



HAL
open science

Analyse structurale et fonctionnelle de la réplication et de la transcription des Bunyavirales

Benoît Arragain

► **To cite this version:**

Benoît Arragain. Analyse structurale et fonctionnelle de la réplication et de la transcription des Bunyavirales. Biologie structurale [q-bio.BM]. Université Grenoble Alpes [2020-..], 2021. Français. NNT : 2021GRALV039 . tel-03779594

HAL Id: tel-03779594

<https://theses.hal.science/tel-03779594v1>

Submitted on 17 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : Biologie Structurale et Nanobiologie

Arrêté ministériel : 25 mai 2016

Présentée par

Benoît ARRAGAIN

Thèse dirigée par **Guy SCHOEHN**
et co-encadrée par **Hélène MALET**

préparée au sein du **Laboratoire Institut de Biologie Structurale**
dans **l'École Doctorale Chimie et Sciences du Vivant**

Analyse structurale et fonctionnelle de la réplication et de la transcription des Bunyavirales

Structural and functional analysis of Bunyavirales replication and transcription

Thèse soutenue publiquement le **17 septembre 2021**
devant le jury composé de :

Monsieur Guy SCHOEHN

DIRECTEUR DE RECHERCHE, CNRS délégation Alpes, Directeur de thèse

Madame Célia PLISSON-CHASTANG

CHARGE DE RECHERCHE HDR, CNRS délégation Occitanie Ouest, Rapporteuse

Monsieur Bruno CANARD

DIRECTEUR DE RECHERCHE, CNRS délégation Provence et Corse, Rapporteur

Monsieur Felix REY

DIRECTEUR DE RECHERCHE, Institut Pasteur, Examineur

Monsieur Stephen CUSACK

DIRECTEUR DE RECHERCHE, European Molecular Biology Laboratory, Examineur

Madame Cécile BREYTON

DIRECTRICE DE RECHERCHE, CNRS délégation Alpes, Présidente



Remerciements

Nous y voilà, la dernière partie à écrire, la plus « compliquée », et celle qui sera probablement la plus lue, pour peu que quelqu'un tombe sur cette thèse un jour ! Donc si par mégarde, toi lecteur, tu lis ces quelques lignes, sache que les 4 années et quelques mois passés à l'IBS dans le groupe de microscopie électronique et méthodes, MEM pour les intimes, ont été (de manière générale évidemment) exceptionnelles d'un point de vue des relations humaines et de la curiosité scientifique : d'un point de vue des relations humaines parce que j'ai rencontré et travaillé avec des personnes formidables avec qui j'ai tissé de belles relations, et d'un point de vue scientifique parce que la microscopie électronique et les polymérasés virales, c'est quand même vachement sexy (les nucléocapsides aussi, mais un peu moins je dois l'avouer).

Maintenant, et tu l'auras probablement déjà compris en lisant le premier paragraphe, ce gros pavé de presque 300 pages - un peu inutile, on peut maintenant se l'avouer parce qu'il ne reflète que la partie professionnelle de cette thèse - ne résulte pas seulement du travail que j'ai pu fournir mais des personnes qui m'ont accompagné et soutenu tout au long de cette aventure. C'est donc pour ça que je tiens à remercier chaleureusement pas mal de monde et tout particulièrement :

Ma chef, Hélène, ou CTF pour les intimes, qui m'a (quasiment) tout appris, qui m'a accompagné tout au long de ces années et qui a eu la plus grande des patiences avec moi (j'ai été patient aussi par moment, attention !). Je vais essayer de ne pas trop faire dans l'eau de rose mais tu as toujours été là pour me pousser, me soutenir et faire en sorte que tout se passe pour le mieux, et ce depuis le M2, et je t'en remercie. Au final, je pense qu'on a formé un duo professionnel de fou furieux et tout ce que toi, lecteur, tu t'apprêtes à lire si t'as la motiv', est le fruit d'une belle entente malgré les moments houleux de la deuxième année et tu vois desquels je parle Hélène ;-) !

Mon big boss, Guy, le biologiste de l'ombre, ma figure d'autorité (!) qui a tiré pas mal de ficelles en coulisse et qui a toujours fait en sorte que les choses se passent pour le mieux, dans le respect des règles et tu sais que j'adore ne pas les respecter,

Remerciements

surtout celles qui ne font pas de sens. Plus sérieusement, j'ai eu la chance d'évoluer en même temps que la plateforme de microscopie quand je suis arrivé en 2017 et si cette thèse s'est déroulée correctement c'est grâce à tout le travail que tu as abattu pour que nous, au sens de la communauté ;-), puissions profiter de machines fonctionnelles et de haute qualité. Enfin, je te dis un grand merci pour m'avoir laissé ma chance, avec encore en tête notre tout premier entretien avec Hélène, et fait confiance pendant toutes ces années et de continuer à le faire !

Mon ancienne chef de M2 pour boucler le trio de choc, Irina, qui m'a permis de faire mon stage au sein de l'équipe MMBV puis de me permettre de commencer ma thèse avec Hélène. Évidemment Irina, au-delà du simple aspect financier et tu vois de quoi je parle ;-), je te remercie pour le soutien que tu as pu apporter à Hélène et moi tout au long de ces années.

Les membres de mon jury de thèse qui ont eu la patience de lire ces nombreuses pages et accepté d'évaluer ce travail : Célia Plisson-Chastang, Bruno Canard, Félix Rey, Stephen Cusack et Cécile Breyton.

Les membres de mon CSI, Juan Reguera et Rob Ruigrok, qui ont toujours été de bons conseils au fil des ans, je n'ai pas rendu l'argent et j'ai bien fait.

Les loufous du bureau 364, bureau le plus excitant et animé de l'IBS sans aucun doute, qui sont tous aussi fous les uns que les autres et avec qui j'ai passé la majeure partie de mon temps. Je pense notamment à Alister avec qui j'ai commencé ma thèse et partagé des discussions enflammées sur le système administratif français, l'encadrement de nos thèses respectives ou un peu plus orientées traitements d'images, parti trop tôt au pays des haricots et membre de la #teamtomo basse résolution ; Yoan, notre nouvelle *sex machine* à la Rasputin qui adore dormir avec Martine ; Ladislav, qui a passé un petit bout de temps à nos côtés et dont le *stick* est fantastique ; notre voisin de bureau qui a toujours passé plus de temps dans le nôtre que dans le sien, Romain (au moins tu as terminé le « matériel et méthodes » avant que j'écrive les remerciements, l'honneur est sauf...) et le meilleur pour la fin, notre gourou au grand cœur, notre maître à tous, un exemple de ponctualité, de rigueur et d'efficacité (les deux dernières qualités sont vraies quand il est là), Ambroise, l'un des membres fondateurs de la mafia, toujours à l'écoute et présent dans les bons comme les mauvais moments

surtout quand tu ramènes un shawarma qui empeste le bureau, ou pire.... Au-delà des conneries que je raconte juste au-dessus, j'ai énormément appris à tes côtés, que ça soit scientifiquement ou humainement, donc un grand merci mon loulou et pleins de bisous. Maintenant, malgré le poste permanent et le statut de chef de groupe par intérim, j'espère que les apéros bureau perdureront encore longtemps ! Et pas de panique, ce qui s'est passé au chalet, reste au chalet.

Je ne peux pas remercier Ambroise sans remercier Leandro parce qu'ils font la paire et qu'il est également un membre fondateur de la mafia. Merci d'avoir été notre physicien-psychologue-danseur pendant ces quelques années ainsi que pour tous les bons moments qu'on a passés ensemble et qu'on passe encore.

Mes deux anges gardiens du labo, Karine et Angélique, qui ont toujours été là pour m'épauler et qui m'ont énormément appris du M2 jusqu'à la fin de ma thèse.

Les membres du groupe MEM/MMBV-MICA et alumni, je pense notamment à Matt, Jan et Clarissa qui étaient là à mes débuts et qui se sont maintenant envolés vers de nouveaux horizons, le général Moriscot, Catherine (Manue) et Liliane (Daphna), Benoît le gendre idéal, Grég' la force tranquille, Dominique notre aîné, Pascal et Christopher dont l'équipe n'a toujours pas été rapatriée au château, Quentin la relève, Camille M. étudiante en M1 qui a dû me supporter (ou l'inverse), Lorenzo, Lorène, Maria et Madalen !

Le chevreuil libérateur, fidèle compagnon nocturne avec qui j'ai fermé l'IBS tard le soir de nombreuses fois quand il était encore postdoc mais qui préfère maintenant faire la petite sardine sur le canapé parce qu'il est permanent...

Aymeric, le maître du cluster, toujours disponible pour aider quand on a une m**** informatique et toujours prêt à dégainer des *scancel* ou à donner un coup de main pour imprimer des manuscrits de thèse à la dernière minute parce que les imprimantes font des siennes...

Les responsables de la plateforme d'expression en cellules eucaryotes à l'EMBL en commençant par Alice, qui a toujours été présente pour me dépanner quand je lançais mes cultures à l'arrache et Martin le coquin qui a pris la suite je n'ai pas encore eu l'occasion de t'emmerder mais ça viendra peut-être pendant le postdoc.

Remerciements

Michael, Grég' pour les différentes collectes qu'on a pu réaliser sur le Krios CM01 à l'ESRF et Simon pour celle d'Heidelberg.

Jacques, notre ancien chef d'installation, qui a été plutôt compréhensif et conciliant avec ma présence les weekends, jours fériés ou pendant les HNO... et parce que le courant ne passe pas tout le temps avec tout le monde, merci à vous, vous vous reconnaitrez si vous lisez ces lignes, pour les leçons que vous m'avez transmises parce que l'on n'apprend pas seulement au contact des personnes que l'on apprécie et/ou qui sont dotés d'un minimum d'empathie !

Stephen pour m'avoir embauché en qualité de postdoc et me permettre de continuer à travailler sur des polymérases virales ainsi que les membres de mon nouveau groupe et les alumni qui ont pu me dépanner ou avec qui j'ai pu collaborer pendant ma thèse : Petra, Alice, Anna, Tomas, Joanna et Sissy.

Les ami(e)s de longue date, copains, copines, anciens et nouveaux collègues professionnels ou de beuveries, je cite : Charlie, Vincent, Andrea, Elena, Guillaume F., Sylvain, Barbara, Shekhar, Catarina, Laura T., Laura L., Audrey, François, Guillaume A., Cindy C., Nico, Martin et j'en oublie probablement quelques-uns...

Nina, pour avoir réussi à canaliser l'animal que je suis, notamment en fin de deuxième année..., et pour m'avoir apporté son soutien sans faille jusqu'au bout de cette aventure.

Tous les groupes à l'IBS auxquels j'ai pu piquer des trucs quand on n'avait pas trop de soussous au tout début de la thèse et un peu plus tard à l'EMBL, mentions spéciales à la caverne d'Alibaba du 2^{ème} étage et au stock d'ARNs de Piotr.

Les sponsors qui m'ont accompagné tout au long de ma thèse : Fleur du pays, OCB, Bic, Aberlour, Triple Carmélite, Jacques Vabre et le café gratuit du groupe MEM (promis, j'amènerai un chèque avant que je parte de Grenoble pour rembourser mes dettes).

Enfin, j'aimerais quand même dédier cette thèse à ma famille et tout particulièrement à mes parents : ma petite mum et mon père, qui, depuis que j'ai été parachuté sur cette planète, ont toujours fait en sorte de me donner les moyens de mon épanouissement.

J'allais oublier de remercier mon chat, Diesel, à qui j'ai pu me confier sans crainte pendant toutes ces années, une oreille attentive.

Et pour finir cette section en beauté, j'ajouterai juste une petite citation trouvée dans une papillote, attribuée à Oscar Wilde, et qui de mon point de vue correspond parfaitement à cette thèse :

« Les folies sont les seules choses qu'on ne regrette jamais »

Merci à toutes et tous, des gros bisous !

(Et bonne lecture pour les plus courageux...)

Table des matières

<i>Remerciements</i>	5
<i>Table des matières</i>	11
<i>Résumé en français</i>	15
<i>Résumé en anglais</i>	17
<i>Virologie et microscopie, à la croisée des chemins</i>	21
Chapitre I : Les Bunyavirus	25
1. Classification des virus	27
1.1. Les virus à ADN	28
1.2. Les virus à transcriptase inverse	29
1.3. Les virus à ARN	29
2. L'ordre des <i>Bunyavirales</i>	33
2.1. Des virus émergents.....	34
2.2. Organisation du virus et de son génome.....	37
2.3. Les NTRs du génome viral et leur implication dans la synthèse d'ARN	41
2.4. Morphologie	43
2.5. Le cycle viral.....	44
3. Les ARN polymérase ARN-dépendantes virales des sNSVs	47
3.1. Caractéristiques et fonctions principales.....	47
3.2. Le cœur catalytique des RdRps	47
3.3. Domaines catalytiques additionnels	59
4. Nucléoprotéines et nucléocapsides chez les sNSVs	68
4.1. Fonctions principales et structures cristallographiques.....	68
4.2. Organisation générale des NPs chez les sNSVs	70
4.3. Flexibilité des RNPs natives chez les sNSVs.....	71
Chapitre II : La cryo-ME en bref !	75
1. Le principe de congélation et de vitrification en cryo-ME	80
2. Formation de l'image en MET, contraste et CTF	83
3. Dommages liés aux radiations, effets de charge et mouvements ..	88
4. Les récentes améliorations matérielles et logiciels en cryo-ME	93
4.1. Au niveau du MET	93
4.2. Au niveau des filtres d'énergie	94
4.3. Au niveau des logiciels et stratégies de collectes automatiques utilisées.....	95
4.4. Au niveau des caméras	96
4.5. Conclusion.....	98
5. Le traitement d'images appliqué à l'analyse de particules uniques en cryo-ME	100
5.1. La correction des mouvements induits par le faisceau d'électron	100
5.2. L'estimation de la CTF.....	102

5.3.	La sélection, l'extraction et la classification 2D des particules	103
5.4.	La détermination d'une reconstruction 3D initiale	106
5.5.	Le principe de l'affinement 3D d'un modèle initial	110
5.6.	Les différentes options de classification 3D	113
5.7.	Post-processing et sharpening	117
5.8.	Le cas particulier des objets hélicoïdaux.....	119
Objectifs de thèse		125
Chapitre III : Résultats		129
1. Obtention de la structure haute résolution par cryo-ME de la HTNV-NC 131		
1.1.	Introduction	131
1.2.	Article n°1 : High resolution cryo-EM structure of the helical RNA-bound Hantaan virus nucleocapsid reveals its assembly mechanisms	135
1.3.	Potentiel mécanisme de déstabilisation de la HTNV-NC	161
1.4.	Optimisation de l'expression, de la purification et de la congélation de la HTNV-L	163
1.5.	Caractérisation structurale de la HTNV-L liée à l'extrémité 5' de l'ARNv en cryo-ME	167
1.6.	La HTNV-L et la LACV-L.....	171
2. La structure à haute résolution par cryo-ME de la LACV-L173		
2.1.	Introduction	173
2.2.	Article n°2 : Pre-initiation and elongation structures of full-length La Crosse virus polymerase reveal functionally important conformational changes	175
2.3.	Discussion	208
3. Caractérisation structurale et fonctionnelle de la LACV-L209		
3.1.	Introduction	209
3.2.	Article n°3 : Structural snapshots of La Crosse virus polymerase reveal the mechanisms underlying <i>Peribunyaviridae</i> replication and transcription	211
3.3.	Discussion	267
Chapitre IV : Conclusion et perspectives générales.....		275
Références.....		285

Résumé en français

L'ordre des *Bunyvirales* englobe plus de 450 espèces de virus à ARN négatif segmenté (sNSV). Répandus dans le monde entier et considérés comme émergents, l'infection de l'homme par certains Bunyavirus peut entraîner de graves maladies. Chez la famille des *Hantaviridae*, le virus Hantaan (HTNV) provoque des fièvres hémorragiques tandis que chez les *Peribunyaviridae*, le virus La Crosse (LACV) entraîne des cas d'encéphalites chez les enfants. De plus, l'ordre des *Bunyvirales* est proche d'autres sNSVs comme le virus influenza (*Orthomyxoviridae*), un agent infectieux responsable d'épidémies saisonnières.

Les génomes d'HTNV et de LACV sont constitués de trois segments codant pour quatre protéines structurales : les glycoprotéines (Gn et Gc), encodées sur le segment M, la nucléoprotéine (NP), encodée sur le segment S et la polymérase, aussi appelée protéine L, encodée sur le segment L. Au cours d'une infection, les NPs protègent le génome viral des dégradations et de la réponse immunitaire innée de la cellule infectée en multimérisant autour du génome viral. Ces complexes protéine-ARN forment des assemblages hélicoïdaux appelés nucléocapsides (NC). Les extrémités 5' et 3' de chacun des segments du génome viral (ARN_v) sont liées à une protéine L. Le complexe L-NP-ARN, appelé particule ribonucléoprotéique (RNP), représente l'unité fonctionnelle nécessaire à la réalisation de deux étapes essentielles à la suite de l'infection virale : la réplication et la transcription des trois segments qui composent l'ARN_v.

La réplication de l'ARN_v débute par la production d'un génome viral complémentaire (ARN_c) en prenant l'ARN_v comme modèle, suivie de la genèse de nouveaux ARN_v en se basant sur les ARN_c. La protéine L, et en particulier son domaine ARN polymérase ARN-dépendante (RdRp), joue un rôle central dans cette réaction. Au cours de l'infection, la protéine L permet également d'initier la production d'ARN messagers (ARN_m) codant pour les quatre protéines virales. Pour cela, elle « vole » à la cellule infectée un segment d'ARN coiffé en 5' via son domaine de liaison de la coiffe (CBD). Cet ARN coiffé est ensuite clivé par un domaine endonucléase (ENDO) également présent au sein de la protéine L et

permettant *in fine* d'initier la synthèse de l'ARNm viral au niveau de son domaine RdRp : c'est la transcription.

Actuellement, il n'existe aucun traitement pour combattre les infections dues à ces virus. Il est donc crucial d'analyser les étapes essentielles de leur cycle viral décrites ci-dessus afin de comprendre leur mécanisme de fonctionnement et de potentiellement permettre le développement de nouveaux traitements antiviraux.

L'objectif global de cette thèse était donc de tirer profit des derniers développements en cryo-microscopie électronique (cryo-ME) afin de caractériser au niveau structural puis au niveau fonctionnel, deux acteurs majeurs de la réplication et de la transcription chez les Bunyavirus en utilisant comme modèle d'étude les NPs et les protéines L du HTNV et du LACV. Au cours de cette thèse, j'ai tout d'abord déterminé la structure à haute résolution de la nucléocapside recombinante du HTNV (HTNV-NC) afin de caractériser les différentes interactions entre les NPs et de comprendre le mécanisme d'encapsidation de l'ARNv au sein des HTNV-NCs. Puis, j'ai déterminé les structures à résolution quasi-atomique de la protéine L du HTNV (HTNV-L) et du LACV (LACV-L) en complexe avec les régions promotrices 5' et 3' de l'ARNv. Enfin, j'ai caractérisé fonctionnellement la LACV-L via la mise en place de tests d'activités réalisés *in vitro*. Ceci a permis de bloquer la LACV-L à plusieurs étapes clés de la réplication et de la transcription, un prérequis qui m'a permis de déterminer sept structures à haute résolution de la LACV-L en action. La combinaison de ces instantanés structuraux montre avec grande précision les mécanismes moléculaires sous-tendant la réplication et la transcription des Peribunyavirus.

Mots clés : Virus, cryo-microscopie électronique, nucléocapside, polymérase virale, réplication, transcription.

Résumé en anglais

The *Bunyavirales* order encompasses over 500 species of segmented negative stranded RNA viruses (sNSV). Widespread throughout the world and considered as emerging viruses, the infection of humans by Bunyaviruses can lead to serious diseases. In the *Hantaviridae* family, hemorrhagic fevers have been reported after Hantaan virus (HTNV) infection, while in the *Peribunyaviridae* family, La Crosse virus (LACV) infections have resulted in cases of encephalitis in children. In addition, the order *Bunyavirales* is related to other sNSVs such as influenza virus (*Orthomyxoviridae*), an infectious agent responsible for seasonal epidemics.

