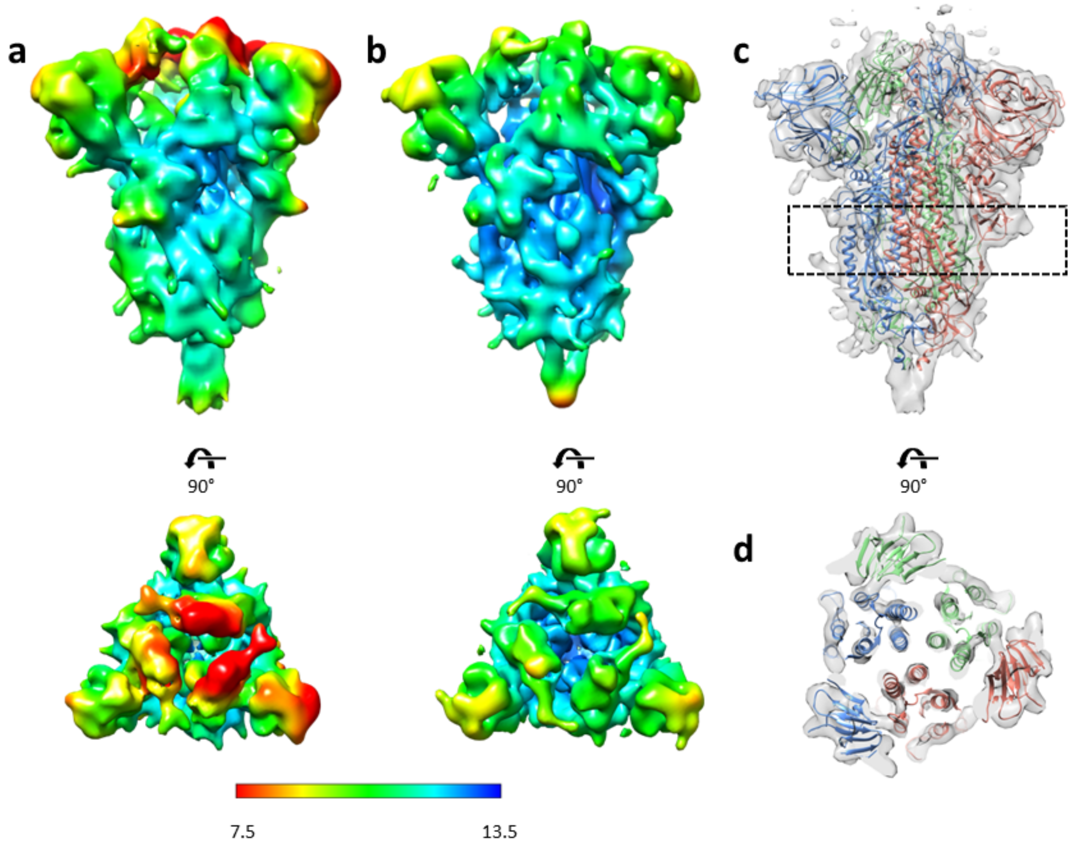


Figure 28 STA of ChAdOx spike 19E and 19E6 (a-b) Orthogonal views of the STA density of ChAdOx spike (a) 19E and (b) 19E6 with C3 symmetry applied. Colored by local resolution. (c) Spike 19E6 density overlayed with a spike model (PDB 6ZGG) (*214*) with each chain colored differently. (d) Orthogonal close-up of the slice marked in (c).

3.4.3 SARS-CoV-2 assembly & egress

To study the late parts of the SARS-CoV-2 life cycle, we grew Vero cells on indexed EM grids and infected them with SARS-CoV-2. The indexed grids allowed us to track the same cells during different analysis techniques and compare the results. A diagram of the complete workflow is depicted in Figure 29. The combination of high resolution cryoET at the cell surface with cyroFIB milling to image the inside of the cell as well as low resolution techniques such as serial FIB/SEM and X-ray cryo-tomography allowed us to correlate detailed structural findings outside and inside the cell with ultrastructural information. During data collection almost 300 tilt series were collected. I helped to sort through them to find infected cells. I also motion corrected the infected cells and aligned the tomograms collected at the cell periphery using eTomo with manual correction during bead tracking.

Serial cryoFIB/SEM and X-ray cryo-tomography showed many cytopathological changes in infected cells. We found many tunnel-like structures built by the plasma membrane and reaching deep into the cytoplasm (Figure 30A/B). The tunnels resembled similar structures found during HIV-1 infection and known to be the places where viruses exit the cell (*215*). The cytoplasm of infected cells contained large viral-containing vesicles (LVCVs) and shorter mitochondria. Furthermore, the nucleus was heavily damaged by invaginated cytoplasm.

CryoET of lamellae, which were milled into the cytoplasm of infected cells, revealed the presence of many DMVs. DMVs are likely needed to evade immune detection during vRNA replication and help to sequester transcripts inside. Many cells also contained larger vesicle packets (VPs) which had been previously suggested to be the result of the fusion of DMVs into larger structures and are necessary for budding of the viruses (*207*, *217*). However,we could neither confirm nor disprove this hypothesis as we could not find any assembly intermediates nor budding viruses in DMVs or VPs.

To better understand the tunnels found in serial cryoFIB/SEM volumes, we took a closer look at the cell periphery using cryoET. And indeed, we could find these tunnels in the tomograms (Figure 30C). The tunnels also contained viruses, which

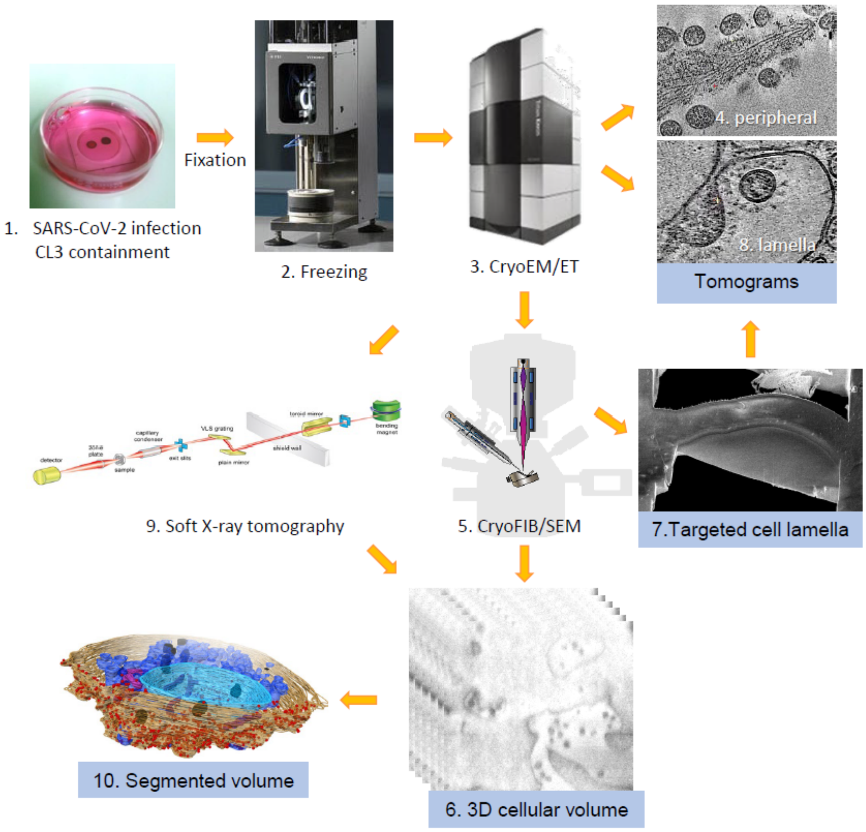


Figure 29 A workflow for correlative multi-modal multi-scale imaging of SARS-CoV-2 infected cells 1) Cells are grown on indexed EM grids, infected with SARS-CoV-2 and fixed with paraformaldehyde. 2) Grids are plunge-frozen in liquid ethane and 3) imaged by cryoEM/ET to locate the infected cells. 4) Tomograms are collected on the cell periphery of infected cells. 5) Infected cells are subjected to processing and imaging in a cryoFIB/SEM dualbeaminstrument for 6) serial cryoFIB/SEM volume imaging and 7) targeted cell lamella. 8) Tomograms are collected from cell lamellae. 9) Alternatively, infected cells are imaged by soft-X-ray cryo-tomography. 10) Cellular volume data are manually segmented. This figure is taken from (*187*) where it is licensed under a Creative Commons Attribution-ShareAlike License Creative Commons Attribution license (CC BY 4.0).

is consistent with these tunnels resulting from the fusion of large virus-containing vesicles with the PM and expelling the viruses in an exocytosis-like mechanism.