The genomes of HTNV and LACV consist of three segments that encode four structural proteins: the glycoproteins (Gn and Gc), encoded on the M segment, the nucleoprotein (NP), encoded on the S segment and the polymerase, also known as L protein, encoded on the L segment. During infection, NPs protect the viral genome from damage and against innate immune response of the infected cell by multimerizing around the viral genome. These protein-RNA complexes form helical assemblies called nucleocapsids (NC). Both the 5' and 3' ends of each segment of the viral genome (vRNA) are respectively bound to an L protein. Interactions between L, NP and vRNA form a circularized complex called ribonucleoprotein (RNP). The RNPs present in the virus are the functional units that perform the replication and the transcription of each of the three viral segments.

The replication of the vRNA first consists in the production of the complementary genome (cRNA) using vRNA as template, followed by the generation of new vRNA based on the cRNA template. The L protein, and in particular its RNA-dependent RNA polymerase (RdRp) region, plays an essential role in this reaction. During infection, the L protein also initiates the production of messenger RNAs (mRNA) that encode viral proteins. It first “steals” a 5' capped RNA segment from the infected cell via its cap-binding domain (CBD). This capped RNA is then cleaved by an endonuclease domain (ENDO) also present in the L

protein, which ultimately initiates the synthesis of viral mRNA thanks to the RdRp catalytic domain: this is transcription.

Currently, there is no treatment to counteract Bunyavirus infections. It is therefore crucial to analyze the essential steps of their viral cycle described above in order to understand their mechanism of operation and potentially allow the development of new antiviral treatments.

The overall objective of this thesis was to take advantage of cryo-electron microscopy to structurally and functionally characterize two major actors of replication and transcription in Bunyaviruses, NPs and L proteins, using HTNV and LACV as models. I first determined the high-resolution structure of the recombinant HTNV nucleocapsid (HTNV-NC) to characterize the different interactions between the different NPs and understand the mechanism of encapsidation of the vRNA within the HTNV-NCs. Secondly, I determined the cryo-EM structures of both HTNV and LACV L proteins (respectively called HTNV-L and LACV-L) in complex with the 5' and 3' promoter regions. Finally, I performed the functional characterization of LACV-L via the implementation of *in vitro* activity assays. This enabled to stall LACV-L in several key stages of both replication and transcription, a pre-requisite that permitted the determination of seven high-resolution structures of LACV-L in action. The combination of these structural snapshots shows with unprecedented details the mechanisms underlying Peribunyaviruses replication and transcription.

Key words: Virus, cryo-electron microscopy, nucleocapsid, viral polymerase, replication, transcription.

Virologie et microscopie, à la croisée des chemins

Depuis leur découverte, les virus font partie intégrante de notre société et plus récemment de notre quotidien comme le prouve la pandémie mondiale causée par le SARS-CoV2. Il est donc dans l'intérêt de tous de mieux comprendre leur fonctionnement afin de prévenir et guérir les maladies qu'ils induisent ou de pouvoir les utiliser à des fins thérapeutiques.

Bien qu'omniprésents dans la totalité des écosystèmes sur notre planète et disposant de la capacité d'infecter tout être vivant (bactéries, archées, algues, champignons, plantes et animaux), ils sont cependant restés invisibles aux yeux de l'homme pendant de nombreux siècles et leur découverte ne remonte qu'à une centaine d'années.

L'une des premières mentions de l'existence de ces agents infectieux remonte à la première moitié du XIX^{ème} siècle. Jean Hameau, alors médecin français, présente une étude « prémonitoire » à ses collègues sur le fait que certaines maladies infectieuses pourraient être induites par autre chose que des microorganismes. Balayés en premier lieu par ses confrères, ces travaux amènent pourtant des idées novatrices. Il décrit en détail les quatre phases d'une infection virale et discute de premières notions de classification. Il introduit également les concepts d'anticorps et évoque la possibilité d'utiliser des *virus* atténués comme solution face à l'infection (Hameau, 1895), ou vaccins, dont le premier a été développé par Edward Jenner contre la variole et dont le processus a été généralisé quelques dizaines d'années plus tard par les travaux du célèbre Louis Pasteur.

En 1887, John Buist est le premier à observer des particules virales de la varicelle par microscopie photonique mais sans le savoir, les interprètent comme étant des spores bactériennes à cause de leurs tailles similaires, de l'ordre du dixième de micromètre. Dmitri Ivanovski fait la même erreur en 1892 lorsqu'il parvient à isoler l'agent causal de la maladie de la mosaïque du tabac. C'est finalement entre 1897 et 1898 que les travaux de Friedrich Löffler, Paul Frosch et Martinus Beijerinck ouvrent la voie vers un nouveau domaine d'étude – la virologie

– en soutenant que ces agents infectieux sont « *suffisamment petits pour passer à travers les pores d'un filtre imperméable à la plus petite des bactéries connues, si petits que même le meilleur système optique à immersion rend son identification impossible à l'aide de nos microscopes* ». Ces particules élémentaires sont alors pour la première fois décrites et nommées : *Contagium vivum fluidum*, littéralement *fluide vivant contagieux* (Beijerinck, 1898).

Malgré le développement de nouvelles méthodes de purifications à l'aide de filtrations successives ou l'apparition de l'ultracentrifugation analytique, aucune technique ne permit une caractérisation précise, rapide et fiable des différentes propriétés morphologiques des virus découverts au début du XX^{ème} siècle. Les microscopes photoniques ne disposaient en effet pas d'une résolution suffisante pour les distinguer.

C'est en 1933 que le domaine des microscopies va prendre un tournant majeur avec la conception et la construction du premier microscope électronique par Ernst Ruska (Figure 1). A défaut d'utiliser des photons comme en microscopie optique conventionnelle, ce sont les électrons qui composent le faisceau lumineux. La focalisation et le degré de condensation du faisceau d'électrons sont quant à eux contrôlés grâce à des lentilles électromagnétiques. Six ans plus tard, les chemins des domaines de la microscopie électronique et de la virologie se rejoignent. Les frères Ruska, Kausche et von Borries parviennent à générer les premiers clichés du virus de la variole et de la mosaïque du tabac à l'aide d'un microscope électronique (Figure 1) (Gelderblom and Krüger, 2014).

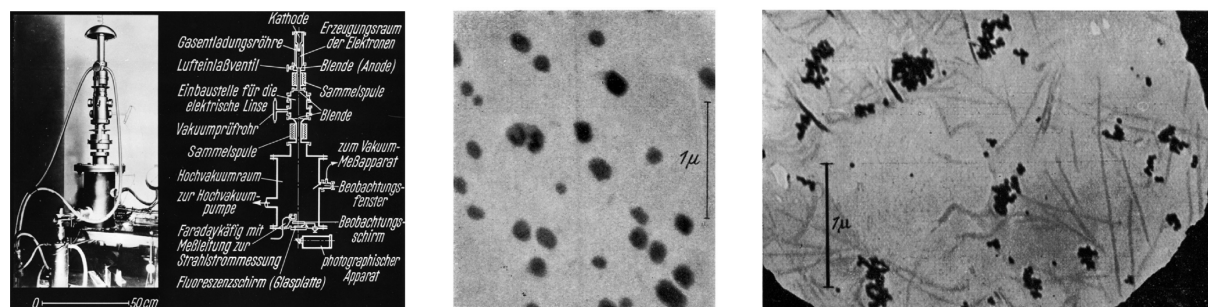


Figure 1. Photographie du premier microscope électronique et premières images de particules virales (*Orthopoxvirus* au centre, virus de la mosaïque du tabac à droite). (Barre d'échelle = 1 μm). Figure adaptée de Gelderblom and Krüger, 2014.

Ce nouvel instrument et ces seules images permettent alors de découvrir un tout nouveau monde et apportent les premières caractérisations structurales venant supporter et compléter des années d'expérimentation préalablement effectuées. Enfin, depuis la deuxième moitié du XX^{ème} siècle jusqu'à nos jours, les améliorations techniques, le développement des diverses méthodes de préparation de l'échantillon, de nouveaux détecteurs et des logiciels de traitement d'images permettent aujourd'hui à la microscopie électronique d'être l'un des fers de lance des domaines de la biologie structurale et de la virologie. Elle permet de mieux comprendre comment ces virus fonctionnent et cela, aussi bien au niveau cellulaire que moléculaire.

Chapitre I : Les Bunyavirus

1. Classification des virus

André Lwoff a été l'un des premiers à défendre l'existence d'un monde viral – ou virome – tout en appelant à la nécessité d'établir une classification pertinente (Lwoff, 1957). Par cette réflexion, il entendait dissocier les notions de microorganisme – qui est soumis à une division cellulaire, est pourvu d'une machinerie enzymatique de production d'énergie et est considéré dans son entité comme agent infectieux – et virus, qui est tributaire de différentes étapes de l'infection précises se déroulant au sein de la cellule hôte, sans division possible ni production d'énergie autonome et dont seul le matériel génétique peut être considéré comme agent infectieux.

C'est en 1971 qu'une classification est instaurée avec la publication des travaux de David Baltimore (Baltimore, 1971). Elle divise le virome en sept groupes principaux basés sur le type d'acide nucléique présent au sein du virus et sur sa nature mono ou bicaténaire (Figure 2).

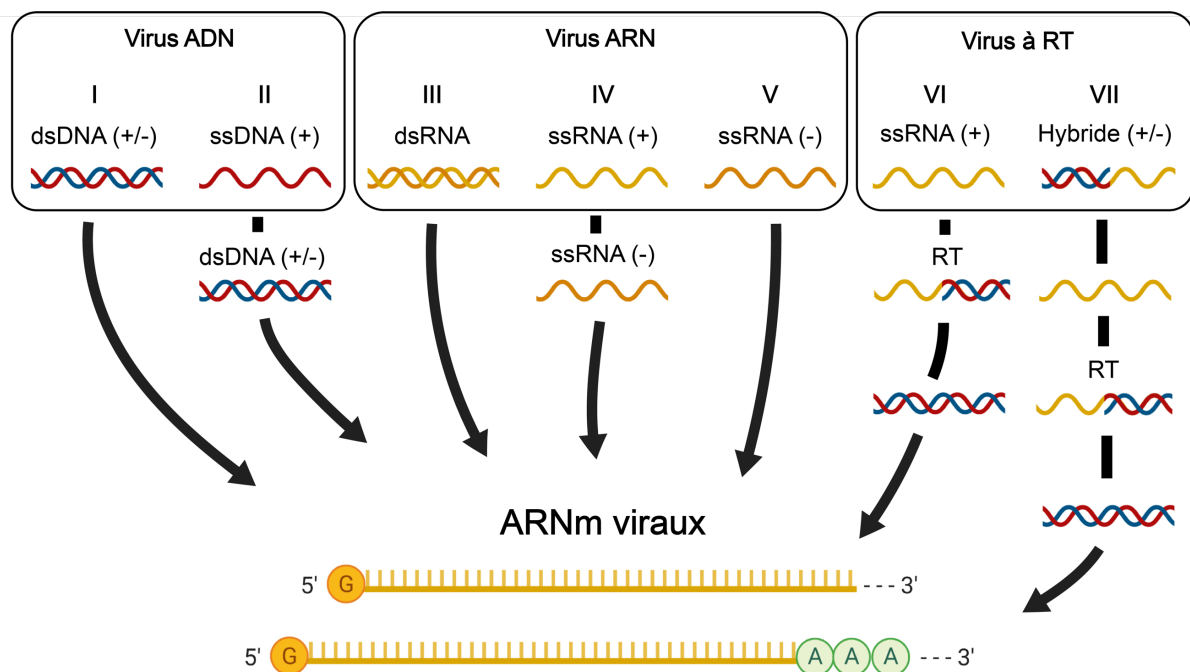


Figure 2. Classification de Baltimore. Les virus dont les principales caractéristiques vont être abordées et développées par la suite appartiennent aux virus à ARN simple brin de polarité négative du groupe V.

Le génome viral peut ainsi être composé d'acide désoxyribonucléique (ADN) ou d'acide ribonucléique (ARN) sous forme de double ou simple brin de topologie

linéaire ou circulaire et comprenant un ou plusieurs segments. La polarité du génome monocaténaire peut être négative (antisens), positive (sens), ou ambisens. Chacune de ces caractéristiques entraîne des répercussions sur les mécanismes de réplication et de transcription, l'expression des gènes, l'assemblage des nouvelles particules virales et sur la localisation de ces événements au sein de la cellule infectée. Certains virus disposent également d'enzymes spécifiques impliquées dans leur réplication que l'on ne retrouve ni dans les cellules eucaryotes, ni dans les procaryotes. C'est en particulier le cas des ARN polymérases ARN-dépendantes (RdRp) qui catalysent la biosynthèse d'un brin d'ARN complémentaire (ARNc) à partir d'un brin d'ARN matrice.

Malgré ces différences, les virus partagent tous le but de répliquer leur génome et d'exprimer leurs gènes en tant qu'ARN messagers (ARNm) afin de garantir la production de nouveaux virions en piratant les différentes machineries cellulaires de l'hôte.

1.1. Les virus à ADN

Les groupes I et II contiennent respectivement les virus à ADN double brin et à ADN simple brin de polarité positive. En raison de la grande stabilité de l'ADN, la taille de leurs génomes varie énormément, allant de 2 kb pour les virus à ADN simple brin à 375 kb pour les virus à ADN double brin et dépassant même le Mb chez la famille des *Mimiviridae*. De par leur variété de structures et de taille, leurs stratégies de réplication sont donc différentes. Les virus à ADN de petite taille comme les *Parvoviridae* (4-6 kb) dépendent presque entièrement de la machinerie de réplication de l'hôte, tandis que les virus à ADN de grande taille comme les *Poxviridae* (130-375 kb) codent pour une grande partie de leur propre machinerie. Ainsi, le génome des virus à ADN est répliqué par des ADN polymérases et transcrit par des ARN polymérases codées par la cellule hôte ou par le virus. En fonction de la famille virale, ces deux mécanismes peuvent être réalisés au sein du noyau et/ou du cytoplasme de la cellule infectée (Payne, 2017).

1.2. Les virus à transcriptase inverse

Les virus des groupes VI (rétrovirus à ARN simple brin) et VII (pararétrovirus à ADN double brin) font appel à une transcriptase inverse (RT) afin d'effectuer leur cycle viral.

La RT du groupe VI possède trois activités séquentielles et complémentaires :

1. Une activité ADN polymérase ARN-dépendante qui permet de synthétiser un brin d'ADN à partir du brin d'ARN simple brin.
2. Une activité ribonucléase H qui permet la dégradation du brin d'ARN initial.
3. Une activité ADN polymérase ADN-dépendante qui permet au virus de créer un génome ADN double brin.

Collectivement, ces activités permettent de convertir l'ARN simple brin du génome initial en ADN double brin, qui peut ensuite être intégré dans le génome de la cellule hôte devenant ainsi un provirus. L'ARN polymérase II de la cellule infectée utilise cet ADN pour la réplication du génome viral et sa transcription en ARNm. L'une des familles de ce groupe, les *Retroviridae*, comprend notamment le virus de l'immunodéficience humaine.

Le génome des virus du groupe VII est circulaire mais pas totalement bicaténaire ce qui complexifie leur cycle viral. A l'aide de l'ARN polymérase II de la cellule hôte, un ARN simple brin de polarité positive est synthétisé à partir du génome initial. La RT réalise ensuite la transcription inverse et synthétise un ADN simple brin de polarité négative à partir de l'ARN simple brin de polarité positive. Un court brin d'ADN de polarité positive est ensuite généré pour reconstituer le génome ADN double brin partiel initial. La famille des *Hepadnaviridae* appartient à ce groupe et contient notamment le virus de l'hépatite B (Fermin, 2018).

1.3. Les virus à ARN

A la différence des virus à ADN et des virus à ARN possédant une RT précédemment décrits, les virus à ARN appartenant aux groupes III, IV et V

disposent d'un arsenal de réplication et de transcription totalement différent et font appel à des RdRp virales.

Le groupe III englobe les virus à ARN double brin. La RdRp virale synthétise un ARN simple brin de polarité positive servant de substrat pour la traduction des protéines virales ainsi que pour la réplication du brin complémentaire de polarité négative. La famille des *Reoviridae* fait partie de ce groupe et inclue différents rotavirus.

Le groupe IV contient les virus à ARN simple brin de polarité positive. De par sa nature, le génome initial permet aux ribosomes de la cellule hôte de synthétiser directement les protéines virales. La RdRp permet la réplication du génome initial en deux étapes via une étape de synthèse intermédiaire d'un brin complémentaire de polarité négative. On y trouve les familles des *Coronaviridae* (SARS-CoV, MERS-CoV, SARS-CoV-2), *Flaviviridae* (Zika, virus de l'hépatite C, dengue) ou *Picornaviridae* (rhinovirus, poliovirus).

Enfin, le groupe V concerne les virus à ARN simple brin de polarité négative. A la différence du groupe IV, la RdRp virale effectue une étape de transcription du génome en ARNm. La réplication quant à elle, passe par une étape intermédiaire de synthèse d'un brin de polarité positive. On y retrouve trois grands ordres :

1. Les *Bunyivirales* contenant notamment les familles des *Peribunyaviridae* (virus de La Crosse, LACV), *Hantaviridae* (virus d'Hantaan, HTNV), *Arenaviridae* (virus de Lassa, LASV).
2. Les *Articulavirales* contenant notamment la famille des *Orthomyxoviridae* (virus influenza).
3. Les *Mononegavirales* dont les familles principales sont les *Paramyxoviridae* (virus de la rougeole), *Filoviridae* (virus d'Ebola, virus de Marburg) et *Rhabdoviridae* (virus de la rage).

Au sein de ce groupe, le génome peut être segmenté ou non (voir paragraphe suivant).

1.3.1. Les virus à ARN simple brin de polarité négative

Comme vu précédemment, les virus à ARN simple brin de polarité négative (NSVs) appartiennent au groupe V de la classification de Baltimore. Ils sont divisés en deux sous-groupes principaux qui se différencient par le degré de segmentation de leur génome. Le premier sous-groupe comprend les virus à ARN simple brin de polarité négative non segmenté (nsNSVs). Ils appartiennent à l'ordre des *Mononegavirales* qui se divise en 5 familles : *Bornaviridae*, *Filoviridae*, *Paramyxoviridae*, *Nyamiviridae* et *Rhabdoviridae*.

Le deuxième groupe concerne les virus à ARN simple brin de polarité négative segmenté (sNSVs). Comme leur dénomination l'indique, leurs génomes se composent de 2 à 8 segments et chacun d'entre eux code pour une ou plusieurs protéines virales. Ils sont divisés en deux ordres principaux : Les *Articulavirales* qui contient la famille des *Orthomyxoviridae* (virus influenza) et les *Bunyavirales* qui contient 12 familles.

La propriété commune des génomes des NSVs consiste en leur organisation hautement structurée sous forme de complexes ribonucléoprotéiques (RNP) dans lesquels l'ARN génomique, lié à la RdRp par ses extrémités 5' et 3', est associé à de multiples monomères de nucléoprotéine (NP). Les NPs protègent intégralement l'ARN en le recouvrant.

Chez les *Mononegavirales*, les gènes viraux sont regroupés sur un segment unique et sont exprimés séquentiellement, tandis que chez les sNSVs, chacun des segments est associé à une RNP distincte, fonctionnellement indépendante, dans le cadre de la transcription et de la réplication du génome viral.

Les NSVs constituent d'importants agents pathogènes pour l'homme provoquant de fréquentes épidémies comme les virus influenza (sNSV, *Orthomyxoviridae*), le virus de la rougeole (nsNSV, *Paramyxoviridae*), les virus des oreillons (nsNSV, *Paramyxoviridae*) ou les virus respiratoires syncytium qui provoquent des bronchiolites (nsNSV, *Pneumoviridae*). D'autres virus peuvent provoquer des infections ou des épidémies plus occasionnelles comme les virus de la rage (nsNSV, *Rhabdoviridae*), Ebola (nsNSV, *Filoviridae*), Nipah (nsNSV,

Paramyxoviridae), LASV (sNSV, *Arenaviridae*), HTNV (sNSV, *Hantaviridae*) et sont fréquemment associés à un taux de mortalité élevé.

Dans l'ensemble, le groupe des NSVs est responsable de maladies présentant une morbidité élevée dans le monde entier, avec des implications sanitaires et économiques importantes et constitue une menace permanente de nouvelles épidémies aux conséquences potentiellement dévastatrices.

2. L'ordre des *Bunyavirales*

Initialement isolé à partir de moustiques du genre *Aedes* en 1943 durant l'étude de la fièvre jaune en Ouganda (Smithburn, Haddow and Mahaffy, 1946), le virus Bunyamwera (BUNV) donnera son nom à la famille des *Bunyaviridae* quelques années plus tard (Murphy, Harrison and Whitfield, 1973). Le BUNV constitue une cause importante de maladie fébrile aiguë chez l'homme et a longtemps été considéré comme le virus prototype de la famille des *Bunyaviridae* jusqu'en 2017 où cette famille est reconsidérée en tant qu'ordre et nommée *Bunyavirales* par le comité international de taxonomie des virus (TaxoProp 2016.030a-vM) (Maes *et al.*, 2018). Regroupant plus de 450 espèces de virus en son sein, cet ordre est le plus abondant du groupe V de la classification de Baltimore suivi par l'ordre des *Articulavirales* qui contient notamment la famille des *Orthomyxoviridae*.

L'ordre des *Bunyavirales* est divisé en 8 familles principales : les *Peribunyaviridae*, les *Hantaviridae*, les *Arenaviridae*, les *Nairoviridae*, les *Fimoviridae*, les *Phasmaviridae*, les *Phenuiviridae* et les *Tospoviridae* (Maes *et al.*, 2018). Chaque famille contient différents genres qui eux-mêmes incluent différentes espèces contenant chacune un ou plusieurs virus en leur sein. Bien qu'appartenant au même ordre, des diversités au niveau biologique (hôte, mode de transmission, maladies) et au niveau moléculaire (encodage du génome, stratégies de réplication) peuvent exister entre ces différentes familles (Figure 3).

Au sein de l'ordre des *Bunyavirales* la majorité des virus est considérée comme étant des arbovirus car les vecteurs de transmission sont des arthropodes comme les moustiques (*Peribunyaviridae*, *Phenuiviridae*), les tiques (*Nairoviridae*, *Phenuiviridae*), les acariens (*Fimoviridae*), les thrips (*Tospoviridae*), les cicadelles (*Phenuiviridae*) ou les mouches (*Phasmaviridae*) (Horne and Vanlandingham, 2014; Herath *et al.*, 2020). A la différence des familles d'arbovirus citées ci-dessus, les vecteurs de transmission de la famille des *Hantaviridae* sont des rongeurs, infectés de manière persistante.

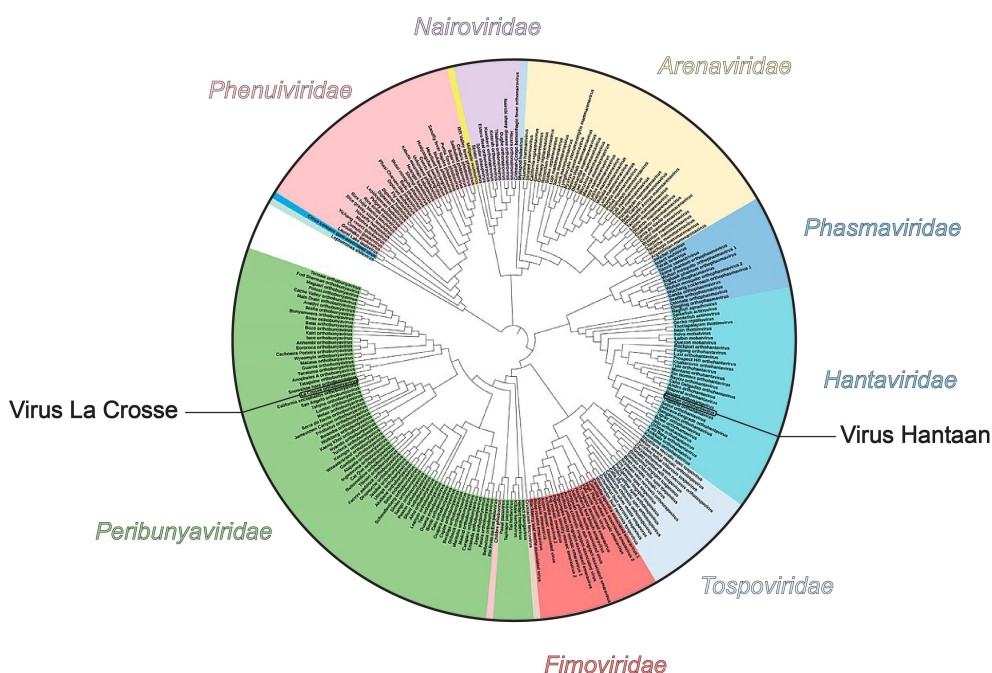


Figure 3. Les 8 familles principales de l'ordre des *Bunyvirales*. Les virus La Crosse et Hantaan, sujets de cette thèse, sont mis en évidence. Figure adaptée de Herath *et al.*, 2020.

2.1. Des virus émergents

Les Bunyavirus utilisent des vecteurs de transmission variés et induisent diverses maladies chez l'homme, chez les animaux domestiques ou d'élevage et chez les plantes. Leur distribution géographique est mondiale mais certains virus sont présents de manière plus importante dans certains pays ou continents pouvant y induire des problèmes de santé publique et économique majeurs. Sont considérés comme émergents les virus associés à certains critères tels que :

- L'évolution récente du virus, *i.e* leur taux de mutation ou les recombinaisons possibles des différents segments de leur génome.
- Les changements de pratiques dans l'agriculture et l'élevage animal qui entraînent une déforestation de plus en plus importante ou de nouveaux projets d'irrigation qui amènent inévitablement à une urbanisation extrême et un contact accru avec les vecteurs de transmission.
- Les changements climatiques actuels dus aux activités humaines et les catastrophes naturelles amènent à bouleverser certains écosystèmes influençant drastiquement l'émergence ou la réémergence de certains de ces virus (Elliott, 1997; Sutherland, Anyamba and LaBeaud, 2013).

Chez l'Homme, quatre familles de Bunyavirus (*Peribunyaviridae*, *Nairoviridae*, *Hantaviridae* et *Phenuiviridae*) sont susceptibles de causer des maladies importantes :

- Le virus de La Crosse (LACV) (*Peribunyaviridae*), présent en Amérique du Nord, peut entraîner des encéphalites qui dans 1% des cas peuvent être mortelles chez les jeunes enfants (Haddow and Odoi, 2009; McJunkin *et al.*, 2009).
- Le virus de la fièvre hémorragique du Congo de Crimée (CCHFV) (*Nairoviridae*) présent en Europe de l'Est, en Afrique et en Asie peut entraîner des hémorragies importantes induisant un taux de mortalité allant jusqu'à 30% (Nasirian, 2020).
- Le virus Hantaan (HTNV), présent en Asie, ou le virus Sin nombre, présent en Amérique (*Hantaviridae*) induisent également des fièvres hémorragiques avec syndrome rénal atteignant des taux de mortalité respectif de 5 à 15% et 50% (Jonsson, Figueiredo and Vapalahti, 2010).
- Le virus de la fièvre de la vallée du Rift (RVFV) présent en Afrique (*Phenuiviridae*) peut également induire encéphalites, fièvres hémorragiques et rétinites chez l'Homme tout comme le virus de la fièvre sévère avec thrombocytopénie (STFSV) récemment découvert en 2009 en Chine (Yu *et al.*, 2011).

Pour le monde animal et en particulier dans le cas de l'élevage animal, on notera le virus de Schmollenberg (*Peribunyaviridae*) apparu en Allemagne et aux Pays-Bas en 2011 qui provoque fièvre, diminution de la production de lait, diarrhées et augmentation de la mortalité chez les bovins (Hoffmann *et al.*, 2012). En ce qui concerne l'agriculture, les Tospovirus (*Tospoviridae*) infectent cultures légumières, aromatiques ou ornementales et sont responsables de pertes économiques considérables (Rotenberg *et al.*, 2015). Les Tenuivirus (*Phenuiviridae*) entraînent quant à eux la destruction de cultures de maïs et de riz (Falk and Tsai, 1998).

De leur découverte à nos jours, le nombre de Bunyavirus recensés n'a cessé d'augmenter et l'émergence de nouveaux Bunyavirus n'est pas à exclure comme en

atteste la découverte récente du virus Cristoli en France (Rodriguez *et al.*, 2020). Certains d'entre eux pourtant déjà connus, occasionnent toujours des épidémies dans certaines régions du monde et des pertes économiques colossales. De plus, le mode de fonctionnement et de production de notre société tendant vers l'urbanisation extrême et vers une agriculture de plus en plus intensive entrainera inévitablement une réémergence d'anciens et/ou l'émergence de nouveaux Bunyavirus et pour lesquels aucun traitement n'est actuellement disponible.

2.2. Organisation du virus et de son génome

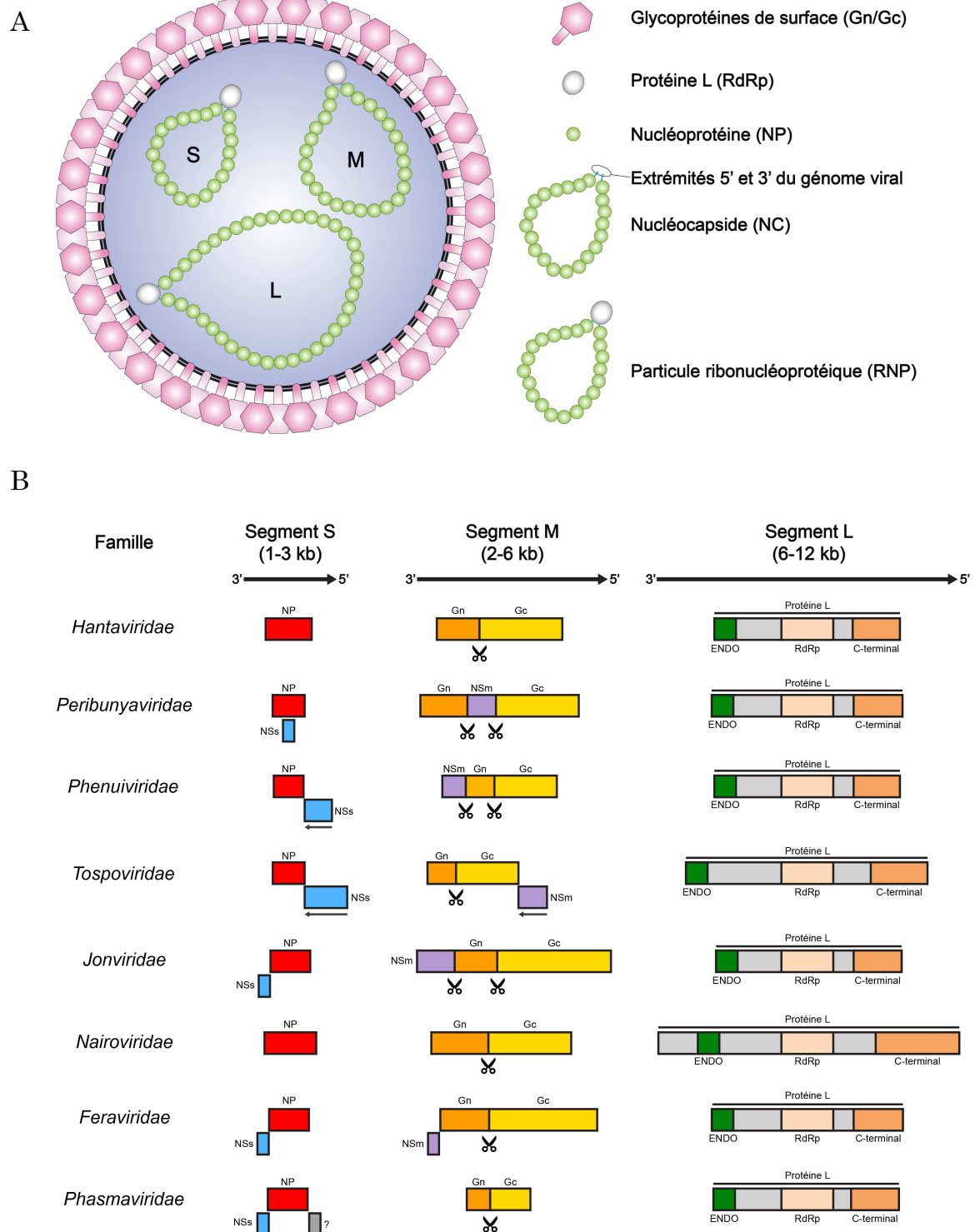


Figure 4. Schéma général représentant un Bunyavirus et organisation de son génome viral (chez les familles de l'ordre des *Bunyvirales* disposant de trois segments d'ARN de polarité négative). **A**, Les glycoprotéines sont ancrées dans la bicouche lipidiques (double traits noir). **B**, Les différentes stratégies de production des protéines sont indiquées. Chacun des domaines principaux des protéines L (ENDO, RdRp, C-terminal) sont indiqués. Les familles des *Hantaviridae* et des *Peribunyaviridae* sont les principaux sujets d'études de cette thèse.

Les Bunyavirus disposent d'un génome à ARN simple brin de polarité négative. A la différence de l'ordre des *Mononegavirales*, le génome de l'ordre des *Bunyavirales* est dit segmenté, car il est composé de 2 à 8 segments. Cet ordre s'apparente à celui des *Articulavirales* qui inclus la famille des *Orthomyxoviridae*. La taille totale du génome des Bunyavirus varie entre 10 et 23 kb. Dans la majorité des cas, le génome des Bunyavirus se répartit en trois segments de tailles différentes (S pour *small*, M pour *medium* et L pour *large*) (Figure 4). Les familles des *Arenaviridae* et des *Fimoviridae* font exception à cette règle avec respectivement deux et quatre segments (Mielke-Ehret and Mühlbach, 2012). Pour la suite de la description de la composition des virions, seules les familles comprenant trois segments seront présentées.

Les segments S, M et L codent pour quatre protéines structurales conservées qui composent les virus et les nouveaux virions produits : la NP sur le segment S, les glycoprotéines (Gn, Gc) sur le segment M et la protéine L (RdRp) sur le segment L (Figure 4). Deux protéines non-structurales NSs est NSm, encodées sur les segments respectifs S et M, sont produites au cours de l'infection virale mais absentes au sein des virions. L'organisation générale du génome viral est hautement conservée entre les différentes familles appartenant à l'ordre des *Bunyavirales*.

2.2.1. Le segment S

Le segment S code pour la NP et pour une protéine non structurale NSs, cette dernière étant uniquement présente chez les *Peribunyaviridae*, les *Phenuiviridae*, les *Tospoviridae*, les *Feraviridae* et les *Phasmaviridae* (Bishop *et al.*, 1982; R.M. Elliott, 1989).

La NP est la protéine virale la plus abondante au sein du virus car de multiples copies recouvrent chaque segment du génome viral pour former un complexe ARN-protéine nommée nucléocapside (NC) (Figure 4). Les NPs jouent un rôle critique dans la réplication du virus en recouvrant les ARN viraux de polarité négative (ARN_v) et positive (ARN_c) nouvellement synthétisés par la protéine L, *a contrario* des ARN_m viraux, produits de transcription, qui ne le sont pas (Sun *et al.*, 2018). Cette encapsidation de l'ARN viral permet de le protéger des

dégradations possibles *in vivo* et de limiter la réponse immunitaire induite par la présence d'ARN simple brin mono ou tri-phosphorylés en 5' dans les cellules infectées (Randall and Goodbourn, 2008).

La protéine NSs joue un rôle prépondérant dans le contrôle de la réponse du système immunitaire innée. En effet, l'entrée du virus dans une cellule entraîne différentes réponses et différentes cascades biochimiques complexes. L'infection virale active notamment le facteur-3 de régulation d'interféron (IRF-3) menant à l'initiation de la transcription des interférons (IFN) α et β . La protéine kinase R entraîne la phosphorylation et l'activation de la sous-unité α du facteur d'initiation eucaryote 2 pouvant mener à l'autophagie, l'apoptose et/ou l'arrêt du cycle cellulaire afin de limiter la propagation du virus. Pour se prémunir de ces différentes réponses à l'infection virale, NSs inhibe le IRF-3 en diminuant la transcription des IFN- β et empêche la phosphorylation de la sérine 2 du domaine C-terminal de l'ARN polymérase II, responsable de l'élongation des ARNm cellulaires et de leur terminaison en 3' (Randall and Goodbourn, 2008; Eifan *et al.*, 2013).

A la différence des Peribunyavirus, des Phenuivirus et des Tosopovirus, les virus appartenant aux familles *Hantaviridae* et *Nairoviridae* sont dépourvus du gène codant pour la NSs. Afin de prévenir l'activation des récepteurs de reconnaissance de motifs – tel que l'hélicase RIG-I – spécialisés dans la détection des ADNs ou ARNs (simple ou double brin) du non soi (Hornung *et al.*, 2006; Pichlmair *et al.*, 2006), une stratégie potentielle d'évasion du virus résiderait dans la modification du 5' tri-phosphorylés en 5' mono-phosphate (Garcin *et al.*, 1995) (Yoneyama and Fujita, 2008).

2.2.2. Le segment M

Le segment M d'environ 4,5 kb code pour deux glycoprotéines de surface Gn et Gc. Chez certaines familles ce segment contient également l'information pour produire une protéine non structurale NSm.

Les gènes *Gn* et *Gc* appartiennent au même cadre de lecture et codent pour deux protéines synthétisées en une protéine précurseur clivée entre Gn et Gc par des protéases cellulaires lors de l'infection. Gn et Gc sont ancrées dans la bicouche

lipidique virale et disposent d'un domaine protubérant externe d'une longueur comprise entre 5 et 10 nm (Figure 4) (Lees, Pringle and Elliott, 1986; Serris *et al.*, 2020). Les glycoprotéines insérées dans l'enveloppe virale des Bunyavirus leur permettent ainsi de reconnaître et de se lier à différents récepteurs spécifiques cellulaires et induisent la pénétration du virus dans la cellule hôte par fusion membranaire (Welsch, Müller and Kräusslich, 2007; Cifuentes-Muñoz, Salazar-Quiroz and Tischler, 2014; Albornoz *et al.*, 2016).

Peu de choses sont connues au regard des fonctions exactes de la protéine non structurale NSm mais celles-ci semblent spécifiques à chaque famille. Chez les *Peribunyaviridae*, il a été démontré que la partie N-terminale de NSm est requise pour l'assemblage de nouvelles particules virales. Chez les *Phenuiviridae*, elle a un effet anti-apoptotique dans les cellules de mammifères et son absence entraîne une atténuation de la virulence. Enfin chez les *Tospoviridae*, NSm est considérée comme une protéine de mouvement permettant aux virus d'infecter d'autres cellules végétales en traversant la paroi cellulaire (Eifan *et al.*, 2013).

2.2.3. Le segment L

Le segment L code pour la polymérase, aussi appelée protéine L. Cette protéine dispose d'un cœur catalytique d'RdRp (R. M. Elliott, 1989). La protéine L assure les deux mécanismes fondamentaux du virus : (i) la réplication *de novo* du génome passant par une étape intermédiaire de production d'ARNc de polarité positive et (ii) la transcription de chacun des segments d'ARNv en ARNm. Ces ARNm sont utilisés ensuite par la machinerie de traduction de la cellule hôte pour produire de nouvelles protéines virales.

A la différence des virus influenza dont la polymérase est synthétisée à partir de trois segments d'ARNv codant pour les sous-unités : PA (*Polymerase Acidic protein*), PB1 (*Polymerase Basic protein 1*) et PB2 (*Polymerase Basic protein 2*), les protéines L des Bunyavirus sont traduites à partir d'un seul ARNm codant pour une protéine d'un poids moléculaire variant de 250 (*Hantaviridae*, *Peribunyaviridae* et *Phenuiviridae*) à 450 kDa (*Nairoviridae*) soit l'équivalent d'environ 2000 à 4000 acides aminés.

Les alignements multiples de séquence des protéines L démontrent une conservation systématique de différents motifs catalytiques responsables de la synthèse d'ARN et indispensable à l'activité des protéines L (Poch *et al.*, 1989, 1990; Jin and Elliott, 1992; Müller *et al.*, 1994). Ces motifs sont localisés au sein du cœur catalytique (RdRp) des polymérases virales (Figure 4).

Chez les sNSVs, des domaines catalytiques additionnels sont également présents aux extrémités N- et C-terminales des protéines L et sont nécessaires à la transcription des ARN_v en ARN_m.