Many tomograms at the cell periphery showed the plasma membranes being dis-

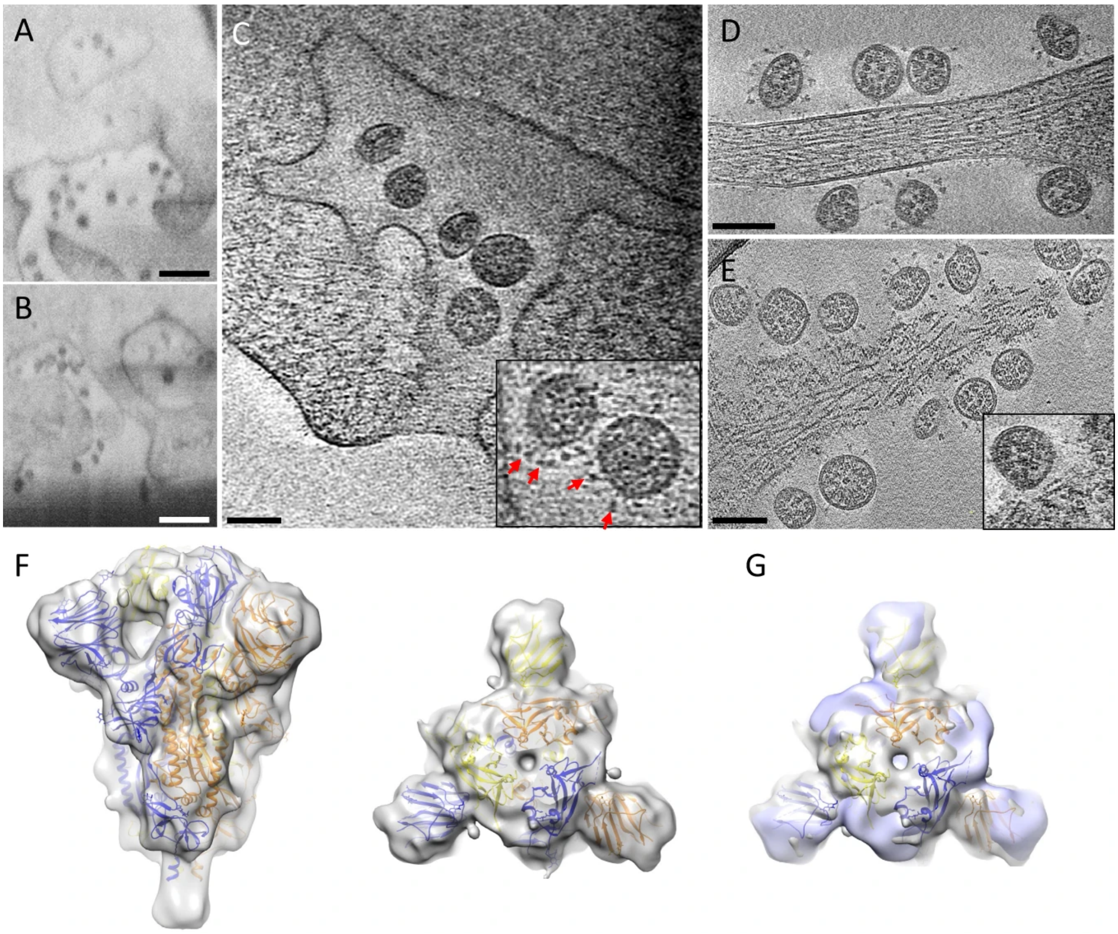


Figure 30 SARS-CoV-2 egress (A,B) CryoFIB/SEM images of cell periphery, depicting virus particles exiting through extended tunnels connected to external ofthe cell. (C) CryoET of the SARS-CoV-2-exiting tunnel. Inset shows a close-up view of viruses with prefusion spikes (red arrows). (D,E) Viruses released from the cell, close to an intact cell membrane (D) or next to membrane lesion sites (E). Inset, close-up view of a membrane lesion site. (F) Subtomogram average of spikes on released viruses at $10.6\,\text{Å}$ resolution fitted with an atomic model of spike trimer (PDB 6VXX) (*216*), viewed from side and top. (G) Comparison of spike structures from intracellular assembled viruses (transparent blue) and extracellular released viruses (transparent gray), viewed from top. Scale bar is $300\,\text{nm}$ in (A, B) and $100\,\text{nm}$ in (C–E). This figure is taken from (*187*) where it is licensed under a Creative Commons Attribution license (CC BY 4.0).

rupted or containing membrane lesions (Figure 30D/E). When counted, 44.6% of tomograms collected on infected cells showed membrane lesions, compared to only 18.7% in uninfected cells. The cell lesions hint at the plasma membrane becoming more fragile upon infection and could also be found previously using atomic force microscopy (*218*).

Subtomogram averaging done on extracellular and intracellular spike proteins yielded density maps at 10.6 Å and 16 Å respectively (Figure 30F). The S protein at the PM as well as inside the cell are in the prefusion state and very similar to structures derived from purified virions. There is only a slight difference between intracellular and extracellular densities in how the RBDs are oriented but no major rearrangements took place during viral egress (Figure 30G).

3.5 Discussion

Our first study of the spike protein produced upon transfection by the ChAdOx1 nCoV-19 vaccine shows native-like spike structures. We could show that the spike proteins produced are still in their prefusion state and it is thus a good target to elicit an immune response. Our data can be compared directly to spike proteins produced by different vaccines. However, the effect of the prefusion state on vaccine efficacy is not yet determined and comparisons to other vaccines needed to allow for further development of these vaccines.

As the vaccine got adapted to further strains of the virus, we repeated our analysis. The two second-generation ChAdOx constructs both also expressed spike in its prefusion state. We found that the HexaPro 19E6 version increased the one-RBD-up conformation and is more stable. Our results, together with an antibody reaction (*219*), confirm the suitability of the HexaPro stabilized ChAdOx vector for vaccine trials. As new SARS-CoV-2 emerge, vaccines will need to be adapted accordingly and regular boosters will be needed. Structural characterizations like ours reflect the state of the spike protein *in situ* and provide a base to develop the next-generation vaccine.

Our approach of using orthogonal techniques at different spatial scales was successful in gaining comprehensive information on SARS-CoV-2 infection. We were able to characterize cytopathic effects in frozen cellular samples and correlate these findings with *in situ* structural information. We thus propose a new model of SARS-CoV-2 infection with its replication being spatially separated inside the cell. RNA replication takes place in DMVs which makes it hard to detect by the immune system. After replication, the vRNAs is transported through transmembrane portals in the DMVs to viral assembly sites and mRNA exits to the cytoplasm for the translation of viral proteins. The S proteins are then transported using small vesicles to the assembly sites as well. The vesicles then fuse with the SMV membranes and the presence of S, vRNA and N at the assembly sites leads to membrane curvature and subsequently budding off into the SMV. The viral particles can exit the cell by exocytosis, which leads to long tunnels.

A good understanding of the complete viral life cycle is needed to better understand newly emerging viruses. Our unique approach can also be applied to other viruses and can help to better grasp infection strategies. We could provide a meaningful step to understand SARS-CoV-2 but there are still parts of its life cycle we do not yet understand. The roles of M and E, as well as interactions with host proteins and transport machinery, is still poorly understood and needs further investigation.

Chapter 4

Future Work

4.1 HERV-K

HERV-K is still a very unknown virus and much research is still needed to understand its complete viral life cycle and role in disease. Taking my research as a base, there are a few things still to be done:

- I managed to elucidate the structure of the immature CA lattice and was thus able to compare it to its mature form and other retroviruses. However, no efforts have been done so far to do the same with the MA lattice. It would be important to know if this layer also shows a structural conservation to modern retroviruses to confirm my hypothesis of a similar assembly and maturation process across evolutionary time.

- Even though the density atop the 6HB looks like an IP6 molecule and molecular dynamic simulations as well as comparisons to HIV-1 and RSV support this so far, we have not yet disproven that this molecule is not another charged small molecule. Thus, assembly studies of CA with and without IP6 still need to be done.

- My work was done solely based on the HERV-K_{con} Gag protein. I have not yet tested how my results compare to the complete construct, which also contains *Env* and *Pol*. The plasmid containing a non-functioning protease could be used for a direct comparison of the immature lattice to my VLPs.

- A lot of work has been done to understand the role of HERV-K in neurological diseases and cancers. However, designing drugs to combat these has been difficult. So far, the only solution was using trial-and-error methods or trying to adapt constructs used against HIV-1. The immature CA structure can now be combined with the mature CA structure for computational docking experiments to inhibit HERV-K maturation. This would allow for a more targeted approach and can hopefully help develop drugs faster.

4.2 SARS-CoV-2

In contrast to the almost unknown class of endogenous retroviruses, a lot of research has been done SARS-CoV-2 since the first case of Covid-19. A big collaborative effort helped to release several different vaccines in record time and to understand how the virus could spread so quickly across the globe. We have proven that it is possible to study ChAdOx spike proteins *in situ* and it can be used to aid the development of new vaccines. However, as the virus will continue to mutate, there is a need to also continue adapting our vaccines. Our technique must thus be repeated when a next-generation vaccine needs to be tested.

We were successful in using a multimodal, multiscalar approach to study SARS-CoV-2 infection. It can now be used to study different viruses and help understand processes across large distances in detail. However, there are still a few questions left unanswered during SARS-CoV-2 infection. We could see neither M nor E in our data. Further tagging of these proteins is needed to understand their role and incorporate them into our data. Also, labeling of the cellular compartments associated with the viral life cycle and correlative microscopy would be needed to dissect different viral pathways.

My Publications

Krebs, A.-S., Mendonça, L. M. & Zhang, P. (2022). Structural Analysis of Retrovirus Assembly and Maturation. *Viruses* 14, 1–12

Watanabe, Y., Mendonça, L., Allen, E. R., Howe, A., Lee, M., Allen, J. D., Chawla, H., Pulido, D., Donnellan, F., Davies, H., Ulaszewska, M., Belij-Rammerstorfer, S., Morris, S., Krebs, A.-S., Dejnirattisai, W., Mongkolsapaya, J., Supasa, P., Screaton, G. R., Green, C. M., Lambe, T., Zhang, P., Gilbert, S. C. & Crispin, M. (2021) Native-like SARS-CoV-2 Spike Glycoprotein Expressed by ChAdOx1 nCoV-19/AZD1222 Vaccine. *ACS Central Science* 7, 594–602

Mendonça, L., Howe, A., Gilchrist, J. B., Sheng, Y., Sun, D., Knight, M. L., Zanetti-Domingues, L. C., Bateman, B., Krebs, A.-S., Chen, L., Radecke, J., Li, V. D., Ni, T., Kounatidis, I., Koronfel, M. A., Szynkiewicz, M., Harkiolaki, M., Martin-Fernandez, M. L., James, W. & Zhang, P. (2021) Correlative multi-scale cryo-imaging unveils SARS-CoV-2 assembly and egress. *Nature communications* 12

Krebs, A.-S., Liu, H.-F., Zhou, Y., Rey, J. S., Levintov, L., Shen, J., Howe, A., Perilla., J. R., Bartesaghi, A. & Zhang, P. (2023) Molecular architecture and conservation of an immature human endogenous retrovirus. *Nature communications* Accepted

Ni, T., Mendonça, L., Zhu, Y., Howe, A., Radecke, J., Shah, P., Sheng, Y., Krebs, A.-S., Duyvesteyn, H., Allen, E., Lambe, T., Bisset, C., Spencer, A., Mor-

ris, S., Stuart, D. I., Gilbert, S. & Zhang, P. (Submitted) ChAdOx1 COVID vaccines express RBD open prefusion SARS-CoV-2 spikes on the cell surface. *Nature Structural & Molecular Biology*

Bibliography

(*1*) Krebs, A. S., Mendonça, L. M., and Zhang, P. (2022). Structural Analysis of Retrovirus Assembly and Maturation. *Viruses 14*, 1–12.