2.3. Les NTRs du génome viral et leur implication dans la synthèse d'ARN

Les régions 5' et 3' du génome viral sont non traduites (NTRs) et leurs extrémités sont directement liées à un exemplaire de la protéine L pour former un complexe circularisé appelé RNP contenant des copies de l'ARN_c (cRNP) ou de l'ARN_v (vRNP) (Figure 4). La longueur de ces NTRs varie selon les familles de virus et les segments associés, variant de 100 à 175 nucléotides pour l'extrémité 5' et de 50 à 85 nucléotides pour l'extrémité 3'. Pour chacune des familles de l'ordre des *Bunyavirales*, les extrémités 5' et 3' NTRs de chacun des segments ont une complémentarité totale sur les huit premiers nucléotides et la séquence de ces extrémités est conservée entre chaque segment d'une même famille virale. La complémentarité des extrémités 5' et 3', également appelée régions promotrices, peut varier, allant de 9 nucléotides chez les *Nairoviridae* jusqu'à 19 nucléotides chez les *Arenaviridae* (Figure 5) (Ferron *et al.*, 2017).

<i>Peribunyaviridae</i>	3' UCA UCA CAU GA 5' AGU AGU GUG CU	<i>Tospoviridae</i>	3' UCU CGU UA 5' AGA GCA AUC
<i>Hantaviridae</i>	3' AUC AUC AUC UG 5' UAG UAG UA	<i>Phenuiviridae</i>	3' UGU GUU UC 5' ACA CAA AG
<i>Arenaviridae</i>	3' GCG UGU CAC CUA GGA UCC G 5' GCGC ACA GUG GAU CCU AGG C	<i>Nairoviridae</i>	3' AGA GUU UCU 5' UCU CAA AGA

Figure 5. Les séquences conservées des extrémités 3' (en rouge) et 5' (en bleu) de l'ARN_v chez différentes familles virales de l'ordre des *Bunyavirales*. Les bases non complémentaires sont encadrées.

Chez le BUNV, prototype de la famille des *Peribunyaviridae*, la complémentarité des régions promotrices 5'/3' est indispensable à la synthèse d'ARN par le virus, toute mutation entraînant une diminution drastique de l'activité de répllication et de transcription laissant à supposer un fonctionnement identique chez les autres familles de l'ordre des *Bunyavirales* (Kohl *et al.*, 2004; Barr and Wertz, 2005).

La complémentarité du génome viral s'étend au-delà des régions promotrices 5'/3' mais à des degrés divers pour chacun des segments. Elle permet notamment de réguler la transcription et/ou la répllication des gènes viraux dans l'ordre suivant : M > L > S. Cet ordre est conservé dans la majorité des différentes familles de l'ordre des *Bunyavirales* et a également été observé *in cellulo* chez le LACV (Rossier, Raju and Kolakofsky, 1988; Barr *et al.*, 2003). Les NTRs seraient également responsables de l'encapsidation des nouvelles copies du génome produites pendant la répllication (Mohl and Barr, 2009).

Les NTRs 5' et 3' ont donc des fonctions importantes et variées dans la régulation de l'activité de répllication et de transcription des différents segments du génome viral. Cette régulation va dépendre du degré de complémentarité de séquence des régions promotrices 5'/3' mais peut également avoir lieu via des motifs de reconnaissance de nucléotides spécifiques.

2.4. Morphologie

Les Bunyavirus sont des virus enveloppés et sphériques d'un diamètre moyen allant de 70 à 140 nm. La visualisation de virions dans des sections fines par microscopie électronique (ME) permet de distinguer chacune des quatre protéines structurales encodées sur les trois segments d'ARN de polarité négative S, M et L. Chaque famille de Bunyavirus présente des caractéristiques morphologiques différentes au regard du diamètre des virions, de l'organisation des glycoprotéines à leur surface ou de l'aspect des RNPs (Figure 6).

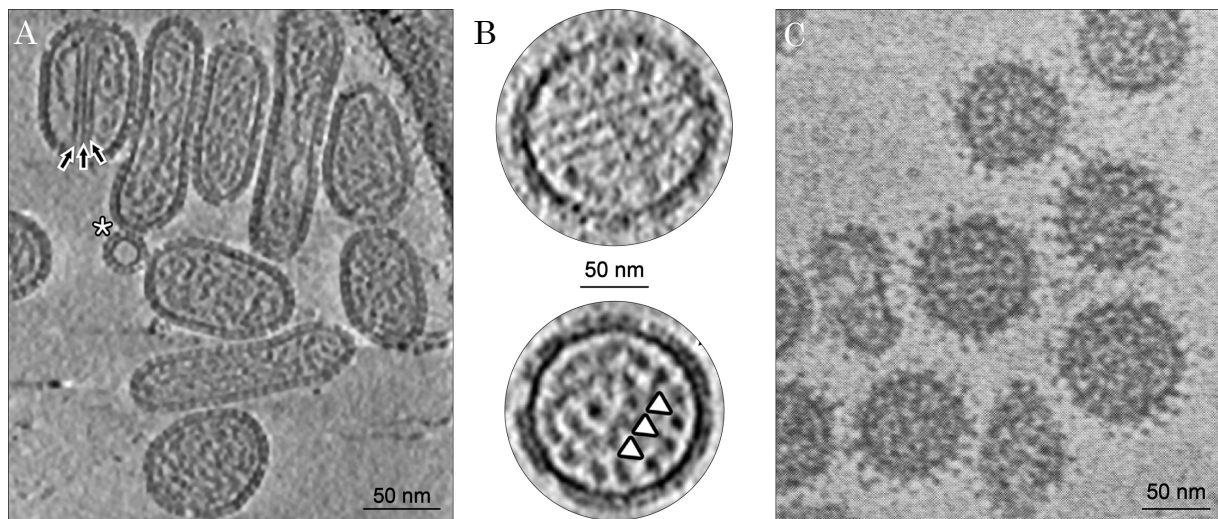


Figure 6. Morphologie des virus Tula, Hantaan et La Crosse observés par cryo-ME. **A**, Particules virales du virus Tula, section centrale d'un tomogramme. Les RNPs rigides sont indiquées par des flèches et la particule virale dépourvue de RNPs est indiquée par un astérisque. Figure adaptée à partir de Huiskonen *et al.*, 2010. **B**, Particules virales du virus Hantaan, section centrale d'un tomogramme. Les RNPs rigides sont visibles et indiquées par des triangles lorsqu'elles sont vues de face. Figure adaptée à partir de Battisti *et al.*, 2011. **C**, Micrographie de particules virales du virus La Crosse obtenue par cryo-ME. Les RNPs sont flexibles et ne présentent pas d'arrangement spécifique. Figure adaptée à partir de Talmon *et al.*, 1987.

Chez les *Hantaviridae* et notamment les virus Hantaan et Tula (TULV), les RNPs ont une forme filamenteuse et rigide (Huiskonen *et al.*, 2010; Battisti *et al.*, 2011). Elles ont un diamètre d'environ 10 nm et sont arrangées de manière parallèle dans les virions. En revanche, chez les *Peribunyaviridae* (LACV) (Talmon *et al.*, 1987) et les *Phenuiviridae* (RVFV, virus de Uukuniemi) (Överby *et al.*, 2008), les RNPs sont beaucoup plus flexibles (Figure 6).

De même, des diversités morphologiques existent entre les différentes familles virales concernant l'arrangement des glycoprotéines à la surface du virus.

Chez les *Phenuiviridae*, les hétéro-dimères Gn/Gc forment un assemblage symétrique icosaédrique. C'est l'unique exemple de ce type trouvé jusqu'à présent au sein de l'ordre des *Bunyvirales* (Freiberg *et al.*, 2008; Huiskonen *et al.*, 2009). Au contraire, la surface des *Hantaviridae* et des *Nairoviridae* est couverte de patches ordonnés d'hétéro-dimères suivant une symétrie C4 (Huiskonen *et al.*, 2010; Battisti *et al.*, 2011; Punch *et al.*, 2018; Serris *et al.*, 2020). Enfin, chez les *Peribunyaviridae* et les *Arenaviridae*, les glycoprotéines de surface forment un assemblage trimérique (Bowden *et al.*, 2013; Li *et al.*, 2016).

Comme évoqué précédemment, la morphologie au sein des différentes familles de Bunyavirus est variée, que ça soit au niveau de l'organisation des RNPs dans le virus, de l'aspect des NCs ou de la disposition des glycoprotéines de surface. Malgré ces différences, les Bunyavirus partagent tous des caractéristiques similaires au niveau de leur cycle viral.

2.5. Le cycle viral

La reconnaissance et l'attachement du virus à une cellule hôte sont les premières étapes de l'infection virale. Les glycoprotéines Gn/Gc localisées à la surface du virus interagissent avec des récepteurs spécifiques présents sur la membrane cellulaire cible. Des interactions de faibles affinités peuvent également avoir lieu avec des lipides et des glycanes. Les différents modes de transmission, *i.e* par morsures d'arthropodes ou l'inhalation d'aérosols contaminés, impliquent également l'utilisation de récepteurs différents. Chez les arbovirus de la famille des *Peribunyaviridae*, *Phenuiviridae* et *Nairoviridae*, les récepteurs DC-SIGN hautement exprimés à la surface des cellules dermiques dendritiques sont leur point d'ancrage principal (Lozach *et al.*, 2011; Hofmann *et al.*, 2013; Suda *et al.*, 2016). Chez les *Hantaviridae*, ce sont les intégrines $\beta 1/2/3$, présentes à la surface des cellules endothéliales qui constituent les récepteurs principaux (Figure 7) (Gavrilovskaya *et al.*, 1998, 1999).

La seconde étape du cycle de réplication des virus est l'internalisation de ces derniers dans la cellule sur laquelle ils se sont fixés. Pour la majorité des Bunyavirus (*Peribunyaviridae*, *Hantaviridae* et *Nairoviridae*), ce mécanisme d'endocytose est clathrine dépendant à l'exception des *Phenuiviridae* dont

l'endocytose est cavéoline dépendante (Albornoz *et al.*, 2016). L'acidification interne des endosomes précoces (pH 6) à tardifs (pH 5.5-5) entraîne alors une modification de la conformation des glycoprotéines de surface afin de faciliter la fusion des membranes virales et cellulaires. Cette fusion permet le relargage des unités fonctionnelles du virus – les RNPs – dans le cytoplasme de la cellule infectée.

En raison de la polarité négative du génome des Bunyavirus, la première étape effectuée par les RNPs relarguées dans le cytoplasme est la transcription. Celle-ci va permettre la production d'ARNm codant pour de nouvelles protéines virales qui seront par la suite nécessaires à la réplication du génome et à la production des virus néosynthétisés. Chez les Bunyavirus, la synthèse des ARNm viraux est directement couplée à la traduction afin de produire des ARNm fonctionnels suggérant une coordination importante entre la machinerie virale et la machinerie cellulaire (Figure 7) (Raju and Kolakofsky, 1986; Bellocq and Kolakofsky, 1987; Ikegami *et al.*, 2005; Barr, 2007). La terminaison de la transcription est quant à elle médiée par des motifs de séquences conservées présents en 5' NTR du génome sans l'ajout de queue poly-A (Collett, 1986; Bouloy *et al.*, 1990; Hutchinson, Peters and Nichol, 1996), à la différence des ARNm produits par les Orthomyxovirus qui sont polyadénylés (Poon *et al.*, 1999).

Une fois la production de protéines virales initiée par la transcription de l'ARNv en ARNm elle-même couplée à la traduction par la machinerie cellulaire, la RdRp effectue une étape intermédiaire de synthèse *de novo* de nouvelles copies de l'ARNc : c'est la réplication. Ces ARNc néosynthétisés sont encapsidés par les NPs et utilisés par la protéine L comme modèle afin de produire de nouvelles copies de l'ARNv qui seront intégrés dans de nouveaux virions (Figure 7).

Dans la cellule, les produits de transcription et de réplication se concentrent au niveau des membranes de l'appareil de Golgi et forment des usines virales tubulaires stabilisées par le cytosquelette. Celles-ci permettent de concentrer et centraliser les composants viraux, *i.e* RNPs, glycoprotéines de surface etc..., au niveau des différents organites cellulaires nécessaires à la formation de nouveaux virions : (i) le réticulum endoplasmique rugueux qui intervient dans la biosynthèse

des protéines avec la présence importante de ribosomes et (ii) l'appareil de Golgi qui permet notamment les modifications post-traductionnelles des glycoprotéines de surface (Fontana *et al.*, 2008).

L'empaquetage des RNPs dans les nouveaux virions survient au niveau de l'appareil de Golgi mais le mécanisme permettant de contrôler la ségrégation spécifique des trois segments S, M et L reste encore inconnu. Il a été montré que les extensions cytoplasmiques des glycoprotéines de surface interagissent avec les NPs chez les *Hantaviridae* (Hepojoki *et al.*, 2010) ou que les régions NTRs de chaque segment` peuvent coordonner leur regroupement chez les *Phenuiviridae* (Terasaki *et al.*, 2011), ce dernier mécanisme ayant déjà été observé chez les virus influenza (Zheng and Tao, 2013).

Finalement, les nouveaux virions sont sécrétés dans des vésicules et migrent jusqu'à la membrane plasmique où ils bourgeonnent à l'extérieur de la cellule pour poursuivre l'infection (Figure 7).

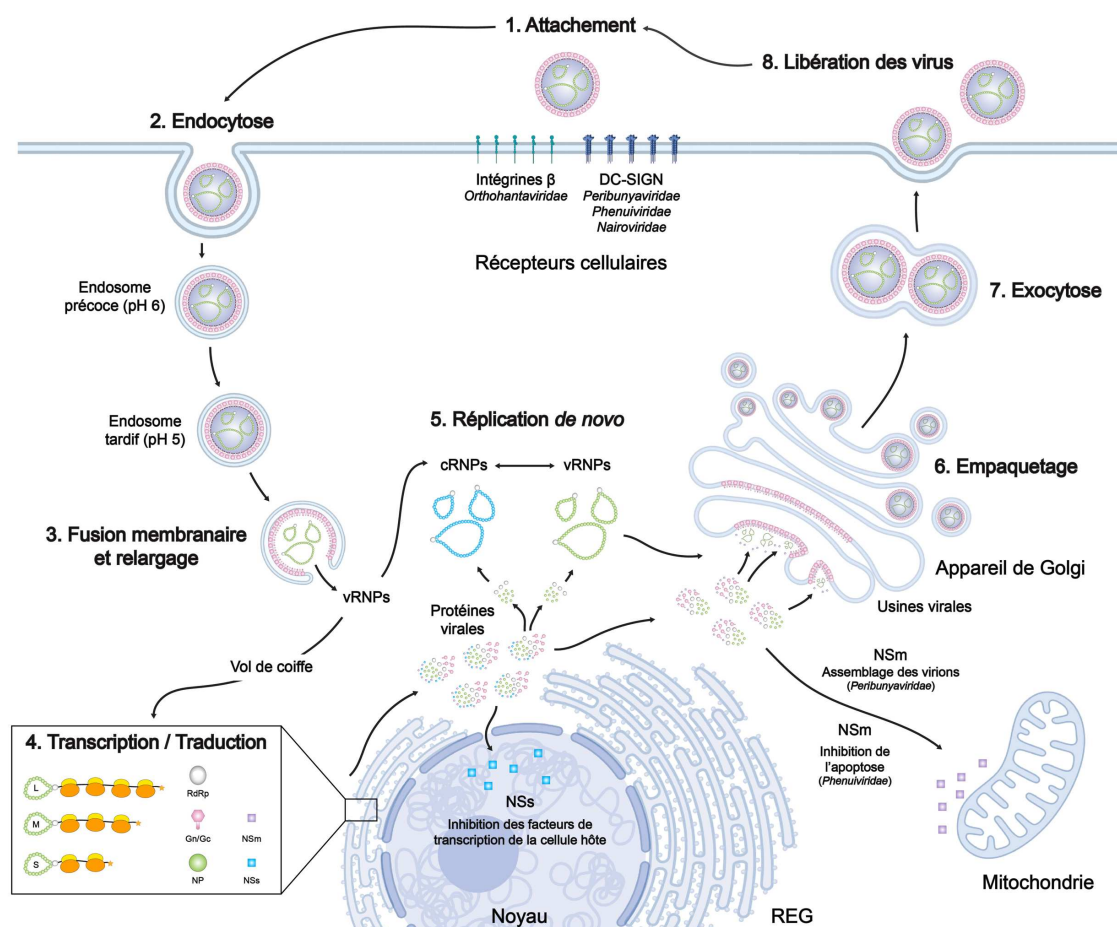


Figure 7. Cycle viral général des Bunyavirus. Chacune des étapes est numérotée de manière chronologique en commençant par l'attachement (1) jusqu'à la libération des nouveaux virus (8).

3. Les ARN polymérases ARN-dépendantes virales des sNSVs

3.1. Caractéristiques et fonctions principales

Le premier mécanisme, commun à toutes les protéines L, permet la sélection et l'incorporation d'un nucléoside triphosphate (NTP) supplémentaire, et cela de manière séquentielle, au brin nouvellement synthétisé en coordonnant une attaque nucléophile à l'aide de deux ions métalliques. C'est le cœur catalytique (RdRp) qui en est responsable. Il permet la réplication du génome viral en passant par une étape intermédiaire de production d'un ARNc de polarité positive.

Le deuxième mécanisme, lié à la transcription, permet la production d'ARNm codant pour les protéines virales. Chez les sNSVs (ordre des *Articulavirales*, famille des *Orthomyxoviridae* (virus influenza), ordre des *Bunyavirales*), les protéines L disposent de domaines catalytiques additionnels : le domaine de liaison de la coiffe (CBD) et l'endonucléase (ENDO). Ces domaines permettent respectivement de voler un ARNm cellulaire en fixant sa coiffe puis en la clivant afin de l'utiliser comme amorce pour produire des ARNm viraux. L'élongation du produit de transcription est quant à elle réalisée au niveau de la RdRp et fait appel au premier mécanisme décrit ci-dessus.

3.2. Le cœur catalytique des RdRps

Le cœur catalytique est la partie la plus conservée des protéines L. D'une longueur d'environ 400 acides aminés, cette région dispose d'une cavité centrale connectée au solvant par 4 tunnels. Ceux-ci permettent l'entrée des NTPs et des cations divalents vers le site actif, l'entrée et la sortie de l'ARN modèle ainsi que du produit. Le cœur catalytique présente une forme de main droite refermée d'où découle l'appellation des différents domaines qui le compose : les doigts (*fingers* et *fangertips*), la paume (*palm*) et le pouce (*thumb*). Ces domaines présentent 6 motifs d'acides aminés conservés et sont désignés par les lettres A à F. Ils sont directement impliqués dans les multiples étapes du cycle catalytique et de la réaction de transfert nucléotidique.

3.2.1. Cycle catalytique et réaction de transfert nucléotidiques

Il existe 5 étapes pour effectuer un cycle complet d'incorporation d'un NTP par la RdRp. Ce cycle dépend en grande partie de la coordination de deux ions métalliques communément appelés métal A et métal B, généralement des ions Mg^{2+} , ainsi que d'important changements de conformations des différents motifs qui composent le cœur catalytique de la RdRp.

La première étape consiste en la formation d'un complexe enzymatique entre la polymérase (L) et l'ARN modèle (ARN). Le métal A est alors coordonné par (i) deux acides aspartiques absolument conservés appartenant aux motifs A et C du cœur catalytique et (ii) le 3'-OH du produit (après la réalisation d'un premier cycle d'incorporation). Un NTP dont les phosphates β et γ sont stabilisés par le métal B, complémentaire au nucléotide présent en position +1 de l'ARN modèle (par définition, la position +1 correspond à la position de fixation et d'incorporation du nouveau NTP), se fixe au complexe $L:ARN_n+NTP$ formant $L:ARN_n:NTP$ (Figure 8).