(*2*) Coffin, J., Blomberg, J., Fan, H., Gifford, R., Hatziioannou, T., Lindemann, D., Mayer, J., Stoye, J., Tristem, M., and Johnson, W. (2021). ICTV Virus Taxonomy Profile: Retroviridae 2021. *Journal of General Virology 102*, DOI: `10.1099/JGV.0.001712/CITE/REFWORKS`.

(*3*) Jern, P., Sperber, G. O., and Blomberg, J. (2005). Use of endogenous retroviral sequences (ERVs) and structural markers for retroviral phylogenetic inference and taxonomy. *Retrovirology 2*, DOI: `10.1186/1742-4690-2-50`.

(*4*) Regenmortel, V., Carstens, C. M. L., Estes, E. B., Lemon, M. K., Mayo, J., and Wickner, C. R. (2000). ICTV 7th Report. *AcademicPress 1024*.

(*5*) Fauquet, C. M., and Mayo, M. A. (2001). The 7th ICTV Report. *Archives of Virology 146*, 189–194.

(*6*) Xiong, Y., and Eickbush, T. H. (1990). Origin and evolution of retroelements based upon their reverse transcriptase sequences. *The EMBO journal 9*, 3353–3362.

(*7*) Hayward, A., Cornwallis, C. K., and Jern, P. (2015). Pan-vertebrate comparative genomics unmasks retrovirus macroevolution. *Proceedings of the National Academy of Sciences of the United States of America 112*, 464–469.

(*8*) Kraus, B., Boller, K., Reuter, A., and Schnierle, B. S. (2011). Characterization of the human endogenous retrovirus K Gag protein: Identification of protease cleavage sites. *Retrovirology 8*, DOI: 10.1186/1742-4690-8-21.

(*9*) George, M., Schwecke, T., Beimforde, N., Hohn, O., Chudak, C., Zimmermann, A., Kurth, R., Naumann, D., and Bannert, N. (2011). Identification of the protease cleavage sites in a reconstituted Gag polyprotein of an HERV-K(HML-2) element. *Retrovirology 8*, 30.

(*10*) Martin, J. L., Cao, S., Maldonado, J. O., Zhang, W., and Mansky, L. M. (2016). Distinct Particle Morphologies Revealed through Comparative Parallel Analyses of Retrovirus-Like Particles. *Journal of Virology 90*, 8074–8084.

(*11*) Tang, C., Loeliger, E., Luncsford, P., Kinde, I., Beckett, D., and Summers, M. F. (2004). Entropic switch regulates myristate exposure in the HIV-1 matrix protein. *Proc Natl Acad Sci U S A 101*, 517–522.

(*12*) Mendonça, L., Sun, D., Ning, J., Liu, J., Kotecha, A., Olek, M., Frosio, T., Fu, X., Himes, B. A., Kleinpeter, A. B., Freed, E. O., Zhou, J., Aiken, C., and Zhang, P. (2021). CryoET structures of immature HIV Gag reveal six-helix bundle. *Communications biology 4*, DOI: 10.1038/s42003-021-01999-1.

(*13*) Amarasinghe, G. K., De Guzman, R. N., Turner, R. B., Chancellor, K. J., Wu, Z. R., and Summers, M. F. (2000). NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition. *Journal of Molecular Biology 301*, 491–511.

(*14*) Fossen, T., Wray, V., Bruns, K., Rachmat, J., Henklein, P., Tessmer, U., Maczurek, A., Klinger, P., and Schubert, U. (2005). Solution structure of the human immunodeficiency virus type 1 p6 protein. *Journal of Biological Chemistry 280*, 42515–42527.

(*15*) Bharat, T. A. M., Castillo Menendez, L. R., Hagen, W. J. H., Lux, V., Igonet, S., Schorb, M., Schur, F. K. M., Kräusslich, H.-G., and Briggs, J. A. G. (2014). Cryo-electron microscopy of tubular arrays of HIV-1 Gag resolves structures essential for immature virus assembly. *Proc Natl Acad Sci U S A 111*, 8233–8238.

(*16*) Qu, K., Glass, B., Dolezal, M., Schur, F. K., Murciano, B., Rein, A., Rumlová, M., Ruml, T., Kräusslich, H. G., and Briggs, J. A. (2018). Structure and architecture of immature and mature murine leukemia virus capsids. *Proceedings of the National Academy of Sciences of the United States of America 115*, E11751–E11760.

(*17*) Schur, F. K., Hagen, W. J., Rumlová, M., Ruml, T., Müller, B., Kraüsslich, H. G., and Briggs, J. A. (2015). Structure of the immature HIV-1 capsid in intact virus particles at 8.8 Å resolution. *Nature 517*, 505–508.

(*18*) Schur, F. K. M., Obr, M., Hagen, W. J. H., Wan, W., Jakobi, A. J., Kirkpatrick, J. M., Sachse, C., Kräusslich, H.-G., and Briggs, J. A. G. (2016). Assembly and Maturation. *Science 353*, 506–508.

(*19*) Briggs, J. A., and Kräusslich, H. G. (2011). The molecular architecture of HIV. *Journal of Molecular Biology 410*, 491–500.

(*20*) Kelly, B. N., Howard, B. R., Wang, H., Robinson, H., Sundquist, W. I., and Hill, C. P. (2006). Implications for viral capsid assembly from crystal structures of HIV-1 Gag(1-278) and CA(N)(133-278). *Biochemistry 45*, 11257–66.

(*21*) Morellet, N., Druillennec, S., Lenoir, C., Bouaziz, S., and Roques, B. P. (2005). Helical structure determined by NMR of the HIV-1 (345-392)Gag sequence, surrounding p2: implications for particle assembly and RNA packaging. *Protein Science 14*, 375–386.

(*22*) Tang, C., Ndassa, Y., and Summers, M. F. (2002). Structure of the N-terminal 283-residue fragment of the immature HIV-1 Gag polyprotein. *Nature structural biology 9*, 537–543.

(*23*) Campos-Olivas, R., Newman, J. L., and Summers, M. F. (2000). Solution structure and dynamics of the Rous sarcoma virus capsid protein and comparison with capsid proteins of other retroviruses. *Journal of molecular biology 296*, 633–649.

(*24*) Kingston, R. L., Fitzon-Ostendorp, T., Eisenmesser, E. Z., Schatz, G. W., Vogt, V. M., Post, C. B., and Rossmann, M. G. (2000). Structure and self-association of the Rous sarcoma virus capsid protein. *Structure 8*, 617–628.

(*25*) Wagner, J. M., Zadrozny, K. K., Chrustowicz, J., Purdy, M. D., Yeager, M., Ganser-Pornillos, B. K., and Pornillos, O. (2016). Crystal structure of an HIV assembly and maturation switch. *Elife 5*, e17063.

(*26*) Himes, B. A., and Zhang, P. (2018). emClarity: software for high-resolution cryo-electron tomography and subtomogram averaging. *Nature Methods 15*, 955–961.

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

(*28*) Wright, E. R., Schooler, J. B., Ding, H. J., Kieffer, C., Fillmore, C., Sundquist, W. I., and Jensen, G. J. (2007). Electron cryotomography of immature HIV-1 virions reveals the structure of the CA and SP1 Gag shells. *The EMBO journal 26*, 2218–2226.

(*29*) Bharat, T. A., Davey, N. E., Ulbrich, P., Riches, J. D., De Marco, A., Rumlova, M., Sachse, C., Ruml, T., and Briggs, J. A. (2012). Structure of the immature retroviral capsid at 8Å resolution by cryo-electron microscopy. *Nature 487*, 385–389.

(*30*) Krupovic, M., and Koonin, E. V. (2017). Homologous Capsid Proteins Testify to the Common Ancestry of Retroviruses, Caulimoviruses, Pseudoviruses, and Metaviruses. *Journal of virology 91*, DOI: `10.1128/JVI.00210-17`.

(*31*) Llorens, C., Fares, M. A., and Moya, A. (2008). Relationships of gag-pol diversity between Ty3/Gypsy and Retroviridae LTR retroelements and the three kings hypothesis. *BMC Evolutionary Biology 8*, DOI: `10.1186/1471-2148-8-276`.

(*32*) Campbell, S., Fisher, R. J., Towler, E. M., Fox, S., Issaq, H. J., Wolfe, T., Phillips, L. R., and Rein, A. (2001). Modulation of HIV-like particle assembly in vitro by inositol phosphates. *Proceedings of the National Academy of Sciences of the United States of America 98*, 10875–10879.

(*33*) Novikova, M., Zhang, Y., Freed, E. O., and Peng, K. (2019). Multiple Roles of HIV-1 Capsid during the Virus Replication Cycle. *Virologica Sinica 34*, 119–134.

(*34*) Dick, R., Xu, C., Morado, D., Kravchuk, V., Ricana, C., Lyddon, T., Broad, A., Feathers, J., Johnson, M., Vogt, V., Perilla, J., Briggs, J., and Schur, F. (2020). Structures of immature EIAV Gag lattices reveal a conserved role for IP6 in lentivirus assembly. *PLoS pathogens 16*, e1008277.

(*35*) Obr, M., Ricana, C. L., Nikulin, N., Feathers, J.-P. R., Klanschnig, M., Thader, A., Johnson, M. C., Vogt, V. M., Schur, F. K. M., and Dick, R. A. (2021). Structure of the mature Rous sarcoma virus lattice reveals a role for IP6 in the formation of the capsid hexamer. *Nature Communications 12*, 1–12.

(*36*) Wang, M. et al. (2017). Quenching protein dynamics interferes with HIV capsid maturation. *Nature Communications 8*, 1779.