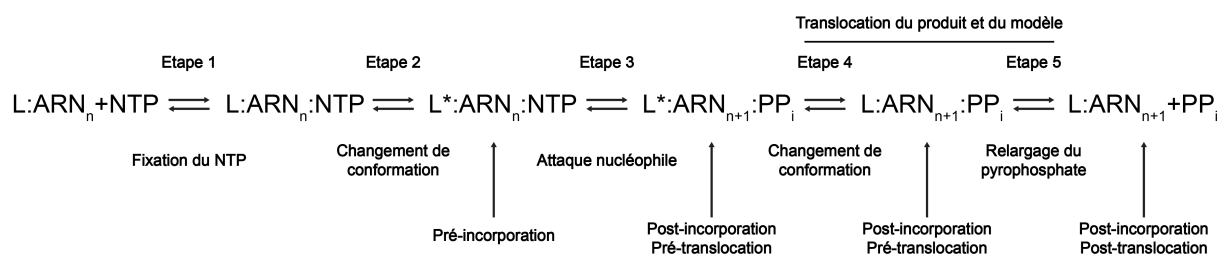


Figure 8. Cycle catalytique et réaction de transfert nucléotidiques. Cinq étapes permettent à la RdRp virale (L) d'incorporer des nucléotides triphosphates (NTP) complémentaire de l'ARN modèle et de relarguer en fin de réaction un pyrophosphate (PPi).

Les métaux A et B permettent alors de positionner de manière optimale le triphosphate du NTP fixé tout en induisant la « fermeture » du site actif dans une position propice à l'attaque nucléophile, c'est l'état de pré-incorporation $L*:ARN_n:NTP$ (Figure 9).

En diminuant le pK_a du 3'-OH du produit, le métal A facilite sa déprotonation en 3'-O \cdot et permet d'effectuer une attaque nucléophile sur le phosphate α du NTP fixé. Celle-ci induit la formation d'un état transitoire lors duquel chacun des métaux A et B forment une liaison pentavalente et est notamment coordonné par des molécules d'eau. Cette conformation entraîne la

formation d'une liaison phosphodiester entre le 3'-O du produit et le phosphate α du NTP incorporé. Le pyrophosphate (PPi) produit du clivage des phosphates β et γ est stabilisé par un acide aminé chargé positivement, K1228 chez la protéine L du LACV (LACV-L), appartenant au motif D : c'est l'état de post-incorporation pré-translocation L*:ARN_{n+1}:PPi.

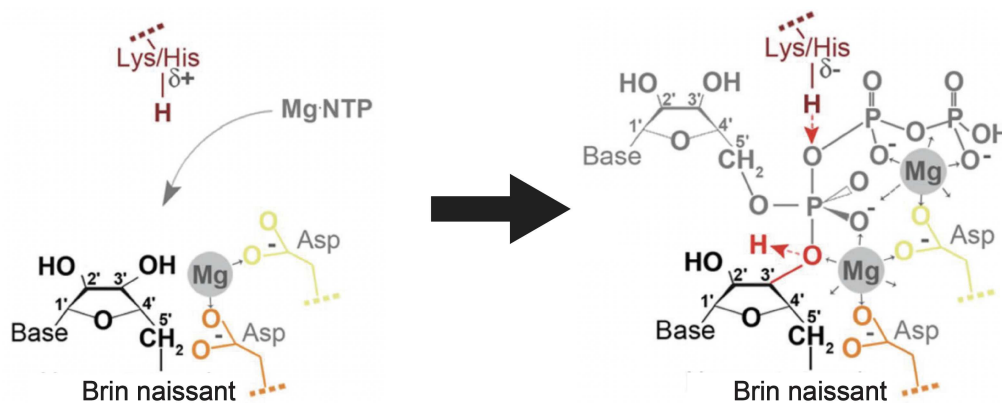


Figure 9. Représentation schématique du site actif des RdRps. Les aspartates (Asp) des motifs A (jaune) et C (orange) lient des ions métalliques divalents (Mg^{2+}), qui sont utilisés pour coordonner la formation d'une nouvelle liaison phosphodiester sur le 3'-OH du brin naissant. Le résidu acide (Lys/His) est positionné près du β -phosphate du NTP entrant afin de protoner le groupe PPi. Figure adaptée de te Velthuis, 2014.

La relaxation (ou « réouverture ») du site actif (L:ARN_{n+1}:PPi) prépare alors le relargage du PPi (L:ARN_{n+1}+PPi) et entraîne la translocation de l'ARN (*i.e* NTP incorporé et ARN modèle) en position -1 dans la cavité du cœur catalytique (L:ARN_{n+1}), c'est l'état de post-incorporation post-translocation. La RdRp est alors prête à effectuer un nouveau cycle catalytique permettant l'incorporation du NTP complémentaire suivant (Figure 8) (Castro *et al.*, 2009; Murakami and Trakselis, 2014; te Velthuis, 2014).

3.2.2. Motifs catalytiques

Le cœur catalytique des RdRps virales est composé de 6 motifs (A à F) conservés en termes de structure et de séquence. Ils entrent en jeu de manière coordonnée afin de réaliser les étapes nécessaires au cycle catalytique permettant à la RdRp de :

1. Sélectionner le NTP complémentaire à l'ARN modèle.
2. Stabiliser le NTP entrant.
3. Coordonner les ions métalliques A et B.

4. Permettre une attaque nucléophile sur le phosphate α du NTP fixé.
5. Stabiliser le PPi avant son relargage pour permettre la translocation du produit et de l'ARN modèle.
6. Réinitialiser le cycle catalytique.

Au niveau structural, les motifs A à E sont localisés dans le domaine de la paume tandis que le motif F fait partie du domaine des doigts. Chez les sNSVs, il existe deux motifs supplémentaires appelées G et H qui font également partie du domaine des doigts (Figure 10).

Le motif A, formé par un brin de feuillet β de la paume, renferme le motif catalytique DX_{2-4}D . Le premier acide aspartique est absolument conservé chez les RdRps des sNSVs. Il coordonne les ions métalliques A et B impliqués dans la réaction de transfert de nucléotides. Le deuxième, de concert avec une asparagine conservée du motif B, joue un rôle crucial dans la différenciation des NTPs/désoxy-NTPs (dNTPs) en formant une liaison hydrogène avec le 2'-OH du NTP entrant. Chez les sNSVs, une lysine conservée remplace le deuxième acide aspartique (Venkataraman, Prasad and Selvarajan, 2018).

Le motif B est composé d'une boucle flexible connectant un brin β du domaine des doigts à une hélice α localisée dans la partie N-terminale du domaine de la paume. La boucle est composée d'une glycine conservée qui stabilise l'ARN modèle. L'hélice α contient une asparagine conservée qui stabilise le 2'-OH du NTP entrant et permet de discriminer les NTPs des dNTPs. Le motif B compose la partie inférieure de la cavité du site actif et permet la fixation de l'ARN modèle à la RdRp. Le motif B est également impliqué dans l'étape de translocation du produit et de l'ARN modèle. Toute mutation en son sein induit une diminution drastique de l'activité (Sholders and Peersen, 2014; Venkataraman, Prasad and Selvarajan, 2018).

Le motif C est le plus conservé d'entre tous. Il est formé de deux brins β au niveau du domaine de la paume reliés par une boucle constituée des acides aminés SDD chez les sNSVs. Les acides aspartiques conservés du motif C (XDXX) et du motif A (DX_{2-4}D) sont impliqués dans la fixation des ions métalliques. L'acide aspartique terminal du motif C (SDD) est quant à lui impliqué dans le bon

positionnement du phosphate α du NTP entrant tout en coordonnant le métal A. Il permet ensuite d'amorcer l'attaque nucléophile du NTP en diminuant le pK_a du 3'-OH du produit.

Le **motif D** est constitué d'une hélice α et d'une boucle flexible localisées à la périphérie du feuillet β du domaine paume. Il est composé en grande partie d'acides aminés chargés positivement formant un environnement basique à l'entrée du tunnel des NTPs et intervient dans leur sélection. La « fermeture » du site actif qui permet la poursuite du cycle catalytique nécessite la formation d'une liaison hydrogène entre une lysine conservée, protonée et exposée au solvant, et le phosphate β du NTP entrant. La fixation d'un NTP non complémentaire de l'ARN modèle n'entraîne pas la « fermeture » du site actif et bloque la réaction (Yang *et al.*, 2012). Cette lysine permet également de protoner le PPi lors de la réaction de transfert nucléotidique et a pour effet d'augmenter l'activité de la RdRp (Castro *et al.*, 2007, 2009).

Le **motif E** est une épingle à cheveux β localisée à la jonction des domaines de la paume et du pouce. Sa fonction supposée est de permettre le positionnement et la stabilisation du 3'-OH du NTP initiateur lors de la réplication, ou de l'amorce en transcription. Un résidu aromatique conservé et présent en N-terminal stabilise le motif C (Venkataraman, Prasad and Selvarajan, 2018).

Le **motif F** est constitué de deux brins β reliés par une boucle flexible surmontant le domaine de la paume. Il appartient au domaine des doigts et contribue à la formation du tunnel d'entrée des NTPs avec le motif D. Il est composé d'acides aminés chargés positivement, dont une arginine conservée, qui neutralisent les charges négatives des groupement phosphates du NTP entrant. La fixation du promoteur 5' à la RdRp chez les *Peribunyaviridae* et chez les virus influenza entraîne la stabilisation de ce motif qui est sinon désordonné. Celle-ci est nécessaire à l'activité de réplication et de transcription (Pflug *et al.*, 2014; Reich *et al.*, 2014; Gerlach *et al.*, 2015). *A contrario*, le motif F des RdRps des Arenavirus ou des Phenuivirus ne nécessite pas la fixation du promoteur 5' pour être ordonné (Peng *et al.*, 2020; Wang *et al.*, 2020).

Chez les RdRps des sNSVs, le **motif G** est composé d'une hélice α qui contient une arginine conservée interagissant avec le NTP entrant tandis que le **motif H** forme un brin β incluant une lysine conservée qui stabilise, par liaison hydrogène, le motif B (Figure 10) (Gerlach *et al.*, 2015; Venkataraman, Prasad and Selvarajan, 2018).

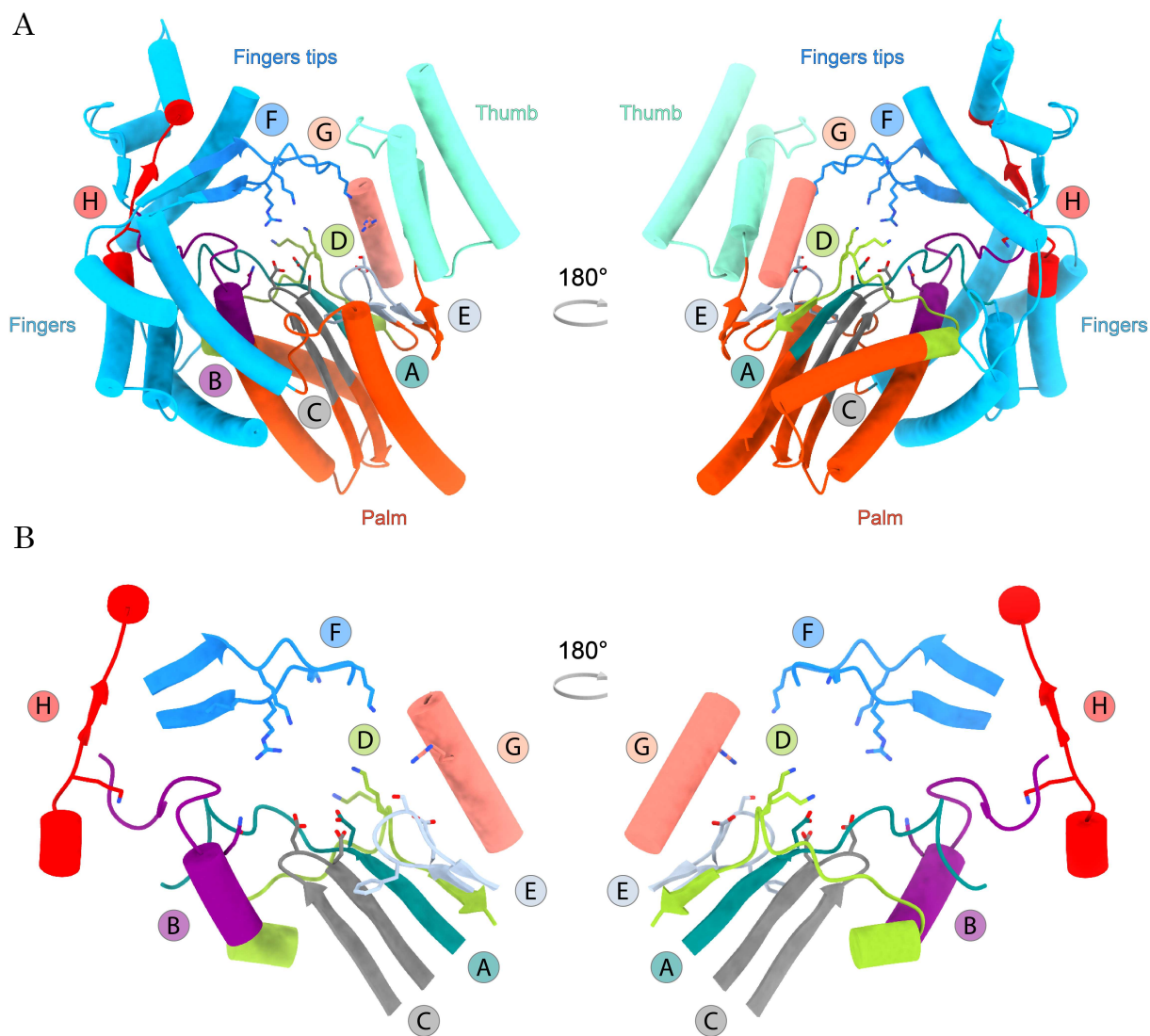


Figure 10. Structure et motifs conservés des RdRps des sNSVs. Le modèle utilisé est la protéine LACV-L1-1750 Δ C-terminal (PDB: 5AMQ). **A**, Vue d'ensemble du site catalytique conservé formé par les domaines des *fingers*, des *fingers tips*, du *thumb*, et de la *palm*. Les motifs de A à H sont représentés et les chaînes latérales des résidus impliqués dans la réaction de transfert nucléotidique sont montrés. **B**, Vue spécifique des différents motifs catalytiques colorés de manière identique à A.

3.2.3. Sites de liaisons des régions 5' et 3' promotrices

L'une des caractéristiques majeures des RdRps des sNSVs est leur capacité à s'associer aux régions conservées et quasi-complémentaires des extrémités 5' et 3' de chacun des segments du génome anti-sens (ARNv) ou sens (ARNc) viral.

Chez les sNSVs et en particulier pour les virus influenza, dont la RdRp a été extensivement étudiée ces dernières années, l'extrémité 5' de l'ARNv (et de l'ARNc) se fixe de manière spécifique entre les sous-unités PA et PB1 à proximité du tunnel d'entrée de l'ARN modèle. Seuls les 10 à 11 premiers nucléotides du 5' se lient à la RdRp en prenant une forme de crochet (*hook*) et deux d'entre eux s'apparient, permettant la stabilisation de cet arrangement (Pflug *et al.*, 2014; Reich *et al.*, 2014; Thierry *et al.*, 2016; Fan *et al.*, 2019; Peng *et al.*, 2019). Les extrémités 5' des ARNv d'autres sNSVs comme LACV chez les *Peribunyaviridae* (Gerlach *et al.*, 2015), les virus Machupo (MACV) et LASV chez les *Arenaviridae* (Peng *et al.*, 2020), adoptent une conformation et/ou disposent d'une poche de fixation similaire, suggérant un mode d'interaction « régions promotrices 5' – RdRp » conservé. Chez les virus des familles *Peribunyaviridae* et *Orthomyxoviridae* la fixation de l'extrémité 5' de l'ARNv sur la RdRp est nécessaire à la stabilisation du motif F, appartenant au domaine des doigts, et permet à la RdRp d'être active *in vitro* en transcription et en réplication lors de l'ajout de l'extrémité 3'. Chez les *Arenaviridae*, l'extrémité 5' jouerait également un rôle de régulateur entre ces deux mécanismes (Peng *et al.*, 2020).

Chez les virus influenza et le LACV de la famille des *Peribunyaviridae*, l'extrémité 3' de l'ARNv dispose de deux sites de liaison à la RdRp utilisés conjointement lors des différentes étapes du cycle de réplication et de transcription.

Le premier site (site A), localisé au niveau du domaine des doigts et du pouce permet la stabilisation de l'extrémité 3' de l'ARNv afin de le positionner au niveau du site actif (Figure 11).

Le deuxième site (site B) permet la fixation de manière séquence spécifique de l'extrémité 3' de l'ARNv entre les domaines du pouce PA-C et PB1 chez les RdRps des virus influenza. La fixation de l'extrémité 3' de l'ARNv sur ce site entre

en jeu dans les étapes de pré-terminaison et terminaison du cycle de transcription (Figure 11) (Wandzik *et al.*, 2020).

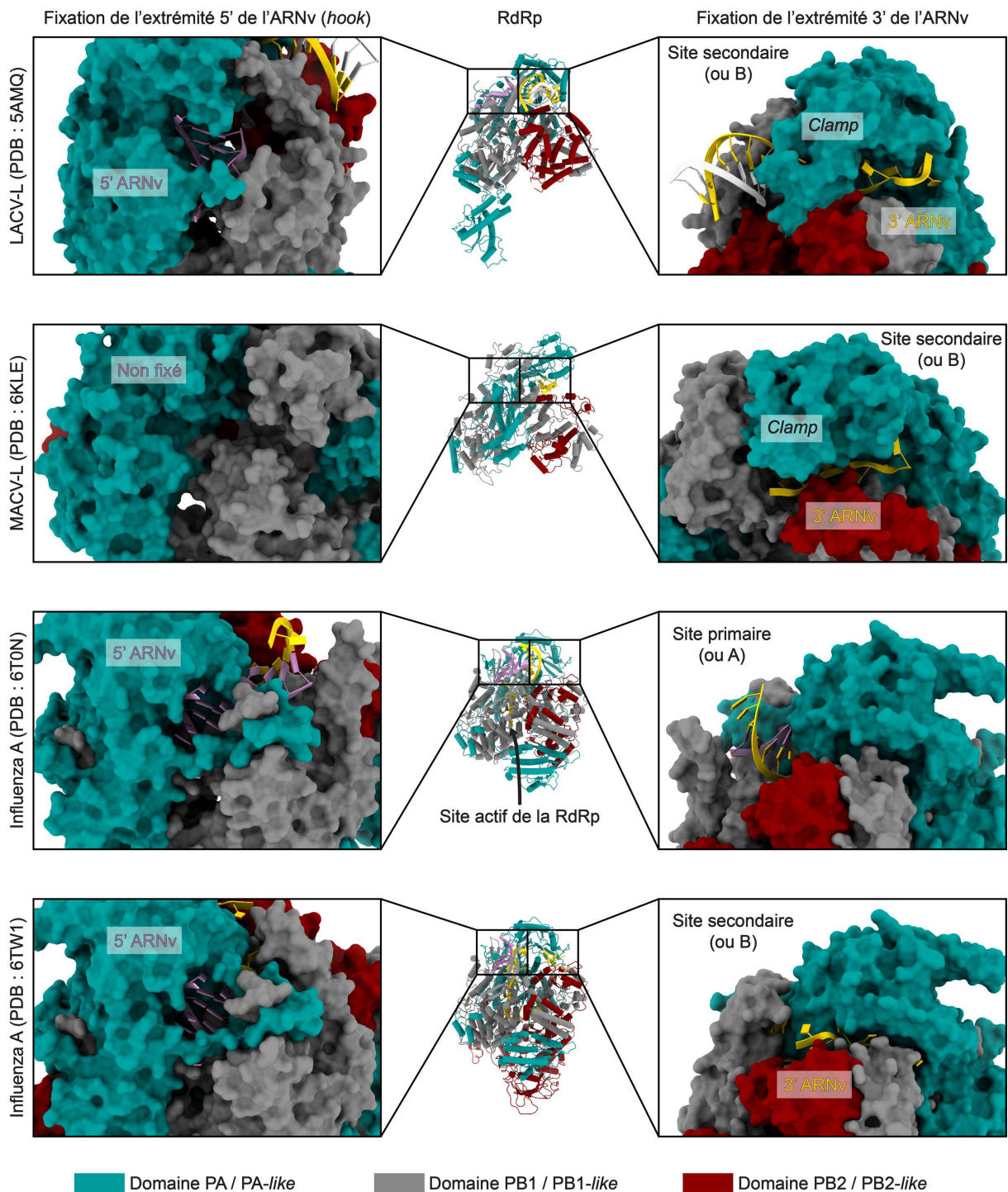


Figure 11. Site de fixation des extrémités 5' et 3' de l'ARNv pour les virus LACV, MACV et Influenza. Le 5' se lie en forme de *hook* dans une poche spécifique localisée entre le domaine PA/PA-like et le domaine PB1/PB1-like. Le 3' dispose de deux sites de fixation : Le site primaire (ou A) au niveau du site actif de la RdRp (PDB : 6T0N) et le site secondaire (ou B) localisé à la surface de la RdRp. Sur le site B, un domaine additionnel, le *clamp* permet la stabilisation du 3'.

Cette poche de fixation est également présente chez les RdRps de la famille des *Peribunyaviridae* (LACV) et des *Arenaviridae* (MACV/LASV). Deux domaines additionnels appelés pince et voute (*clamp/arch*), non présents chez la RdRp d'influenza, se referment sur l'extrémité 3' de l'ARNv lors de sa fixation pour permettre sa stabilisation (Figure 11). En l'absence l'extrémité 3' de l'ARNv, ces régions sont flexibles (Gerlach *et al.*, 2015; Peng *et al.*, 2020).

3.2.4. Réplication du génome viral, initiation de novo et mécanisme d'amorçage - réalignement

Chez les sNSVs, la réplication est un mécanisme nécessitant deux étapes. La première consiste à produire un ARNc à partir de l'ARNv. La seconde consiste à utiliser cet ARNc pour synthétiser de l'ARNv qui sera internalisé dans les virions néosynthétisés. Afin de conserver et maintenir l'intégrité de son génome au cours de ces deux étapes, la RdRp doit coordonner le bon positionnement de l'ARN modèle au niveau du site actif, des motifs qui composent le cœur catalytique et sélectionner les premiers NTPs complémentaires du brin modèle afin de poursuivre l'élongation du produit. A noter que ces RdRp ne possèdent pas de système de relecture, ce qui peut entraîner des erreurs de copies et peut permettre au virus de muter très rapidement.

Comme vu précédemment, l'extrémité 5' du génome est fixée à la RdRp de manière spécifique et permet son activité en ordonnant le motif F. L'extrémité 3' quant à elle dispose de deux sites de fixation, l'un à la surface de la RdRp (site B) et l'autre au niveau du site actif (site A). L'une des premières étapes de la réplication, communément admise pour les sNSVs et supporté par de récentes études sur la RdRp du virus influenza, consiste en la translocation de l'extrémité 3' de l'ARNv du site B au site A (Robb *et al.*, 2016; Wandzik *et al.*, 2020). Une fois l'ARN modèle correctement positionné, l'initiation de la réplication est effectuée *de novo*. Deux scénarios sont alors possibles (Figure 12) :

1. L'initiation peut débuter sur le nucléotide terminal de l'extrémité 3' de l'ARN et permettre l'élongation immédiate du produit ou...
2. L'initiation peut débuter de manière interne à l'ARN modèle en produisant de courtes amorces d'ARNs complémentaires à l'extrémité 3'.

Dans le deuxième cas, afin de pallier le défaut d'incorporation des premiers nucléotides de l'ARN modèle et dans le but de préserver les régions promotrices copiées, ces amorces seront réalignées aux nucléotides terminaux de l'extrémité 3' pour poursuivre l'élongation du produit.

Actuellement, aucune structure n'a été publiée pouvant nous permettre de comprendre de manière exacte le mécanisme d'initiation même si des modèles existent, basés sur des expériences d'activité *in vitro* présentés ci-après.

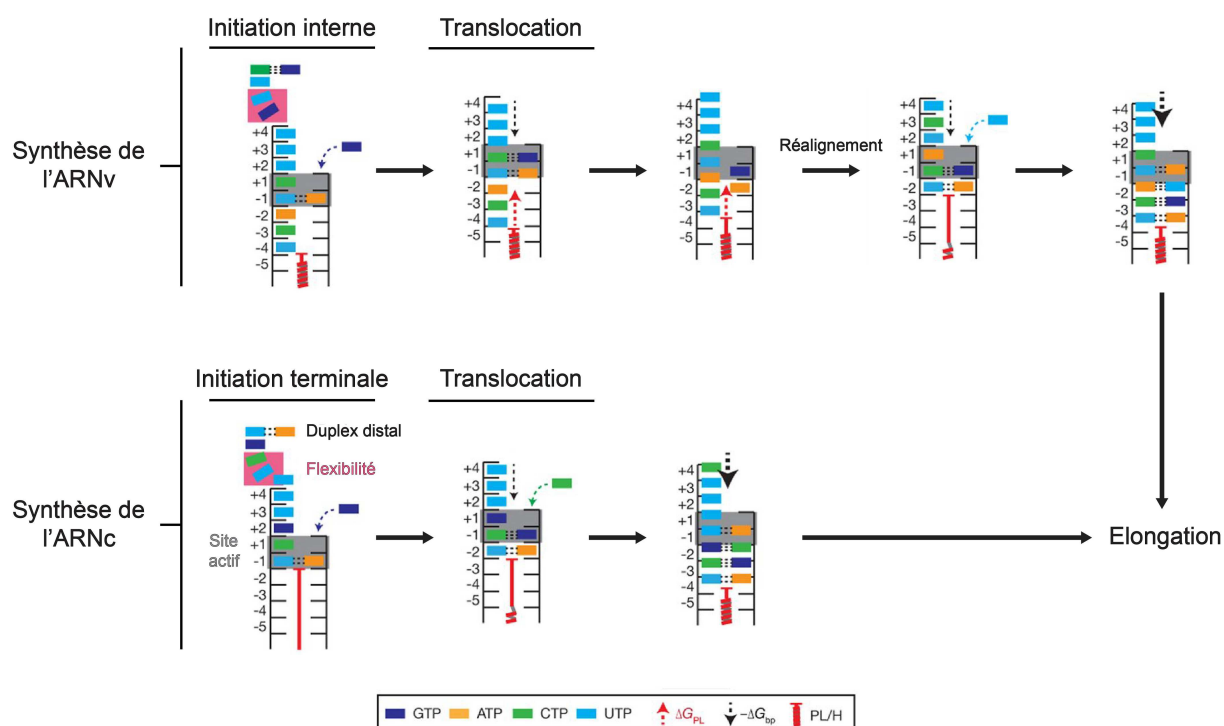


Figure 12. Modèle de l'initiation de la réplication de la RdRp pour le virus influenza A. L'initiation de la synthèse de l'ARNv, à partir de l'ARNc, est catalysée par l'incorporation des nucléotides ATP et GTP sur les nucléotides 4U et 5C du 3' de l'ARNc. Cette initiation interne fait appel à une étape de réalignement, médiée par la *priming loop* et/ou une boucle localisée dans le domaine de la *palm* (PL/H), afin de préserver l'information génétique du virus. La synthèse de l'ARNc, à partir de l'ARNv, est initiée sur les nucléotides 3' terminaux 1U et 2C sans faire appel à un mécanisme de réalignement. Afin de stabiliser l'ATP lors de l'étape d'amorçage, l'extrémité de la *priming loop* est nécessaire. Figure adaptée de Oymans and te Velthuis, 2018.

Chez les virus influenza, les mécanismes de synthèse de l'ARNv et de l'ARNc diffèrent. Il est supposé que, suivant la polarité du brin répliqué, la position de l'extrémité 3' au niveau du site actif varie et adopte deux positions distinctes. Lors de la synthèse de l'ARNv, l'extrémité 3' s'insérerait plus profondément dans la cavité du cœur catalytique plaçant les nucléotides 4U et 5C dans les positions respectives -1 et +1 du site actif. L'initiation serait donc interne au 3'. Elle

débiterait par l'hybridation et l'incorporation des deux NTPs complémentaires A et G du 3' de l'ARNv (3'-UCAUCU...) qui seraient par la suite réalignés sur les nucléotides terminaux 1U et 2C (3'-UCAUCU...). L'utilisation du dinucléotide ApG mime ce mécanisme et augmente l'activité de réplication de la RdRp (Plotch and Krug, 1977). *A contrario*, lors de la synthèse de l'ARNc, les nucléotides 1U et 2C du 3' se placeraient en position -1 et +1 du site actif (Deng, Vreede and Brownlee, 2006; Te Velthuis *et al.*, 2016; Oymans and Te Velthuis, 2018). La boucle d'amorçage du domaine PB1 (ou *priming loop*) permettrait alors de stabiliser l'incorporation des deux premiers NTPs A et G qui pourraient s'hybrider directement avec deux nucléotides terminaux de l'ARNv (3'-UCGUUU) sans avoir à procéder à l'étape de réalignement (Figure 12).

Le mécanisme d'initiation de la réplication varie également en fonction de la complémentarité de séquence des régions promotrices. Chez les *Arenaviridae*, l'extrémité 5' (5'-GCGCA...) de l'ARNv dispose d'un G additionnel non complémentaire du 3' (3'-GCGU...). Cette particularité suggère l'existence chez ces virus d'un mécanisme de réalignement différent des *Orthomyxoviridae* ou d'autres familles de l'ordre des *Bunyvirales*. En effet, afin de permettre le maintien de l'intégrité du génome viral et conserver ce nucléotide additionnel, la RdRp doit incorporer un GTP puis un CTP sur les nucléotides C2 et G3 du 3' (3'-GCGU...). Le C du dinucléotide pppGpC serait ensuite réaligné sur le G1 du 3' (3'-GCGU) laissant le pppGpC dépasser de l'ARN modèle (Garcin and Kolakofsky, 1990, 1992).

Chez les *Peribunyaviridae*, l'extrémité 5' (5'-AGUAGU...) est composée de deux triplets répétés de nucléotides complémentaires à l'extrémité 3' (3'-UCAUCA...). Ceci suggère que l'initiation de la réplication pourrait être réalisée de manière interne sur le nucléotide U4 du 3' (3'-UCAUCA...), puis ensuite allongé via l'incorporation de GTP et UTP pour former un triplet AGU (3'-UCAUCA...) qui est ultérieurement réaligné sur le premier triplet du 3' (3'-UCAUCA...) avant de procéder à l'élongation initiale du produit.

NB : Le dernier article présenté dans cette thèse explique en détail cette étape en se basant sur des résultats d'activité in vitro et sur une structure de la LACV-L

bloquée à l'étape d'initiation de la réplication. Cette structure explique structurellement le fonctionnement de l'initiation chez les Peribunyaviridae et lève le voile sur des fonctionnements possibles chez les autres virus précédemment cités.

Les extrémités des segments du génome de la famille des *Hantaviridae* présentent une organisation similaire à celle des Peribunyavirus. La séquence des extrémités diffère cependant avec un 5'-UAGUAG... et un 3'-AUCAUC.... Le mécanisme d'amorçage et de réalignement proposé par Garcin *et al.*, 1995 suppose l'incorporation d'un G sur le nucléotide C3 du 3' (3'-AUCAUC). L'élongation de l'amorce sur un ou plusieurs nucléotides provoquerait alors son réalignement laissant le pppG(pU(pX)_n) dépasser de l'ARN modèle à la manière du mécanisme proposé pour les Arenavirus. Le pppG ne faisant cependant pas parti du 5' serait alors clivé afin de maintenir la complémentarité du génome.

L'initiation de la réplication des sNSVs est donc en partie dépendante des régions promotrices 5' et 3' qui varient grandement en termes de séquence au sein des différentes familles virales. De plus, la position du 3' au niveau du site actif, les NTPs initiateurs stabilisés par la coordination des motifs et/ou des différents éléments structuraux comme la *priming loop*, sont autant de paramètres qui doivent être coordonnés pour permettre de stabiliser un état transitoire et permettre d'initier la réplication du génome viral.

3.2.5. Implication de(s) priming loop(s) lors de la réplication

Pour que l'initiation *de novo* de la réplication du génome viral puisse se produire, il est nécessaire que le premier nucléotide du produit soit stabilisé afin que son positionnement précis soit compatible avec une attaque nucléophile permettant l'incorporation du nucléotide suivant. La coordination de ces premiers nucléotides est essentielle pour que la réaction puisse débiter. L'un des éléments communs aux protéines L des sNSVs permettant de stabiliser ces NTPs initiateurs est supposée être la boucle d'amorçage ou *priming loop*.

Chez la RdRp du virus influenza, cette boucle est localisée sur le domaine PB1 et forme une épingle à cheveux β traversant la cavité du cœur catalytique en direction du site actif avant l'élongation du produit. *In vivo*, sa délétion entraîne une diminution générale de l'activité de synthèse d'ARN. *In vitro*, il a été démontré

qu'elle est indispensable à la synthèse de l'ARNc, qui fait appel à une initiation terminale au 3', mais semble moins impliquée dans la synthèse de l'ARNv, dont l'initiation est interne au 3'. Les résidus qui composent l'extrémité de l'épingle semblent donc jouer un rôle différent suivant les mécanismes de synthèse de l'ARNc ou de l'ARNv (Te Velhuis *et al.*, 2016). De plus, elle joue également un rôle limitant au moment de l'élongation du produit de réplication et de transcription, récemment confirmé par l'obtention de nouvelles structures des différentes étapes du cycle de transcription de la RdRp d'influenza. Dans ces dernières, *la priming loop* subit d'importants changements de conformation permettant l'accommodation d'un duplex de 9 paires de base au niveau de la cavité du cœur catalytique de la RdRp (Kouba, Drncová and Cusack, 2019; Wandzik *et al.*, 2020).

Récemment, l'importance d'une autre boucle conservée et localisée dans le domaine PB1 de la RdRp d'influenza (266-274) soulève de nouvelles questions quant aux mécanismes d'amorçage et de réalignement. La mutation des résidus qui la composent, notamment la substitution de la V273A, induit des échecs de réalignement plus importants (Oymans and Te Velhuis, 2018).

Chez les *Bunyavirales*, ces deux boucles existent. La *priming loop* varie en longueur et en organisation selon les différentes familles virales, mais elle est toujours structurellement localisée à la même position que chez le virus influenza entre les domaines du *thumb* et du *bridge*. La deuxième boucle existe également. Elle n'est pas conservée en termes de séquence entre les familles virales (sauf chez les *Peribunyaviridae* et les *Hantaviridae*) mais est conservée en position dans le domaine de la *palm*.

NB : Le dernier article de cette thèse explicite clairement le rôle primordial de cette seconde boucle dans l'initiation interne puis le réalignement de l'amorce. Cette boucle a ainsi été nommée Prime and Realign loop (PR loop).

3.3. Domaines catalytiques additionnels

Chez les sNSVs, la protéine L est composée d'un cœur catalytique central RdRp qui permet l'incorporation de nucléotides complémentaires à l'ARN modèle et qui permet d'effectuer la réplication *de novo* du génome viral. Elle dispose

également de deux domaines catalytiques additionnels localisés aux extrémités N et C terminales de la partie RdRp, respectivement l'ENDO et le CBD. Tous deux sont impliqués dans la transcription du génome qui permet de produire des ARNm (codant pour les différentes protéines virales) qui seront ensuite utilisés par la machinerie de traduction de la cellule hôte. La transcription se déroule en deux étapes. Premièrement, le CBD capture un ARNm cellulaire via un site de fixation de la coiffe m⁷GpppX. Cette dernière est une modification co-transcriptionnelle que l'on retrouve à l'extrémité 5' des ARNm dans les cellules eucaryotes. Deuxièmement, cet ARNm cellulaire capturé par le CBD est clivé par l'ENDO. Il en résulte une séquence hétérogène d'une longueur de 10 à 15 nucléotides après la coiffe chez les *Orthomyxoviridae* et de 7 à 25 nucléotides chez la majorité des familles de l'ordre des *Bunyavirales*. Chez les Arenavirus cependant, la longueur maximale de l'ARNm clivé atteint seulement 7 nucléotides (Olschewski, Cusack and Rosenthal, 2020).

Cette amorce est ensuite transloquée au niveau du site actif de la RdRp afin d'initier la transcription. Suivant le degré de complémentarité de l'extrémité 3' de l'amorce avec les premiers nucléotides de l'extrémité 3' de l'ARN modèle et en fonction de la longueur de l'amorce clivée, la RdRp a la possibilité d'effectuer de multiples cycles de *prime and realign* lui permettant de procéder à l'élongation de l'ARNm viral sans perte d'information génétique (Koppstein, Ashour and Bartel, 2015; Amroun, Priet and Querat, 2017; Liu *et al.*, 2018; Te Velthuis and Oymans, 2018).

3.3.1. L'endonucléase

La présence d'un domaine ENDO localisé à l'extrémité N-terminale des protéines L a été démontrée chez de nombreux Bunyavirus et Orthomyxovirus. L'obtention de nombreuses structures cristallographiques de ce domaine exprimé et produit de manière isolée chez différentes familles virales a notamment permis d'apprécier leur conservation structurale. De forme globulaire, l'ENDO est composée de deux lobes. Le premier est principalement composé d'hélices α , additionné d'un feuillet β chez les *Phenuiviridae*. Le deuxième présente un feuillet β majeur composé de trois brins et surmonté d'une hélice α conservée qui contient

la majorité des résidus du site actif (Figure 13) (Yuan *et al.*, 2009; Morin *et al.*, 2010; Reguera, Weber and Cusack, 2010; Wallat *et al.*, 2014; Fernández-García *et al.*, 2016; Reguera *et al.*, 2016; Rosenthal *et al.*, 2017; Jones *et al.*, 2019).

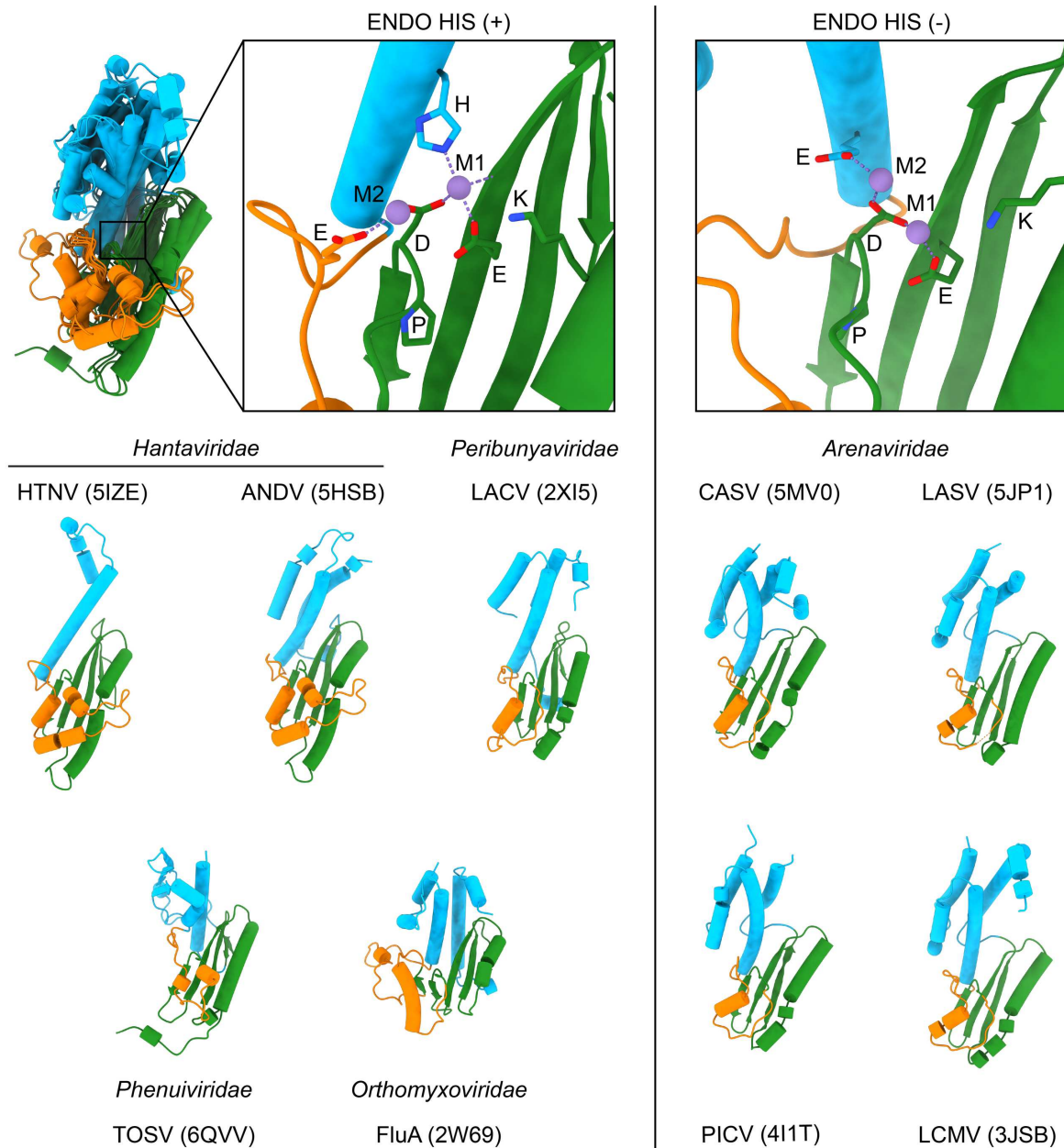


Figure 13. Comparaison des structures connues des domaines ENDO en fonction des groupes auxquels elles appartiennent : ENDO His (+) ou ENDO His (-).