(*37*) Dick, R. A., Zadrozny, K. K., Xu, C., Schur, F. K. M., Lyddon, T. D., Ricana, C. L., Wagner, J. M., Perilla, J. R., Ganser-Pornillos, B. K., Johnson, M. C., Pornillos, O., and Vogt, V. M. (2018). Inositol phosphates are assembly co-factors for HIV-1. *Nature 560*, 509–512.

(*38*) Keller, P. W., Adamson, C. S., Heymann, J. B., Freed, E. O., and Steven, A. C. (2011). HIV-1 maturation inhibitor bevirimat stabilizes the immature Gag lattice. *J Virol 85*, 1420–1428.

(*39*) Spearman, P. (2016). HIV-1 Gag as an Antiviral Target: Development of Assembly and Maturation Inhibitors. *Curr Top Med Chem 16*, 1154–1166.

(*40*) Qu, K., Ke, Z., Zila, V., Anders-Össwein, M., Glass, B., Mücksch, F., Müller, R., Schultz, C., Müller, B., Kräusslich, H. G., and Briggs, J. A. (2021). Maturation of the matrix and viral membrane of HIV-1. *Science 373*, 700–704.

(*41*) Hill, C. P., Worthylake, D., Bancroft, D. P., Christensen, A. M., and Sundquist, W. I. (1996). Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly. *Proceedings of the National Academy of Sciences of the United States of America 93*, 3099–3104.

(*42*) Liu, C. et al. (2016). Cyclophilin A stabilizes the HIV-1 capsid through a novel non-canonical binding site. *Nature Communications 7*, DOI: `10.1038/NCOMMS10714`.

(*43*) Ning, J., Zhong, Z., Fischer, D. K., Harris, G., Watkins, S. C., Ambrose, Z., and Zhang, P. (2018). Truncated CPSF6 Forms Higher-Order Complexes That Bind and Disrupt HIV-1 Capsid. *Journal of Virology 92*, DOI: `10.1128/JVI.00368-18`.

(*44*) Zhong, Z., Ning, J., Boggs, E. A., Jang, S., Wallace, C., Telmer, C., Bruchez, M. P., Ahn, J., Engelman, A. N., Zhang, P., Watkins, S. C., and Ambrose, Z. (2021). Cytoplasmic CPSF6 Regulates HIV-1 Capsid Trafficking and Infection in a Cyclophilin A-Dependent Manner. *mBio 12*, DOI: `10.1128/mBio.03142-20`.

(*45*) Mattei, S., Glass, B., Hagen, W. J., Kräusslich, H.-G., and Briggs, J. A. (2016). The structure and flexibility of conical HIV-1 capsids determined within intact virions. *Science 354*, 1434–1437.

(*46*) Ni, T., Gerard, S., Zhao, G., Dent, K., Ning, J., Zhou, J., Shi, J., Anderson-Daniels, J., Li, W., Jang, S., Engelman, A. N., Aiken, C., and Zhang, P. (2020). Intrinsic curvature of the HIV-1 CA hexamer underlies capsid topology and interaction with cyclophilin A. *Nature Structural & Molecular Biology 27*, 855–862.

(*47*) Meissner, M. E., Mendonça, L. M., Zhang, W., and Mansky, L. M. (2017). Polymorphic Nature of Human T-Cell Leukemia Virus Type 1 Particle Cores as Revealed through Characterization of a Chronically Infected Cell Line. *Journal of Virology 91*, DOI: `10.1128/jvi.00369-17`.

(*48*) Ni, T., Zhu, Y., Yang, Z., Xu, C., Chaban, Y., Nesterova, T., Ning, J., Böcking, T., Parker, M. W., Monnie, C., Ahn, J., Perilla, J. R., and Zhang, P. (2021). Structure of native HIV-1 cores and their interactions with IP6 and CypA. *Science advances 7*, DOI: `10.1126/SCIADV.ABJ5715`.

(*49*) Zhao, G., Perilla, J. R., Yufenyuy, E. L., Meng, X., Chen, B., Ning, J., Ahn, J., Gronenborn, A. M., Schulten, K., Aiken, C., and Zhang, P. (2013). Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature 497*, 643–646.

(*50*) Mallery, D. L., Márquez, C. L., McEwan, W. A., Dickson, C. F., Jacques, D. A., Anandapadamanaban, M., Bichel, K., Towers, G. J., Saiardi, A., Böcking, T., and James, L. C. (2018). IP6 is an HIV pocket factor that prevents capsid collapse and promotes DNA synthesis. *eLife 7*, e35335.

(*51*) Jacques, D. A., McEwan, W. A., Hilditch, L., Price, A. J., Towers, G. J., and James, L. C. (2016). HIV-1 uses dynamic capsid pores to import nucleotides and fuel encapsidated DNA synthesis. *Nature 536*, 349–353.

(*52*) Pornillos, O., Ganser-Pornillos, B. K., and Yeager, M. (2011). Atomic-level modelling of the HIV capsid. *Nature 469*, 424–427.

(*53*) Gres, A. T., Kirby, K. A., Kewalramani, V. N., Tanner, J. J., Pornillos, O., and Sarafianos, S. G. (2015). X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability. *Science 349*, 99–103.

(*54*) Acton, O., Grant, T., Nicastro, G., Ball, N. J., Goldstone, D. C., Robertson, L. E., Sader, K., Nans, A., Ramos, A., Stoye, J. P., Taylor, I. A., and Rosenthal, P. B. (2019). Structural basis for Fullerene geometry in a human endogenous retrovirus capsid. *Nature Communications 10*, 1–13.

(*55*) Medstrand, P., and Blomberg, J. (1993). Characterization of Novel Reverse Transcriptase Encoding Human Endogenous Retroviral Sequences Similar to Type A and Type B Retroviruses: Differential Transcription in Normal Human Tissues. *Journal of Virology 67*, 6778–6787.

(*56*) Hu, W. S., and Hughes, S. H. (2012). HIV-1 Reverse Transcription. *Cold Spring Harbor Perspectives in Medicine 2*, a006882.

(*57*) Baldwin, E. T., Gotte, M., Tchesnokov, E. P., Arnold, E., Hagel, M., Nichols, C., Dossang, P., Lamers, M., Wan, P., Steinbacher, S., and Romero, D. L. (2022). Human endogenous retrovirus-K (HERV-K) reverse transcriptase (RT) structure and biochemistry reveals remarkable similarities to HIV-1 RT and opportunities for HERV-K–specific inhibition. *Proceedings of the National Academy of Sciences of the United States of America 119*, e2200260119.

(*58*) Das, K., Martinez, S. E., and Arnold, E. (2017). Structural Insights into HIV Reverse Transcriptase Mutations Q151M and Q151M Complex That

Confer Multinucleoside Drug Resistance. *Antimicrobial agents and chemotherapy 61*, DOI: 10.1128/AAC.00224-17.

(*59*) Taube, R., Loya, S., Avidan, O., Perach, M., and Hizi, A. (1998). Reverse transcriptase of mouse mammary tumour virus: expression in bacteria, purification and biochemical characterization. *The Biochemical journal 332*, 807–808.

(*60*) Bray, S., Turnbull, M., Hebert, S., and Douville, R. N. (2016). Insight into the ERVK Integrase - Propensity for DNA Damage. *Frontiers in microbiology 7*, DOI: 10.3389/FMICB.2016.01941.

(*61*) Maertens, G. N., Engelman, A. N., and Cherepanov, P. (2022). Structure and function of retroviral integrase. *Nature reviews. Microbiology 20*, 20–34.

(*62*) Ballandras-Colas, A., Brown, M., Cook, N. J., Dewdney, T. G., Demeler, B., Cherepanov, P., Lyumkis, D., and Engelman, A. N. (2016). Cryo-EM reveals a novel octameric integrase structure for betaretroviral intasome function. *Nature 530*, 358–361.

(*63*) Fontana, J., Jurado, K. A., Cheng, N., Ly, N. L., Fuchs, J. R., Gorelick, R. J., Engelman, A. N., and Steven, A. C. (2015). Distribution and Redistribution of HIV-1 Nucleocapsid Protein in Immature, Mature, and Integrase-Inhibited Virions: a Role for Integrase in Maturation. *Journal of virology 89*, 9765–9780.

(*64*) Engelman, A. N., and Kvaratskhelia, M. (2022). Multimodal Functionalities of HIV-1 Integrase. *Viruses 14*, DOI: 10.3390/V14050926.

(*65*) Schommer, S., Sauter, M., Kräusslich, H. G., Best, B., and Mueller-Lantzsch, N. (1996). Characterization of the human endogenous retrovirus K proteinase. *Journal of General Virology 77*, 375–379.

(*66*) Weber, I. T., Wang, Y. F., and Harrison, R. W. (2021). HIV Protease: Historical Perspective and Current Research. *Viruses 13*, DOI: `10.3390/V13050839`.

(*67*) Dunn, B. M., Goodenow, M. M., Gustchina, A., and Wlodawer, A. (2002). Retroviral proteases. *Genome Biology 3*.

(*68*) Ludwig, C., Leiherer, A., and Wagner, R. (2008). Importance of Protease Cleavage Sites within and Flanking Human Immunodeficiency Virus Type 1 Transframe Protein p6* for Spatiotemporal Regulation of Protease Activation. *Journal of Virology 82*, 4573–4584.

(*69*) Tabler, C. O., Wegman, S. J., Chen, J., Shroff, H., Alhusaini, N., and Tilton, J. C. (2022). The HIV-1 Viral Protease Is Activated during Assembly and Budding Prior to Particle Release. *Journal of Virology 96*, DOI: `10.1128/JVI.02198-21/SUPPL{\_}FILE/JVI.02198-21-S0002.XLSX`.

(*70*) Hizi, A., and Herzig, E. (2015). dUTPase: the frequently overlooked enzyme encoded by many retroviruses. *Retrovirology 12*, DOI: `10.1186/S12977-015-0198-9`.

(*71*) Mayer, J., and Meese, E. U. (2003). Presence of dUTPase in the various human endogenous retrovirus K (HERV-K) families. *Journal of molecular evolution 57*, 642–649.

(*72*) Harris, J. M., McIntosh, E. M., and Muscat, G. E. (1999). Structure/function analysis of a dUTPase: catalytic mechanism of a potential chemotherapeutic target. *Journal of molecular biology 288*, 275–287.

(*73*) Garcia-Montojo, M., Doucet-O'Hare, T., Henderson, L., and Nath, A. (2018). Human endogenous retrovirus-K (HML-2): a comprehensive review. *Critical Reviews in Microbiology 44*, 715–738.

(*74*) Vértessy, B. G., and Tóth, J. (2009). Keeping uracil out of DNA: physiological role, structure and catalytic mechanism of dUTPases. *Accounts of Chemical Research 42*, 97–106.

(*75*) Chen, R., Wang, H., and Mansky, L. M. (2002). Roles of uracil-DNA glycosylase and dUTPase in virus replication. *The Journal of general virology 83*, 2339–2345.

(*76*) Barabás, O., Németh, V., and Vértessy, B. G. (2006). Crystallization and preliminary X-ray studies of dUTPase from Mason–Pfizer monkey retrovirus. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications 62*, 399–401.

(*77*) Prasad, G. S., Stura, E. A., Elder, J. H., and Stout, C. D. (2000). Structures of feline immunodeficiency virus dUTP pyrophosphatase and its nucleotide complexes in three crystal forms. *Acta crystallographica. Section D, Biological crystallography 56*, 1100–1109.

(*78*) Dauter, Z., Persson, R., Rosengren, A. M., Nyman, P. O., Wilson, K. S., and Cedergren-Zeppezauer, E. S. (1999). Crystal structure of dUTPase from equine infectious anaemia virus; active site metal binding in a substrate analogue complex. *Journal of molecular biology 285*, 655–673.

(*79*) Foerster, J. et al. (2005). Haplotype sharing analysis identifies a retroviral dUTPase as candidate susceptibility gene for psoriasis. *The Journal of investigative dermatology 124*, 99–102.

(*80*) Ariza, M. E., and Williams, M. V. (2011). A human endogenous retrovirus K dUTPase triggers a TH1, TH17 cytokine response: does it have a role in psoriasis? *The Journal of investigative dermatology 131*, 2419–2427.

(*81*) Lai, O. Y., Chen, H., Michaud, H. A., Hayashi, G., Kuebler, P. J., Hultman, G. K., Ariza, M. E., Williams, M. V., Batista, M. D., Nixon, D. F., Foerster, J., Bowcock, A. M., and Liao, W. (2012). Protective effect of human endogenous retrovirus K dUTPase variants on psoriasis susceptibility. *The Journal of investigative dermatology 132*, 1833–1840.

(*82*) Checkley, M. A., Luttge, B. G., and Freed, E. O. (2011). HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. *Journal of Molecular Biology 410*, 582–608.

(*83*) Zhao, C., Li, H., Swartz, T. H., and Chen, B. K. (2022). The HIV Env Glycoprotein Conformational States on Cells and Viruses. *mBio 13*, DOI: `10.1128/MBIO.01825-21`.

(*84*) Yuan, M., Cottrell, C. A., Ozorowski, G., van Gils, M. J., Kumar, S., Wu, N. C., Sarkar, A., Torres, J. L., de Val, N., Copps, J., Moore, J. P., Sanders, R. W., Ward, A. B., and Wilson, I. A. (2019). Conformational Plasticity in the HIV-1 Fusion Peptide Facilitates Recognition by Broadly Neutralizing Antibodies. *Cell Host & Microbe 25*, 873–883.

(*85*) Ofek, G., Guenaga, F. J., Schief, W. R., Skinner, J., Baker, D., Wyatt, R., and Kwong, P. D. (2010). Elicitation of structure-specific antibodies by epitope scaffolds. *Proceedings of the National Academy of Sciences of the United States of America 107*, 17880–17887.

(*86*) Munro, J. B., Gorman, J., Ma, X., Zhou, Z., Arthos, J., Burton, D. R., Koff, W. C., Courter, J. R., Smith, A. B., Kwong, P. D., Blanchard, S. C., and Mothes, W. (2014). Conformational dynamics of single HIV-1 envelope trimers on the surface of native virions. *Science 346*, 759–763.

(*87*) Herschhorn, A., Ma, X., Gu, C., Ventura, J. D., Castillo-Menendez, L., Melillo, B., Terry, D. S., Smith, A. B., Blanchard, S. C., Munro, J. B., Mothes, W., Finzi, A., and Sodroski, J. (2016). Release of GP120 restraints leads to an entry-competent intermediate state of the HIV-1 envelope glycoproteins. *mBio 7*, 01598–16.

(*88*) Chan, S. M., Sapir, T., Park, S.-S., Rual, J.-F., Contreras-Galindo, R., Reiner, O., and Markovitz, D. M. (2019). The HERV-K accessory protein Np9 controls viability and migration of teratocarcinoma cells. *Plos One 14*, e0212970.

(*89*) Schmitt, K., Heyne, K., Roemer, K., Meese, E., and Mayer, J. (2015). HERV-K(HML-2) rec and np9 transcripts not restricted to disease but present in many normal human tissues. *Mobile DNA 6*, DOI: `10.1186/S13100-015-0035-7`.

(*90*) Armbruester, V., Sauter, M., Roemer, K., Best, B., Hahn, S., Nty, A., Schmid, A., Philipp, S., Mueller, A., and Mueller-Lantzsch, N. (2004). Np9 protein of human endogenous retrovirus K interacts with ligand of numb protein X. *Journal of virology 78*, 10310–10319.

(*91*) Denne, M., Sauter, M., Armbruester, V., Licht, J. D., Roemer, K., and Mueller-Lantzsch, N. (2007). Physical and functional interactions of human endogenous retrovirus proteins Np9 and rec with the promyelocytic leukemia zinc finger protein. *Journal of virology 81*, 5607–5616.

(*92*) Chen, T. et al. (2013). The viral oncogene Np9 acts as a critical molecular switch for co-activating β-catenin, ERK, Akt and Notch1 and promoting the growth of human leukemia stem/progenitor cells. *Leukemia 27*, 1469–1478.

(*93*) Wodrich, H., and Kräusslich, H. G. (2001). Nucleocytoplasmic RNA transport in retroviral replication. *Results and problems in cell differentiation 34*, 197–217.

(*94*) Hanke, K., Hohn, O., Liedgens, L., Fiddeke, K., Wamara, J., Kurth, R., and Bannert, N. (2013). Staufen-1 interacts with the human endogenous retrovirus family HERV-K(HML-2) rec and gag proteins and increases virion production. *Journal of virology 87*, 11019–11030.

(*95*) Langner, J. S., Fuchs, N. V., Hoffmann, J., Wittmann, A., Brutschy, B., Löwer, R., and Suess, B. (2012). Biochemical analysis of the complex between the tetrameric export adapter protein Rec of HERV-K/HML-2 and the responsive RNA element RcRE pck30. *Journal of virology 86*, 9079–9087.

(*96*) Wang, Q., Finzi, A., and Sodroski, J. (2020). The Conformational States of the HIV-1 Envelope Glycoproteins. *Trends in microbiology 28*, 655–667.

(*97*) McDonald, D., Vodicka, M. A., Lucero, G., Svitkina, T. M., Borisy, G. G., Emerman, M., and Hope, T. J. (2002). Visualization of the intracellular behavior of HIV in living cells. *The Journal of cell biology 159*, 441–452.

(*98*) Zila, V., Margiotta, E., Turoňová, B., Müller, T. G., Zimmerli, C. E., Mattei, S., Allegretti, M., Börner, K., Rada, J., Müller, B., Lusic, M., Kräusslich, H. G., and Beck, M. (2021). Cone-shaped HIV-1 capsids are transported through intact nuclear pores. *Cell 184*, 1032–1046.

(*99*) Müller, T. G., Zila, V., Müller, B., and Kräusslich, H.-G. (2022). Nuclear Capsid Uncoating and Reverse Transcription of HIV-1. *Annual Review of Virology 9*, 261–284.

(*100*) Zila, V., Müller, T. G., Laketa, V., Müller, B., and Kräusslich, H. G. (2019). Analysis of CA Content and CPSF6 Dependence of Early HIV-1 Replication Complexes in SupT1-R5 Cells. *mBio 10*, DOI: `10.1128/MBIO.02501-19`.

(*101*) Peng, K., Muranyi, W., Glass, B., Laketa, V., Yant, S. R., Tsai, L., Cihlar, T., Müller, B., and Kräusslich, H. G. (2014). Quantitative microscopy of functional HIV post-entry complexes reveals association of replication with the viral capsid. *eLife 3*, e04114.

(*102*) Burdick, R. C., Li, C., Munshi, M. H., Rawson, J. M., Nagashima, K., Hu, W. S., and Pathak, V. K. (2020). HIV-1 uncoats in the nucleus near sites of integration. *Proceedings of the National Academy of Sciences of the United States of America 117*, 5486–5493.

(*103*) Müller, T. G., Zila, V., Peters, K., Schifferdecker, S., Stanic, M., Lucic, B., Laketa, V., Lusic, M., Müller, B., and Kräusslich, H. G. (2021). HIV-

1 uncoating by release of viral cDNA from capsid-like structures in the nucleus of infected cells. *eLife 10*, DOI: 10.7554/ELIFE.64776.

(*104*) Schifferdecker, S., Zila, V., Müller, T. G., Sakin, V., Anders-Össwein, M., Laketa, V., Kräusslich, H. G., and Müller, B. (2022). Direct Capsid Labeling of Infectious HIV-1 by Genetic Code Expansion Allows Detection of Largely Complete Nuclear Capsids and Suggests Nuclear Entry of HIV-1 Complexes via Common Routes. *mBio 13*, e0195922.

(*105*) Isel, C., Ehresmann, C., and Marquet, R. (2010). Initiation of HIV Reverse Transcription. *Viruses 2*, 213–243.