Toutes les endonucléases des Bunyavirus disposent d'un motif catalytique conservé PD(E/D)K qui coordonne deux cations métalliques divalents nommés M1 et M2. La présence ou l'absence d'une histidine en amont du site actif permet de classer les ENDO en deux groupes : His (+) et His (-). Chez les ENDO His (+) à

l'image des familles *Hantaviridae* et *Peribunyaviridae* par exemple, l'histidine coordonne M1. *A contrario*, chez les ENDO His (-) à l'image de la famille des *Arenaviridae*, elle est remplacée par un résidu acide (aspartate ou glutamate) localisé dans une position similaire mais stabilisant M2. Un résidu aspartique/glutamique conservé chez ces deux groupes et localisé sur une boucle flexible proche du site actif est impliquée dans la coordination de M2 chez les ENDO His (+). *A contrario* chez les ENDO His (-), ce résidu est à une distance trop importante du site actif et reste dispensable à l'activité de transcription (Figure 13). La classification des ENDO His (+) / (-) est également supportée par la caractérisation de l'activité *in vitro* du domaine isolé mais également *in vivo* à l'aide de systèmes mini-réplicons.

En effet, les domaines isolés ENDO His (+) ont une activité *in vitro* plus importante que les ENDO His (-) qui se révèle en fait être nulle. *In vivo*, ces dernières sont cependant totalement actives suggérant une activité dépendante du contexte structural dans lequel elles se trouvent, *i.e* au sein de la protéine L entière et/ou des RNPs. L'activité des ENDO est enfin dépendante de la nature des cations métalliques présent. Chez les ENDO His (+) des familles virales *Peribunyaviridae*, *Hantaviridae*, *Phenuiviridae* et *Orthomyxoviridae*, les ions manganèses induisent une activité *in vitro* supérieure aux ions magnésium (Reguera *et al.*, 2016; Holm *et al.*, 2018; Jones *et al.*, 2019).

3.3.2. Le domaine de liaison de la coiffe

Chez les sNSVs, le CBD permet de capturer un ARNm cellulaire et de le présenter au domaine ENDO où il sera clivé. L'amorce coiffée est alors transloquée vers le cœur catalytique de la RdRp afin de produire des ARNm viraux.

Au sein de la polymérase hétérotrimérique du virus influenza, le CBD est localisée sur la sous-unité PB2. La structure cristallographique de ce domaine a notamment permis de mettre en lumière les résidus spécifiques permettant la stabilisation d'un analogue de coiffe, le m⁷GTP. La base méthylée est stabilisée entre deux résidus aromatiques et un acide glutamique conservés tandis que les triphosphates sont stabilisés par des résidus basiques. L'affinité du domaine isolé pour cet analogue est d'environ 170 μ M (Guilligay *et al.*, 2008).

Chez les différentes familles de l'ordre des *Bunyavirales*, la difficulté à localiser un tel domaine dû à la faible conservation de séquence des régions C-terminales des protéines L a laissé sous-entendre que le vol de coiffe puisse être réalisé par les NPs, notamment chez les Arenavirus et les Hantavirus (Qi *et al.*, 2010). Chez les Arenavirus, cette hypothèse a rapidement été réfutée par des études fonctionnelles (Brunotte *et al.*, 2011) tandis que l'obtention du modèle atomique de la NP₁₁₃₋₄₂₉ du virus Hantaan a révélé l'absence d'un tel site de fixation (Olal and Daumke, 2016).

Par contre, de récents travaux ont également révélé la présence d'un CBD au sein de la région C-terminale des protéines L chez les *Arenaviridae* (Rosenthal *et al.*, 2017) et les *Phenuiviridae* (Gogrefe *et al.*, 2019). Ces CBD ont été isolés puis caractérisés au niveau structural et biochimique. Chez les *Arenaviridae*, aucun analogue de coiffe n'a été localisé dans la structure de CBD bien qu'il ait été ajouté au cours de la cristallisation. En revanche chez les *Phenuiviridae*, le m⁷GTP est bien présent. D'une manière similaire à l'interaction PB2-CBD/m⁷GTP chez la polymérase d'influenza, la base méthylée est stabilisée par deux résidus aromatiques et les triphosphates par une arginine, tous conservés chez les protéines L des Phenuivirus. Étonnamment, l'affinité pour le m⁷GTP est bien plus faible que chez influenza, environ 740 μM, suggérant l'intervention de domaines supplémentaires de la protéine L et qui pourraient intervenir dans la stabilisation de l'ARN coiffé au cours du vol de l'ARNm cellulaire (Figure 14).

Dans le contexte des protéines L entières, la présence d'un tel domaine chez les familles des *Phenuiviridae* et *Arenaviridae* a été récemment démontré confortant cette hypothèse (Peng *et al.*, 2020; Wang *et al.*, 2020). L'obtention de ces structures a révélé d'importantes similarités au niveau de leur organisation. Les CBDs disposent tous d'un feuillet β principal comprenant de 5 à 8 brins flanqué par une longue hélice α. Une insertion spécifique à chaque famille virale est également présente entre l'hélice α et le dernier brin β du CBD.

NB : Au début de cette thèse la position du CBD chez les Peribunyaviridae et Hantaviridae restait inconnus. Les travaux menés et présentés dans l'article 2

montre la localisation et la structure du CBD au sein de la LACV-L. L'article 3 permet de visualiser le CBD lié à un ARN coiffé.

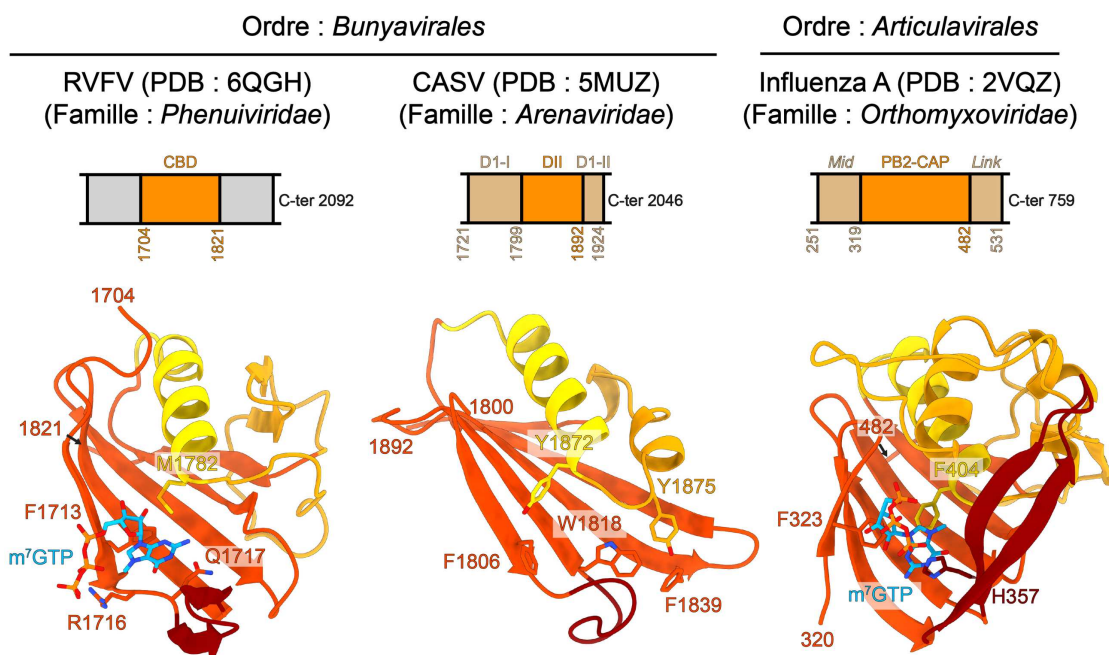


Figure 14. Comparaison des structures des domaines CBD des *Phenuiviridae*, des *Arenaviridae* et des *Orthomyxoviridae* résolues par cristallographie aux rayons X.

3.3.3. Changement de conformations, multimérisation et partenaires cellulaires

Jusqu'au début des années 2015, les données structurales sur les polymérases virales entières des SNSVs étaient encore très rares de par le degré de complexité nécessaire à leur cristallisation. Elles étaient alors cantonnées à des domaines isolés, à l'image de différentes pièces d'un puzzle sans toutefois pouvoir les associer entre elles. Grâce aux développements technologiques récents, notamment en cryo-microscopie électronique (cryo-ME), il est désormais possible non seulement de résoudre les structures des polymérases entières mais également de capturer différents états biologiquement pertinents représentatifs des multiples étapes de réplication et de transcription du génome viral. En effet, de multiples domaines doivent agir de manière coordonnée au sein de la polymérase tout en interagissant avec différents cofacteurs cellulaires engendrant des changements de conformations importants. Ces changements permettent par exemple de

détourner un ARNm cellulaire et de le cliver pour former une amorce coiffée ou de promouvoir la synthèse de l'ARNv et de l'ARNc.

Chez la polymérase du virus influenza, la fixation des régions promotrices 5'/3' de l'ARN est fondamentale à la stabilisation des domaines additionnels ENDO et CBD qui sont liés par des *linker* à la RdRp. De plus, en fonction de leur polarité, la stabilisation du cœur et des domaines catalytiques additionnels diffèrent induisant une activité *in vitro* spécifique au mécanisme de transcription ou de réplication (Thierry *et al.*, 2016; Pflug *et al.*, 2017).

La transcription fait appel à la coordination de multiples domaines afin d'assurer la production d'ARNm viraux. Le vol de coiffe est un mécanisme initié par l'interaction de la polymérase virale avec le domaine C-terminal de la polymérase ARN II. Cette dernière stabilise l'état de pré-initiation permettant au CBD de capturer l'ARNm cellulaire naissant (Lukarska *et al.*, 2017). D'importants changements de conformations de la *priming loop* et du tunnel d'entrée de l'ARN modèle s'opèrent alors pendant l'initiation. Pendant l'élongation, les domaines du *lid*, du *bridge* et du *thumb-ring* s'ouvrent et permettent d'accommoder le duplex produit/ARN modèle dans la cavité de la RdRp (Kouba, Drncová and Cusack, 2019). Enfin, l'étape de terminaison de la transcription lors de laquelle l'extrémité 3' de l'ARNv se fixe sur le site B est suivie d'un état de dissociation permettant séquentiellement,

1. Le relargage du produit de transcription et...
2. Le recyclage de la polymérase.

Ce dernier état permet la réutilisation subséquente de la RdRp en réplication et/ou transcription en raison de la reformation du duplex distal entre les régions promotrices 5' et 3' de l'ARNv (Wandzik *et al.*, 2020). Chez les *Peribunyaviridae*, les différentes étapes initiales de la transcription (*i.e* vol de coiffe, clivage, pré-initiation et initiation) s'accompagnent également de modifications importantes des conformations des domaines ENDO et CBD couplées à une rotation majeure du domaine C-terminal de la protéine L lorsque l'amorce coiffée est transloquée vers le site actif.

Pendant la réplication, l'extrémité 5' de l'ARN nouvellement synthétisé est directement liée à une seconde polymérase qui permettra son encapsidation au cours de l'élongation menant *in fine* à la formation de nouveaux RNPs contenant des copies de l'ARNc (cRNP) ou de l'ARNv (vRNP). *In vitro*, la multimérisation des polymérases virales a déjà été montrée pour les *Arenaviridae* et la polymérase du virus influenza (Chang *et al.*, 2015; Peng *et al.*, 2020).

Jusque-là, seuls deux modes de dimérisation de la polymérase du virus influenza ont été reportés apportant de réelles avancées dans la compréhension du cycle de réplication de ces virus. Le premier mode suggère qu'au cours de la réplication permettant la synthèse de l'ARNv, l'intervention d'une seconde polymérase engendrerait la stabilisation d'un complexe de réplication formé avec une polymérase liée aux extrémités 5'/3' de l'ARNc au sein d'une cRNP, pour permettre *in fine* le positionnement correct des régions promotrices de l'ARNc au niveau du site actif de la RdRp afin d'effectuer une initiation de la réplication terminale. D'un point de vue cellulaire, cette dimérisation diminue la quantité de polymérase virale isolée présente au sein de la cellule infectée et joue donc un rôle de régulateur dans la production de nouvelles copies de l'ARN. Elle limite ainsi la quantité de génome produit ne pouvant être encapsidé dans de nouveaux RNPs tout en prévenant l'activation de la réponse immunitaire de l'hôte (Fan *et al.*, 2019).

Le deuxième mode de dimérisation fait appel à un des nombreux cofacteurs cellulaires interagissant avec la polymérase du virus influenza, ANP32 (Peacock *et al.*, 2019). Cette protéine promeut la formation d'une triade dans laquelle elle stabilise deux polymérases virales formant un dimère asymétrique. Dans cette conformation, la polymérase permettant l'encapsidation est idéalement placée pour se fixer à l'extrémité 5' de l'ARN nouvellement produit par la polymérase en réplication. De plus, l'une des parties flexibles d'ANP32 non résolue pourrait permettre le recrutement des NPs d'une manière similaire à la protéine P lors de la réplication des protéines L des nsNSVs (Carrique *et al.*, 2020).

Chez les Bunyavirus, il est probable que des cofacteurs cellulaires différents interagissent avec les protéines L car la réplication et la transcription se déroulent

dans le cytoplasme et non dans le noyau de la cellule infectée (Rossier, Patterson and Kolakofsky, 1986; Schreur and Kortekaas, 2016).

4. Nucléoprotéines et nucléocapsides chez les sNSVs

4.1. Fonctions principales et structures cristallographiques

Comme introduit précédemment, pour les protéines L des *Bunyavirales* ou au sein des polymérases hétérotrimériques des *Orthomyxoviridae*, les extrémités 5' et 3' de chaque segment des régions promotrices de l'ARN viral sont fixées en tant que simple brin dans deux sites spécifiques localisés à la surface de la polymérase (Pflug *et al.*, 2014; Reich *et al.*, 2014; Gerlach *et al.*, 2015; Peng *et al.*, 2020; Wang *et al.*, 2020). *In vivo*, le reste du génome viral est encapsidé par de multiples copies de NP et l'ensemble ARNv-NPs forment des assemblages généralement flexibles, allant d'une organisation en colliers de perles jusqu'à des hélices appelées NC. En recouvrant le génome viral, les NPs permettent de le protéger des dégradations, empêchent l'hybridation des génomes de polarité négative et positive au cours de la réplication et permettent également de promouvoir le regroupement des RNPs dans de nouveaux virions en interagissant avec plusieurs partenaires cellulaires ou viraux au niveau de la membrane plasmique des cellules (Överby, Pettersson and Neve, 2007; Noton *et al.*, 2009; Hepojoki *et al.*, 2010). Enfin, lors des processus de réplication et de transcription, elles doivent également permettre à la protéine L d'accéder au génome viral au fur et à mesure de sa lecture ce qui suggère un remodelage dynamique des interactions entre les NPs.

Chez les sNSVs, l'étude structurale des NPs a été effectuée dans la majorité des cas par cristallographie aux rayons X. De nombreuses structures ont ainsi été obtenues comme les NPs des virus influenza A, B et D (*Orthomyxoviridae*) (Ye, Krug and Tao, 2006; Ng *et al.*, 2012; Donchet *et al.*, 2019), de RVFV, Toscana (*Phenuiviridae*) (Raymond *et al.*, 2012; Olal *et al.*, 2014), CCHFV (*Nairoviridae*) (Guo *et al.*, 2012), LACV et BUNV (*Peribunyaviridae*) (Ariza *et al.*, 2013; Reguera *et al.*, 2013), LASV (*Arenaviridae*) (Brunotte *et al.*, 2011) ou HTNV (*Hantaviridae*) (Olal and Daumke, 2016).

Les NPs ont été cristallisées sous différents états d'oligomérisations en présence ou en absence d'ARNs mais actuellement aucun lien n'a été établi entre

le degré de multimérisation des NPs et la fixation d'ARN aux NPs. Par exemple, les NPs du virus influenza A, B et D ont été cristallisées en tant que monomères, trimères et tétramères avec ou sans ARN (Ye, Krug and Tao, 2006; Ng *et al.*, 2012; Donchet *et al.*, 2019; Tang *et al.*, 2021), les NPs de RVFV en tant qu'héxamères avec ou sans ARN (Raymond *et al.*, 2010, 2012) tandis que la NP de LACV (LACV-NP) a été cristallisée sous une forme (i) monomérique sans ARN, (ii) tétramérique avec/sans ARN et (iii) pseudo-hélicoïdale sans ARN (Reguera *et al.*, 2013). Bien que ces structures en anneaux montrent comment les NPs interagissent entre elles et avec l'ARN, elles ne permettent pas la couverture totale du génome viral. Pour cela, il est nécessaire de former un assemblage plus long soit en collier de perles extrêmement flexible, soit en assemblage hélicoïdal complexe pouvant prendre différentes conformations (Figure 15).

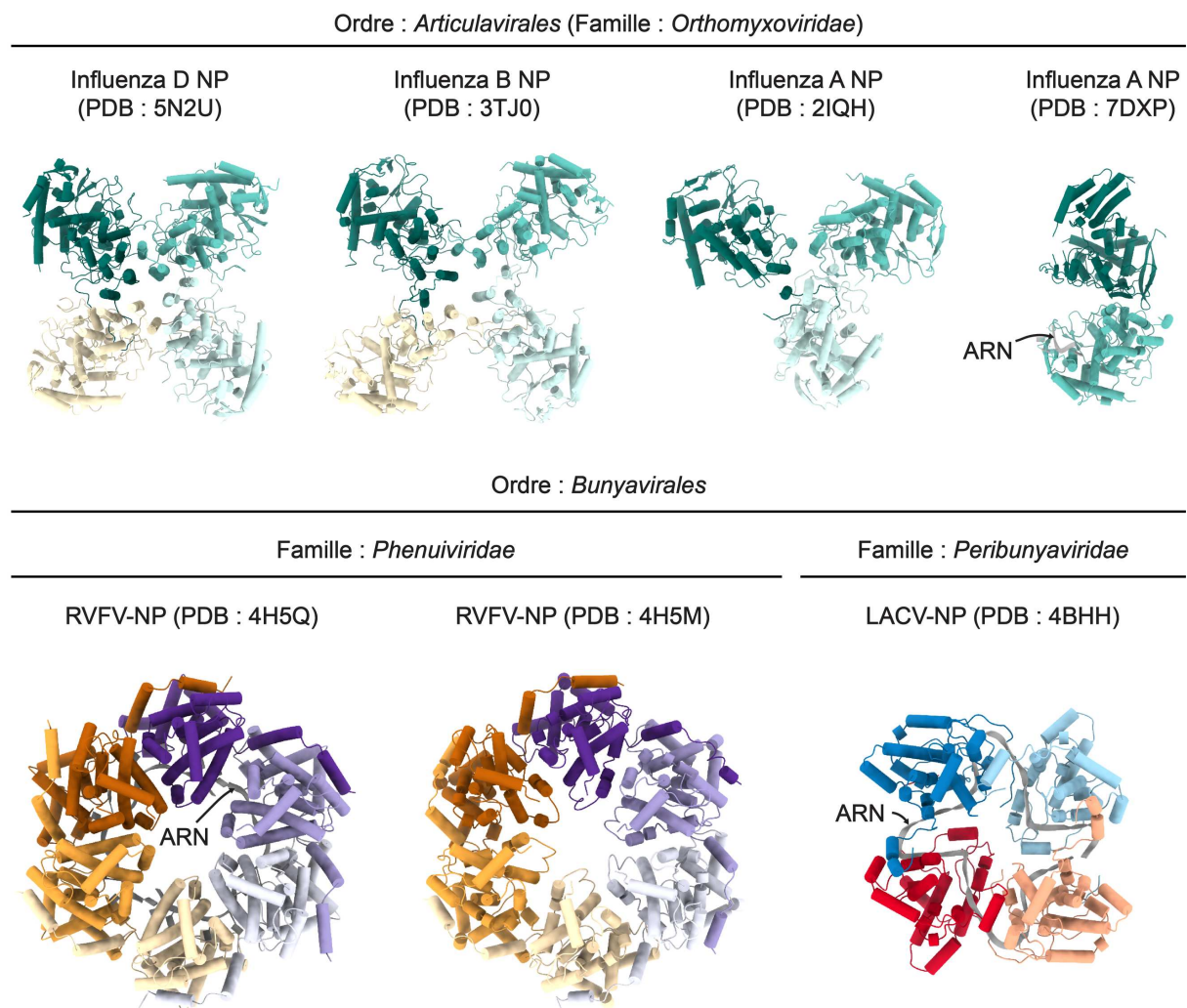


Figure 15. Comparaison des différentes structures cristallographiques de NPs d'*Orthomyxoviridae* (*Articulavirales*) et de représentants des familles des *Phenuiviridae* et des *Peribunyaviridae* (*Bunyvirales*).