(*106*) Lesbats, P., Engelman, A. N., and Cherepanov, P. (2016). Retroviral DNA Integration. *Chemical reviews 116*, 12730–12757.

(*107*) Beemon, K. L. (2022). Retroviral RNA Processing. *Viruses 14*, DOI: 10.3390/V14051113.

(*108*) Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., and Varmus, H. E. (1988). Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature 331*, 280–283.

(*109*) Jacks, T., Madhani, H. D., Masiarz, F. R., and Varmus, H. E. (1988). Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell 55*, 447–458.

(*110*) Carlson, B. A., Kwon, S. Y., Chamorro, M., Oroszlan, S., Hatfield, D. L., and Lee, B. J. (1999). Transfer RNA modification status influences retroviral ribosomal frameshifting. *Virology 255*, 2–8.

(*111*) Sundquist, W. I., and Kräusslich, H.-G. (2012). HIV-1 Assembly , Budding , and Maturation. *Cold Spring Harbor Perspectives in Medicine 2*, DOI: 10.1101/cshperspect.a006924.

(*112*) Flügel, R. M., and Pfrepper, K. I. (2003). Proteolytic processing of foamy virus Gag and Pol proteins. *Current topics in microbiology and immunology 277*, 63–88.

(*113*) Bussienne, C., Marquet, R., Paillart, J.-C., and Bernacchi, S. (2021). Post-Translational Modifications of Retroviral HIV-1 Gag Precursors: An Overview of Their Biological Role. *Int. J. Mol. Sci 22*, DOI: 10.3390/ijms22062871.

(*114*) Inlora, J., Collins, D. R., Trubin, M. E., Chung, J. Y. J., and Ono, A. (2014). Membrane Binding and Subcellular Localization of Retroviral Gag Proteins Are Differentially Regulated by MA Interactions with Phosphatidylinositol-(4,5)-Bisphosphate and RNA. *mBio 5*, DOI: 10.1128/mBio.02202-14.

(*115*) Martin, J. L., Mendonça, L. M., Angert, I., Mueller, J. D., Zhang, W., and Mansky, L. M. (2017). Disparate Contributions of Human Retrovirus Capsid Subdomains to Gag- Gag Oligomerization, Virus Morphology, and Particle Biogenesis. *Journal of Virology 91*, 00298–17.

(*116*) Mailler, E., Bernacchi, S., Marquet, R., Paillart, J. C., Vivet-Boudou, V., and Smyth, R. P. (2016). The life-cycle of the HIV-1 gag–RNA complex. *Viruses 8*, 1–19.

(*117*) Chen, J., Liu, Y., Wu, B., Nikolaitchik, O. A., Mohan, P. R., Chen, J., Pathak, V. K., and Hu, W.-S. (2020). Visualizing the translation and packaging of HIV-1 full-length RNA. *Proceedings of the National Academy of Sciences 117*, 6145–6155.

(*118*) Chen, J., Rahman, S. A., Nikolaitchik, O. A., Grunwald, D., Sardo, L., Burdick, R. C., Plisov, S., Liang, E., Tai, S., Pathak, V. K., and Hu, W.-S. (2016). HIV-1 RNA genome dimerizes on the plasma membrane in the presence of Gag protein. *Proceedings of the National Academy of Sciences of the United States of America 113*, E201–E208.

(*119*) Kutluay, S. B., Zang, T., Blanco-Melo, D., Powell, C., Jannain, D., Errando, M., and Bieniasz, P. D. (2014). Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis. *Cell 159*, 1096–1109.

(*120*) Todd, G. C., Duchon, A., Inlora, J., Olson, E. D., Musier-Forsyth, K., and Ono, A. (2017). Inhibition of HIV-1 Gag-membrane interactions by specific RNAs. *RNA (New York, N.Y.) 23*, 395–405.

(*121*) Byland, R., Vance, P. J., Hoxie, J. A., and Marsh, M. (2007). A conserved dileucine motif mediates clathrin and AP-2-dependent endocytosis of the HIV-1 envelope protein. *Molecular biology of the cell 18*, 414–425.

(*122*) Chudak, C., Beimforde, N., George, M., Zimmermann, A., Lausch, V., Hanke, K., and Bannert, N. (2013). Identification of late assembly domains of the human endogenous retrovirus-K(HML-2). *Retrovirology 10*, DOI: `10.1186/1742-4690-10-140`.

(*123*) Woodward, C. L., Cheng, S. N., and Jensen, G. J. (2015). Electron Cryotomography Studies of Maturing HIV-1 Particles Reveal the Assembly Pathway of the Viral Core. *Journal of Virology 89*, 1267–1277.

(*124*) Carlson, L. A., de Marco, A., Oberwinkler, H., Habermann, A., Briggs, J. A., Kräusslich, H. G., and Grünewald, K. (2010). Cryo electron tomography of native HIV-1 budding sites. *PLoS Pathogens 6*, e1001173.

(*125*) Mattei, S., Anders, M., Konvalinka, J., Kräusslich, H.-G., Briggs, J. A. G., Müller, B., Krausslich, H.-G., Briggs, J. A. G., and Muller, B. (2014). Induced Maturation of Human Immunodeficiency Virus. *Journal of Virology 88*, 13722–13731.

(*126*) Mattei, S., Tan, A., Glass, B., Müller, B., Kräusslich, H.-G., and Briggs, J. A. (2018). High-resolution structures of HIV-1 Gag cleavage mutants determine structural switch for virus maturation. *Proceedings of the National Academy of Sciences of the United States of America 115*, E9401–E9410.

(*127*) Chojnacki, J., Staudt, T., Glass, B., Bingen, P., Engelhardt, J., Anders, M., Schneider, J., Müller, B., Hell, S. W., and Kräusslich, H. G. (2012). Maturation-dependent HIV-1 surface protein redistribution revealed by fluorescence nanoscopy. *Science 338*, 524–528.

(*128*) Frank, G. A., Narayan, K., Bess, J. W., Del Prete, G. Q., Wu, X., Moran, A., Hartnell, L. M., Earl, L. A., Lifson, J. D., and Subramaniam, S. (2015). Maturation of the HIV-1 core by a non-diffusional phase transition. *Nature Communications 6*, DOI: `10.1038/ncomms6854`.

(*129*) Meng, X., Zhao, G., Yufenyuy, E., Ke, D., Ning, J., DeLucia, M., Ahn, J., Gronenborn, A. M., Aiken, C., and Zhang, P. (2012). Protease Cleavage Leads to Formation of Mature Trimer Interface in HIV-1 Capsid. *PLoS Pathogens 8*, 1002886.

(*130*) Ning, J., Erdemci-Tandogan, G., Yufenyuy, E. L., Wagner, J., Himes, B. A., Zhao, G., Aiken, C., Zandi, R., and Zhang, P. (2016). In vitro protease cleavage and computer simulations reveal the HIV-1 capsid maturation pathway. *Nature Communications 7*, 13689.

(*131*) Lee, Y. N., and Bieniasz, P. D. (2007). Reconstitution of an infectious human endogenous retrovirus. *PLoS Pathogens 3*, 0119–0130.

(*132*) Sfakianos, J. N., and Hunter, E. (2003). M-PMV capsid transport is mediated by Env/Gag interactions at the pericentriolar recycling endosome. *Traffic 4*, 671–680.

(*133*) Sfakianos, J. N., LaCasse, R. A., and Hunter, E. (2003). The M-PMV cytoplasmic targeting-retention signal directs nascent Gag polypeptides to a pericentriolar region of the cell. *Traffic 4*, 660–670.

(*134*) Krebs, A.-S., Liu, H.-F., Zhou, Y., Rey, J. S., Levintov, L., Shen, J., Howe, A., Perilla, J. R., Bartesaghi, A., and Zhang, P. (2023). Molecular architecture and conservation of an immature human endogenous retrovirus. *bioRxiv*, DOI: `10.1101/2023.06.07.544027`.

(*135*) Lander, E., Linton, L., Birren, B., and Nusbaum, C. (2001). Initial sequencing and analysis of the human genome. *Nature 409*, 860–921.

(*136*) Mager, D. L., and Stoye, J. P. (2015). Mammalian Endogenous Retroviruses. *Microbiology Spectrum 3*, DOI: `10.1128/MICROBIOLSPEC.MDNA3-0009-2014/ASSET/2D8F7365-58B5-43E8-9BF4-B541CF452408/ASSETS/GRAPHIC/MDNA3-0009-2014-FIG3.GIF`.

(*137*) Sugimoto, J., and Schust, D. J. (2009). Human endogenous retroviruses and the placenta. *Reproductive sciences 16*, 1023–1033.

(*138*) Dupressoir, A., Lavialle, C., and Heidmann, T. (2012). From ancestral infectious retroviruses to bona fide cellular genes: Role of the captured syncytins in placentation. *Placenta 33*, 663–671.

(*139*) Lebedev, Y. B., Belonovitch, O. S., Zybrova, N. V., Khil, P. P., Kurdyukov, S. G., Vinogradova, T. V., Hunsmann, G., and Sverdlov, E. D. (2000). Differences in HERV-K LTR insertions in orthologous loci of humans and great apes. *Gene 247*, 265–277.

(*140*) Bièche, I., Laurent, A., Laurendeau, I., Duret, L., Giovangrandi, Y., Frendo, J. L., Olivi, M., Fausser, J. L., Evain-Brion, D., and Vidaud, M. (2003). Placenta-specific INSL4 expression is mediated by a human endogenous retrovirus element. *Biology of reproduction 68*, 1422–1429.

(*141*) Subramanian, R. P., Wildschutte, J. H., Russo, C., and Coffin, J. M. (2011). Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. *Retrovirology 8*, 1–22.

(*142*) Callahan, R., Drohan, W., Tronick, S., and Schlom, J. (1982). Detection and cloning of human DNA sequences related to the mouse mammary tumor virus genome. *Proceedings of the National Academy of Sciences of the United States of America 79*, 5503–5507.

(*143*) Kumar, S., and Subramanian, S. (2002). Mutation rates in mammalian genomes. *Proceedings of the National Academy of Sciences of the United States of America 99*, 803–808.

(*144*) Stoye, J. P. (2001). Endogenous retroviruses: Still active after all these years? *Current biology 11*, R914–R916.

(*145*) Turner, G., Barbulescu, M., Su, M., Jensen-Seaman, M. I., Kidd, K. K., and Lenz, J. (2001). Insertional polymorphisms of full-length endogenous retroviruses in humans. *Current biology 11*, 1531–1535.

(*146*) Jha, A. R., Pillai, S. K., York, V. A., Sharp, E. R., Storm, E. C., Wachter, D. J., Martin, J. N., Deeks, S. G., Rosenberg, M. G., Nixon, D. F., and Garrison, K. E. (2009). Cross-sectional dating of novel haplotypes of HERV-K 113 and HERV-K 115 indicate these proviruses originated in Africa before Homo sapiens. *Molecular biology and evolution 26*, 2617–2626.

(*147*) Dewannieux, M., Harper, F., Richaud, A., Letzelter, C., Ribet, D., Pierron, G., and Heidmann, T. (2006). Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements. *Genome Research 16*, 1548–1556.

(*148*) Manghera, M., and Douville, R. N. (2013). Endogenous retrovirus-K promoter: A landing strip for inflammatory transcription factors? *Retrovirology 10*, 1–11.

(*149*) Li, W., Pandya, D., Pasternack, N., Garcia-Montojo, M., Henderson, L., Kozak, C. A., and Nath, A. (2022). Retroviral Elements in Pathophysiology and as Therapeutic Targets for Amyotrophic Lateral Sclerosis. *Neurotherapeutics*, 1–17.

(*150*) Douville, R., Liu, J., Rothstein, J., and Nath, A. (2011). Identification of active loci of a human endogenous retrovirus in neurons of patients with amyotrophic lateral sclerosis. *Annals of Neurology 69*, 141–151.

(*151*) Li, W. et al. (2015). Human endogenous retrovirus-K contributes to motor neuron disease. *Science Translational Medicine 7*, DOI: 10.1126/scitranslmed.aac8201.

(*152*) Burn, A., Roy, F., Freeman, M., and Coffin, J. M. (2022). Widespread expression of the ancient HERV-K (HML-2) provirus group in normal human tissues. *PLoS biology 20*, e3001826.

(*153*) Blanco-Melo, D., Gifford, R. J., and Bieniasz, P. D. (2017). Co-option of an endogenous retrovirus envelope for host defense in hominid ancestors. *eLife 6*, e22519.

(*154*) Contreras-Galindo, R., Kaplan, M. H., Markovitz, D. M., Lorenzo, E., and Yamamura, Y. (2006). Detection of HERV-K(HML-2) viral RNA in plasma of HIV type 1-infected individuals. *AIDS research and human retroviruses 22*, 979–984.

(*155*) Bhardwaj, N., Maldarelli, F., Mellors, J., and Coffin, J. M. (2014). HIV-1 infection leads to increased transcription of human endogenous retrovirus HERV-K (HML-2) proviruses in vivo but not to increased virion production. *Journal of virology 88*, 11108–11120.

(*156*) Contreras-Galindo, R., Lorenzo, E., Kaplan, M. H., Gonzalez-Hernandez, M. J., Giusti, F., Dosik, M. H., Gitlin, S. D., Ferlenghi, I., Markovitz, D. M., Yamamura, Y., and Contreras-Galindo, A. C. (2012). Characterization of Human Endogenous Retroviral Elements in the Blood of HIV-1-Infected Individuals. *Journal of Virology 86*, 262–276.

(*157*) Young, G. R., Terry, S. N., Manganaro, L., Cuesta-Dominguez, A., Deikus, G., Bernal-Rubio, D., Campisi, L., Fernandez-Sesma, A., Sebra, R., Simon, V., and Mulder, L. C. F. (2018). HIV-1 Infection of Primary CD4 + T Cells Regulates the Expression of Specific Human Endogenous Retrovirus HERV-K (HML-2) Elements. *Journal of Virology 92*, 01507–17.

(*158*) Monde, K., Contreras-Galindo, R., Kaplan, M. H., Markovitz, D. M., and Ono, A. (2012). Human endogenous retrovirus K Gag coassembles with HIV-1 Gag and reduces the release efficiency and infectivity of HIV-1. *Journal of virology 86*, 11194–11208.

(*159*) Monde, K., Terasawa, H., Nakano, Y., Soheilian, F., Nagashima, K., Maeda, Y., and Ono, A. (2017). Molecular mechanisms by which HERV-K Gag interferes with HIV-1 Gag assembly and particle infectivity. *Retrovirology 14*, 1–16.

(*160*) Maksakova, I. A., Mager, D. L., and Reiss, D. (2008). Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. *Cellular and molecular life sciences 65*, 3329–3347.

(*161*) Kämmerer, U., Germeyer, A., Stengel, S., Kapp, M., and Denner, J. (2011). Human endogenous retrovirus K (HERV-K) is expressed in villous and extravillous cytotrophoblast cells of the human placenta. *Journal of Reproductive Immunology 91*, 1–8.

(*162*) Autio, A., Nevalainen, T., Mishra, B. H., Jylhä, M., Flinck, H., and Hurme, M. (2020). Effect of aging on the transcriptomic changes associated with the expression of the HERV-K (HML-2) provirus at 1q22. *Immunity & Ageing 17*, DOI: `10.1186/S12979-020-00182-0`.

(*163*) Liu, X. et al. (2023). Resurrection of endogenous retroviruses during aging reinforces senescence. *Cell 186*, 1–18.

(*164*) Copley, K. E., and Shorter, J. (2023). Repetitive elements in aging and neurodegeneration. *Trends in genetics ahead of print*, DOI: `10.1016/J.TIG.2023.02.008`.

(*165*) Nakane, T. et al. (2020). Single-particle cryo-EM at atomic resolution. *Nature 587*, 152–156.

(*166*) Pal, S. (2020). Macromolecular assemblies. *Fundamentals of Molecular Structural Biology*, 449–463.

(*167*) Galaz-Montoya, J. G., and Ludtke, S. J. (2017). The advent of structural biology in situ by single particle cryo-electron tomography. *Biophysics Reports 3*, 17–35.

(*168*) Bouvette, J., Liu, H. F., Du, X., Zhou, Y., Sikkema, A. P., da Fonseca Rezende e Mello, J., Klemm, B. P., Huang, R., Schaaper, R. M., Borgnia, M. J., and Bartesaghi, A. (2021). Beam image-shift accelerated data acquisition for near-atomic resolution single-particle cryo-electron tomography. *Nature communications 12*, DOI: `10.1038/S41467-021-22251-8`.

(*169*) Bartesaghi, A., Lecumberry, F., Sapiro, G., and Subramaniam, S. (2012). Protein secondary structure determination by constrained single-particle cryo-electron tomography. *Structure 20*, 2003–2013.

(*170*) Bartesaghi, A., Sapiro, G., and Subramaniam, S. (2005). An energy-based three-dimensional segmentation approach for the quantitative interpretation of electron tomograms. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society 14*, 1314–1323.

(*171*) Bartesaghi, A., Sapiro, G., Lee, S., Lefman, J., Wahl, S., Orenstein, J., and Subramaniam, S. (2004). A new approach for 3D segmentation of cellular tomograms obtained using three-dimensional electron microscopy. *2nd IEEE International Symposium on Biomedical Imaging: Macro to Nano 2*, 5–8.

(*172*) Chen, M., Bell, J. M., Shi, X., Sun, S. Y., Wang, Z., and Ludtke, S. J. (2019). A complete data processing workflow for cryo-ET and subtomogram averaging. *Nature methods 16*, 1161–1168.

(*173*) Stuart, D. I., Levine, M., Muirhead, H., and Stammers, D. K. (1979). Crystal structure of cat muscle pyruvate kinase at a resolution of 2.6 Å. *Journal of Molecular Biology 134*, 109–142.

(*174*) Madeira, F., Pearce, M., Tivey, A. R., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., and Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Research 50*, W276–W279.

(*175*) Zábranský, A., Hoboth, P., Hadravová, R., Štokrová, J., Sakalian, M., and Pichová, I. (2010). The Noncanonical Gag Domains p8 and n Are Critical for Assembly and Release of Mouse Mammary Tumor Virus. *Journal of Virology 84*, 11555–11559.

(*176*) Sakalian, M., and Hunter, E. (1999). Separate Assembly and Transport Domains within the Gag Precursor of Mason-Pfizer Monkey Virus. *Journal of Virology 73*, 8073–8082.

(*177*) Gottwein, E., Bodem, J., Müller, B., Schmechel, A., Zentgraf, H., and Kräusslich, H.-G. (2003). The Mason-Pfizer Monkey Virus PPPY and

PSAP Motifs Both Contribute to Virus Release. *Journal of Virology 77*, 9474–9485.

(*178*) Dilley, K. A., Gregory, D., Johnson, M. C., and Vogt, V. M. (2010). An LYPSL Late Domain in the Gag Protein Contributes to the Efficient Release and Replication of Rous Sarcoma Virus. *Journal of Virology 84*, 6276–6287.

(*179*) Xiang, Y., Cameron, C. E., Wills, J. W., and Leis, J. (1996). Fine mapping and characterization of the Rous sarcoma virus Pr76gag late assembly domain. *Journal of Virology 70*, 5695–5700.

(*180*) Bartusch, C., and Prange, R. (2016). ESCRT Requirements for Murine Leukemia Virus Release. *Viruses 8*, DOI: `10.3390/V8040103`.

(*181*) Obr, M., Hagen, W. J., Dick, R. A., Yu, L., Kotecha, A., and Schur, F. K. (2022). Exploring high-resolution cryo-ET and subtomogram averaging capabilities of contemporary DEDs. *Journal of structural biology 214*, 107852.

(*182*) Song, Y., Dimaio, F., Wang, R. Y. R., Kim, D., Miles, C., Brunette, T., Thompson, J., and Baker, D. (2013). High-resolution comparative modeling with RosettaCM. *Structure 21*, 1735–1742.

(*183*) Strohalmová-Bohmová, K., Spiwok, V., Lepšík, M., Hadravová, R., Křížová, I., Ulbrich, P., Pichová, I., Bednárová, L., Ruml, T., and Rumlová, M. (2014). Role of Mason-Pfizer Monkey Virus CA-NC Spacer Peptide-Like Domain in Assembly of Immature Particles. *Journal of Virology 88*, 14148–14160.

(*184*) Schur, F. K. M., Dick, R. A., Hagen, W. J. H., Vogt, V. M., and Briggs, J. A. G. (2015). The Structure of Immature Virus-Like Rous Sarcoma Virus Gag Particles Reveals a Structural Role for the p10 Domain in Assembly. *Journal of Virology 89*, 10294–10302.

(*185*) Ng, W. M., Stelfox, A. J., and Bowden, T. A. (2020). Unraveling virus relationships by structure-based phylogenetic classification. *Virus evolution 6*, DOI: `10.1093/VE/VEAA003`.

(*186*) Watanabe, Y. et al. (2021). Native-like SARS-CoV-2 Spike Glycoprotein Expressed by ChAdOx1 nCoV-19/AZD1222 Vaccine. *ACS Central Science 7*, 594–602.

(*187*) Mendonça, L. et al. (2021). Correlative multi-scale cryo-imaging unveils SARS-CoV-2 assembly and egress. *Nature communications 12*, DOI: `10.1038/s41467-021-24887-y`.

(*188*) Li, F. (2016). Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annual review of virology 3*, 237–261.

(*189*) Hardenbrook, N. J., and Zhang, P. (2022). A structural view of the SARS-CoV-2 virus and its assembly. *Current opinion in virology 52*, 123–134.

(*190*) Wolf, K. A., Kwan, J. C., and Kamil, J. P. (2022). Structural Dynamics and Molecular Evolution of the SARS-CoV-2 Spike Protein. *mBio 13*, DOI: `10.1128/MBIO.02030-21`.

(*191*) Xu, C. et al. (2021). Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. *Science advances 7*, DOI: `10.1126/SCIADV.ABE5575`.

(*192*) Cai, Y., Zhang, J., Xiao, T., Peng, H., Sterling, S. M., Walsh, R. M., Rawson, S., Rits-Volloch, S., and Chen, B. (2020). Distinct conformational states of SARS-CoV-2 spike protein. *Science 369*, DOI: `10.1126/SCIENCE.ABD4251`.

(*193*) Fan, X., Cao, D., Kong, L., and Zhang, X. (2020). Cryo-EM analysis of the post-fusion structure of the SARS-CoV spike glycoprotein. *Nature communications 11*, DOI: `10.1038/S41467-020-17371-6`.

(*194*) Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., and Wang, X. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature 581*, 215–220.

(*195*) Nieto-Torres, J. L., DeDiego, M. L., Álvarez, E., Jiménez-Guardeño, J. M., Regla-Nava, J. A., Llorente, M., Kremer, L., Shuo, S., and Enjuanes, L. (2011). Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein. *Virology 415*, 69–82.

(*196*) Cao, Y., Yang, R., Lee, I., Zhang, W., Sun, J., Wang, W., and Meng, X. (2021). Characterization of the SARS-CoV-2 E Protein: Sequence, Structure, Viroporin, and Inhibitors. *Protein science 30*, 1114–1130.

(*197*) Liang, J. Q., Fang, S., Yuan, Q., Huang, M., Chen, R. A., Fung, T. S., and Liu, D. X. (2019). N-Linked glycosylation of the membrane protein ectodomain regulates infectious bronchitis virus-induced ER stress response, apoptosis and pathogenesis. *Virology 531*, 48–56.

(*198*) Wang, Q., Zhang, Y., Wu, L., Niu, S., Song, C., Zhang, Z., Lu, G., Qiao, C., Hu, Y., Yuen, K. Y., Wang, Q., Zhou, H., Yan, J., and Qi, J. (2020).

Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell 181*, 894–904.

(*199*) Whittaker, G. R., Daniel, S., and Millet, J. K. (2021). Coronavirus entry: how we arrived at SARS-CoV-2. *Current opinion in virology 47*, 113–120.

(*200*) Tang, T., Bidon, M., Jaimes, J. A., Whittaker, G. R., and Daniel, S. (2020). Coronavirus membrane fusion mechanism offers a potential target for antiviral development. *Antiviral research 178*, DOI: `10.1016/J.ANTIVIRAL.2020.104792`.

(*201*) Matsuyama, S. et al. (2020). Enhanced isolation of SARS-CoV-2 by TMPRSS2-expressing cells. *Proceedings of the National Academy of Sciences of the United States of America 117*, 7001–7003.

(*202*) Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N. H., Nitsche, A., Müller, M. A., Drosten, C., and Pöhlmann, S. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell 181*, 271–280.

(*203*) Simmons, G., Reeves, J. D., Rennekamp, A. J., Amberg, S. M., Piefer, A. J., and Bates, P. (2004). Characterization of severe acute respiratory syndrome-associated coronavirus (SARS-CoV) spike glycoprotein-mediated viral entry. *Proceedings of the National Academy of Sciences of the United States of America 101*, 4240–4245.

(*204*) Doyle, N., Neuman, B. W., Simpson, J., Hawes, P. C., Mantell, J., Verkade, P., Alrashedi, H., and Maier, H. J. (2018). Infectious Bronchitis Virus

Nonstructural Protein 4 Alone Induces Membrane Pairing. *Viruses 10*, DOI: 10.3390/V10090477.

(*205*) Hagemeijer, M. C., Monastyrska, I., Griffith, J., van der Sluijs, P., Voortman, J., van Bergen en Henegouwen, P. M., Vonk, A. M., Rottier, P. J., Reggiori, F., and De Haan, C. A. (2014). Membrane rearrangements mediated by coronavirus nonstructural proteins 3 and 4. *Virology 458*, 125–135.

(*206*) Oudshoorn, D., Rijs, K., Limpens, R. W., Groen, K., Koster, A. J., Snijder, E. J., Kikkert, M., and Bárcena, M. (2017). Expression and Cleavage of Middle East Respiratory Syndrome Coronavirus nsp3-4 Polyprotein Induce the Formation of Double-Membrane Vesicles That Mimic Those Associated with Coronaviral RNA Replication. *mBio 8*, 01658–17.

(*207*) Klein, S., Cortese, M., Winter, S. L., Wachsmuth-Melm, M., Neufeldt, C. J., Cerikan, B., Stanifer, M. L., Boulant, S., Bartenschlager, R., and Chlanda, P. (2020). SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *Nature communications 11*, DOI: 10.1038/S41467-020-19619-7.

(*208*) Snijder, E. J., Limpens, R. W., de Wilde, A. H., de Jong, A. W., Zevenhoven-Dobbe, J. C., Maier, H. J., Faas, F. F., Koster, A. J., and Bárcena, M. (2020). A unifying structural and functional model of the coronavirus replication organelle: Tracking down RNA synthesis. *PLoS biology 18*, DOI: 10.1371/JOURNAL.PBIO.3000715.

(*209*) Goldsmith, C. S., Tatti, K. M., Ksiazek, T. G., Rollin, P. E., Comer, J. A., Lee, W. W., Rota, P. A., Bankamp, B., Bellini, W. J., and Zaki, S. R.

(2004). Ultrastructural characterization of SARS coronavirus. *Emerging infectious diseases 10*, 320–326.

(*210*) Stertz, S., Reichelt, M., Spiegel, M., Kuri, T., Martínez-Sobrido, L., García-Sastre, A., Weber, F., and Kochs, G. (2007). The intracellular sites of early replication and budding of SARS-coronavirus. *Virology 361*, 304–315.

(*211*) Ghosh, S. et al. (2020). β-Coronaviruses Use Lysosomes for Egress Instead of the Biosynthetic Secretory Pathway. *Cell 183*, 1520–1535.

(*212*) Ramasamy, M. N. et al. (2020). Safety and immunogenicity of ChAdOx1 nCoV-19 vaccine administered in a prime-boost regimen in young and old adults (COV002): a single-blind, randomised, controlled, phase 2/3 trial. *The Lancet 396*, 1979–1993.

(*213*) Toelzer, C. et al. (2020). Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein. *Science (New York, N.Y.) 370*, 725–730.

(*214*) Wrobel, A. G., Benton, D. J., Xu, P., Roustan, C., Martin, S. R., Rosenthal, P. B., Skehel, J. J., and Gamblin, S. J. (2020). SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nature structural & molecular biology 27*, 763–767.

(*215*) Bennett, A. E., Narayan, K., Shi, D., Hartnell, L. M., Gousset, K., He, H., Lowekamp, B. C., Yoo, T. S., Bliss, D., Freed, E. O., and Subramaniam, S. (2009). Ion-abrasion scanning electron microscopy reveals surface-connected tubular conduits in HIV-infected macrophages. *PLoS pathogens 5*, DOI: `10.1371/JOURNAL.PPAT.1000591`.

(*216*) Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell 181*, 281–292.

(*217*) Ogando, N. S., Dalebout, T. J., Zevenhoven-Dobbe, J. C., Limpens, R. W., van der Meer, Y., Caly, L., Druce, J., de Vries, J. J., Kikkert, M., Barcena, M., Sidorov, I., and Snijder, E. J. (2020). SARS-coronavirus-2 replication in Vero E6 cells: replication kinetics, rapid adaptation and cytopathology. *The Journal of general virology 101*, 925–940.

(*218*) Ng, M. L., Lee, J. W., Leong, M. L., Ling, A. E., Tan, H. C., and Ooi, E. E. (2004). Topographic changes in SARS coronavirus-infected cells at late stages of infection. *Emerging infectious diseases 10*, 1907–1914.

(*219*) Lu, M. et al. (2022). SARS-CoV-2 prefusion spike protein stabilized by six rather than two prolines is more potent for inducing antibodies that neutralize viral variants of concern. *Proceedings of the National Academy of Sciences of the United States of America 119*, DOI: `10.1073/PNAS.2110105119`.