4.2. Organisation générale des NPs chez les sNSVs

Les différents états de multimérisation de la NP précédemment reportés sont dus à l'organisation générale des NPs. Celles-ci disposent en effet de deux bras localisés aux extrémités N- et C-terminales reliées au cœur de la NP par deux domaines de liaisons flexibles. Suivant les différentes familles virales, ces protrusions ont une longueur en acides aminés (et donc une taille) plus ou moins importante qui peut directement influencer sur les différentes interactions entre les NPs (Figure 16) (Papageorgiou *et al.*, 2020).

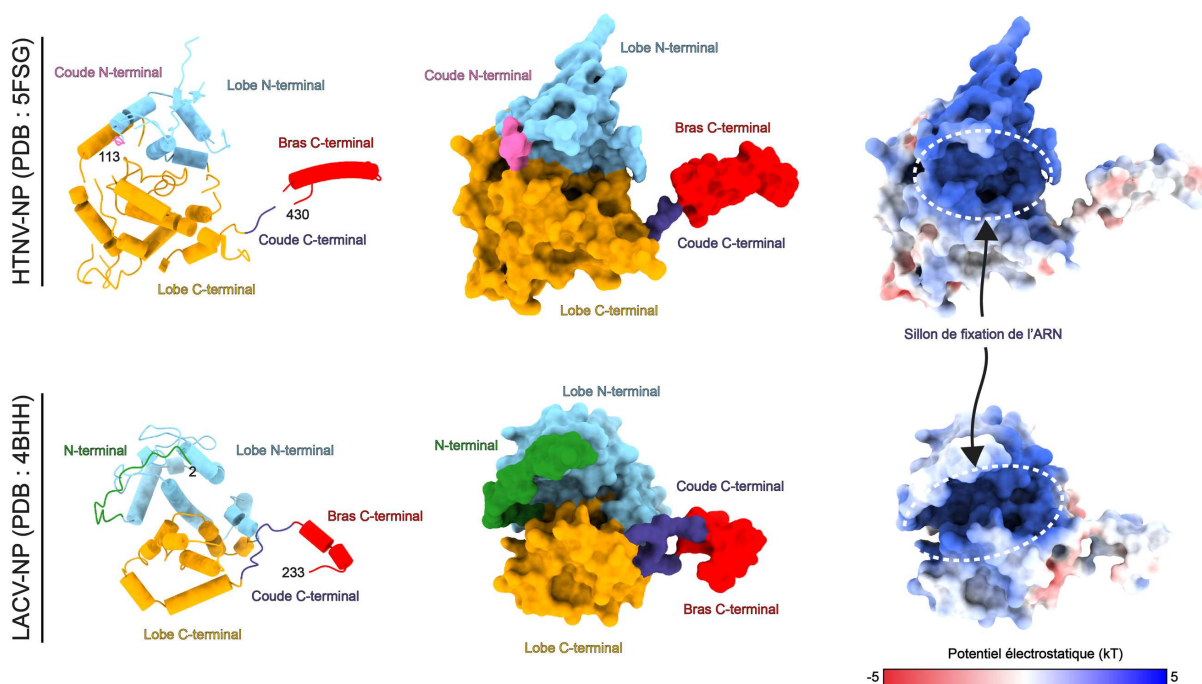


Figure 16. Comparaison des structures des NPs de LACV (LACV-NP) et de HTNV (HTNV-NP). Pour LACV-NP, un monomère a été extrait du tétramère cristallisé et l'ARN n'est pas montré. Les différents domaines sont indiqués. La partie N-terminale de la HTNV-NP est tronquée des résidus 1 à 112 (Olal and Daumke, 2016). Le sillon de fixation de l'ARN chargé positivement est indiqué.

Le cœur des NPs est quant à lui majoritairement composé d'hélices α et partage une architecture globulaire au sein de laquelle un sillon interne permet d'accueillir le génome viral ARN simple brin (Figure 16). Le nombre de nucléotides présents au sein d'une NP varie entre les différentes familles appartenant à l'ordre des *Bunyvirales* : de six à sept chez les Arenavirus, sept chez les Phenuivirus, de six à huit chez les Tospovirus et jusqu'à onze nucléotides chez les Peribunyavirus. Les interactions entre les NPs et l'ARN sont majoritairement réalisées via des résidus chargés comme des lysines et des

arginines qui stabilisent les groupements phosphates du génome. En raison de la diversité des séquences encapsidées, il n'existe aucune interaction base-spécifique mais des liaisons hydrogènes avec le 2'-OH du ribose permettent de préférentiellement sélectionner l'ARN par rapport à l'ADN (Reguera, Cusack and Kolakofsky, 2014). Enfin, dans toutes les structures obtenues, le génome viral est totalement protégé du solvant ce qui indique qu'au cours des mécanismes de réplication et de transcription, l'ARN doit en partie être dissocié des NPs afin d'être rendu accessible à la polymérase virale.

4.3. Flexibilité des RNPs natives chez les sNSVs

Chez les sNSVs, la purification à partir de virus de RNPs natives et leur observation par ME démontre un arrangement flexible et complètement différent des arrangements multimériques des NPs cristallisés (Figure 17) (Ruigrok, Crépin and Kolakofsky, 2011). Il est cependant possible qu'au sein des NCs, certaines régions locales puissent adopter une conformation en forme d'anneau comme il l'a récemment été observé chez les RNPs des virus influenza par cryo-tomographie (Gallagher *et al.*, 2017). Le rôle biologique précis de ces anneaux reste inconnu.

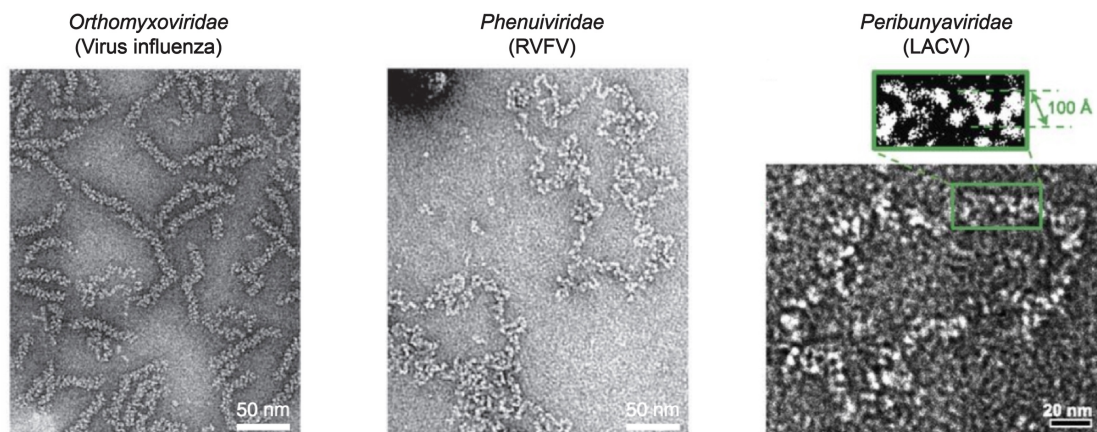


Figure 17. Comparaison des morphologies des RNPs du virus influenza (*Orthomyxoviridae*), du RVFV (*Phenuiviridae*) et du LACV (*Peribunyaviridae*) obtenues par coloration négatives. Figure adaptée de Ruigrok, Crépin and Kolakofsky, 2011 et Reguera *et al.*, 2013.

Chez les *Bunyvirales*, un seul modèle de NC hélicoïdale existe au sein de la famille des *Peribunyaviridae*. Il est basé sur une structure cristallographique de la LACV-NP obtenue à 7.4 Å de résolution qui a la particularité d'avoir cristallisé

dans le groupe d'espace $P4_1$ formant un assemblage hélicoïdal au sein du cristal. Ce dernier dispose d'un diamètre d'environ 100 Å et progresse en suivant une symétrie hélicoïdale de type « main gauche » où quatre NPs sont présentes par tour et dont le *pitch* (pas de l'hélice) a été estimé à environ 54 Å. Bien que l'ARN ne soit pas présent dans cet assemblage hélicoïdal, ces différents paramètres sont en accord avec ceux déduits d'une analyse réalisée en ME par coloration négative sur des RNPs directement purifiées à partir du virus (Figure 17 et 18) (Reguera *et al.*, 2013).

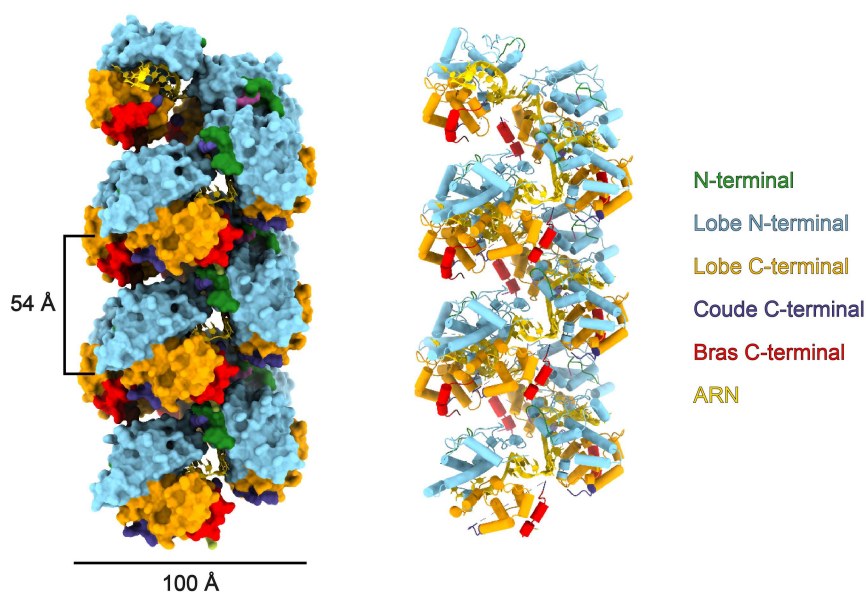


Figure 18. Empaquetage moléculaire de la LACV-NP à partir de sa forme cristallisée dans le groupe d'espace $P4_1$ obtenue à une résolution de 7.4 Å. La largeur et le *pitch* de l'hélice sont indiqués. Les domaines de chaque NPs sont colorés de manière identique. L'ARN a été modélisé à partir de la structure tétramérique de la LACV-NP (PDB : 4BHH). Il est localisé à l'intérieur de la NC et coloré en jaune.

En ce qui concerne l'ordre des *Articulavirales* et en particulier le virus influenza, les RNPs forment une double hélice flexible à l'extrémité de laquelle se fixe la polymérase virale. La flexibilité de ces hélices rend leur analyse structurale particulièrement complexe. Deux études, une basée sur des RNPs purifiées à partir du virus et l'autre basée sur une reconstitution de RNP par transfection avec des plasmides codant pour les composants de la vRNP, ont proposé des modèles de RNP de grippe (Arranz *et al.*, 2012; Moeller *et al.*, 2012). On notera cependant que la main de ces hélices diffère, rendant difficile toute interprétation. Une étude ultérieure réalisée par tomographie, (Gallagher *et al.*, 2017) indique qu'il s'agirait

d'une hélice main droite présentant des courbures, *pitchs* et nombre de sous-unités par tour variable du fait de la grande flexibilité des RNPs.

La flexibilité et les nombreuses conformations potentielles des NCs chez les sNSVs leur permettent ainsi d'adopter des géométries multiples qui doivent nécessairement jouer un rôle important dans l'empaquetage des RNPs mais également permettre le recrutement du bon nombre de segments à incorporer au sein des nouveaux virions.

Chez les virus influenza par exemple, les huit segments individuels sont répliqués au sein du noyau de la cellule infectée avant d'être exportés dans le cytoplasme. Les huit RNPs sont alors empaquetées au niveau de la membrane plasmique au sein des nouveaux virions et se présentent sous forme de huit bâtonnets groupés, comme observé par cryo-tomographie (Noda *et al.*, 2006). Chez les *Bunyavirales*, la réplication et la transcription du génome s'opèrent dans le cytoplasme mais de manière similaire, les RNPs peuvent présenter une certaine rigidité et un arrangement parallèle au sein des virions, notamment chez la famille des *Hantaviridae*, comme en atteste l'analyse des virions des virus Hantaan et Tula par cryo-tomographie (Huisken *et al.*, 2010; Battisti *et al.*, 2011).

Tous ces éléments laissent donc supposer que les différentes formes ou conformations que peuvent prendre les RNPs/NCs jouent un rôle fondamental dans les différentes étapes du cycle de l'infection virale (Reguera, Cusack and Kolakofsky, 2014). En raison de la flexibilité de ces arrangements, l'étude structurale des NCs entières des sNSVs est restée difficile par cryo-ME.

NB : Au début de cette thèse, aucune structure de NCs des sNSVs n'était disponible à haute résolution. Au cours des travaux menés et présentés dans l'article 1, j'ai pu tirer parti de l'apparente rigidité in vivo des NCs de la famille des Hantaviridae pour résoudre la structure de NCs recombinantes du virus Hantaan (HTNV-NC) par cryo-ME à 3.3 Å de résolution.

Chapitre II : La cryo-ME en bref !

La détermination de la structure à haute résolution d'une macromolécule biologique est une étape clé dans la compréhension détaillée des mécanismes qui régissent son fonctionnement au sein d'un organisme. Ces données structurales permettent également d'ouvrir de nouvelles voies thérapeutiques dans la création de molécules visant à inhiber ou bloquer des sites d'activités importants (Renaud *et al.*, 2018).

Pendant de nombreuses années, les méthodes privilégiées de détermination de structure étaient la cristallographie aux rayons X ainsi que la résonance magnétique nucléaire. Elles continuent évidemment à se développer notamment en termes de sources lumineuses, de plus en plus brillantes, de spectromètres à très hauts champs, de logiciels de traitement de données de plus en plus accessibles ou en termes de facilité d'accès à la communauté scientifique.

L'utilisation de la ME, qui est une technique historiquement aussi vieille que la cristallographie, n'a pourtant connu un regain d'intérêt il n'y a que quelques années. Elle était jusqu'à récemment en grande partie limitée à la science des matériaux qui ne souffrent pas des dommages aux radiations *a contrario* des échantillons biologiques qui existent dans un environnement aqueux et qui sont majoritairement composés d'atomes légers, *i.e* carbone, oxygène, azote, hydrogène, ne produisant que très peu de contraste. Elle a tout de même permis d'obtenir les premières structures de virus et de ribosomes à des résolutions comprises entre 20 et 40 Å et représentaient alors les premières images directes et volumes tridimensionnels de ces objets.

En microscopie électronique à transmission (MET), les images obtenues correspondent à des projections bidimensionnelles d'un objet tridimensionnel. Au moment de la préparation de la grille, que ça soit par vitrification ou coloration négative, de multiples copies de notre objet se retrouvent sur la grille de microscopie électronique dans des orientations différentes et, dans le meilleur des cas, aléatoires. Les images collectées regrouperont donc les projections correspondantes de notre objet.

C'est ici que le théorème de la section centrale prend son sens : la transformée de Fourier (TF) d'une projection 2D d'un objet 3D correspond à une

section centrale de la TF 3D de ce même objet. L'orientation de cette section dans la TF 3D dépend de l'orientation de départ de la particule sous le faisceau d'électrons incident. Par conséquent, si les objets étudiés adoptent suffisamment d'orientations différentes dans la glace vitreuse, il est possible de repositionner les différentes tranches des TF 2D correspondantes au sein de la TF 3D. Il sera ainsi possible de recalculer la structure originale de notre objet par TF inverse (De Rosier and Klug, 1968) en ayant bien-sûr conscience que la résolution atteinte dépendra de l'échantillonnage dans la TF 3D.

Les images 2D collectées sont malheureusement extrêmement bruitées et souffrent de détériorations en raison des conditions d'imageries utilisées et des aberrations optiques intrinsèques au MET. Par conséquent, il est nécessaire de collecter plusieurs milliers d'images (contenant chacune plusieurs dizaines voire centaines de particules) dans le but de moyennner les images des particules qui sont similaires et de renforcer le signal qui nous permettra *in fine* de les aligner entre elles (Frank, 1975).

C'est grâce aux multiples avancées matérielles et informatiques réalisées concernant la préparation des échantillons, l'amélioration des sources d'électrons, la mise au point des détecteurs directs, l'automatisation des collectes et des logiciels de traitement d'images que nous avons pu assister à une véritable révolution en terme de résolution atteignable en cryo-ME appliquée à l'étude de protéines isolées (Kühlbrandt, 2014). Cette technique offre désormais la possibilité de déterminer des structures à résolution (quasi-)atomique de manière (quasi-)routinière pour des objets biologiques de tailles variant d'environ 100 kDa à plusieurs méga-daltons. Il est également possible de résoudre des structures d'objets difficilement cristallisables comme des assemblages hélicoïdaux ou des protéines membranaires.

L'un des principaux avantages de la cryo-ME réside également dans la possibilité d'étudier des échantillons hétérogènes, dans la limite du raisonnable, ou intrinsèquement flexibles et de faire varier les conditions de tampon des échantillons. En séparant les images par de multiples classifications il est alors possible d'obtenir différentes structures d'états transitoires. Ces différentes

structures permettent d'obtenir de véritables arrêts sur images des différents changements de conformations de notre objet au sein d'un environnement quasi-natif.

Évidemment, cette révolution n'en n'est qu'à ses débuts. Certaines limitations existent encore en ce qui concerne la méthode de congélation des grilles et ce, malgré l'automatisation de la procédure depuis les travaux de Jacques Dubochet (Dubochet and McDowell, 1981; Dubochet *et al.*, 1988). Les problèmes principaux restent les interactions des protéines à l'interface air-eau qui imposent des contraintes importantes aux protéines et induisent leur dénaturation tout en les poussant à adopter des orientations préférentielles dans la glace. D'autres problèmes, intrinsèques, à la ME persistent également : les dommages liés aux radiations, les mouvements induits par le faisceau d'électrons au contact de l'échantillon et les aberrations optiques des lentilles électromagnétiques du microscope.

D'un point de vue logistique, l'accessibilité des équipements de pointe s'améliore peu à peu et les temps de collectes deviennent de plus en plus courts même si malgré cela, le coût financier d'achats et d'entretiens de ces équipements restent exorbitant pour la majorité des laboratoires souhaitant se lancer dans la cryo-ME. Tous ces facteurs sont à prendre en ligne de compte dans le futur de cette technique qui devrait encore voir de nombreuses avancées se succéder au cours des prochaines années.

Afin d'aborder le sujet de manière « simplifié », les sections qui vont suivre reprennent de manière succincte les différents concepts théoriques et pratiques permettant d'obtenir une structure 3D à haute résolution de notre objet d'intérêt « particules isolées » par cryo-ME. Nous aborderons ainsi la préparation de l'échantillon par vitrification, la formation d'images imparfaites due à la physique des MET et comment corriger ces altérations, les dommages liés aux radiations et les mouvements induit par le faisceau d'électrons. Enfin, nous aborderons parallèlement les récentes avancées technologiques et logiciels qui ont permis à la cryo-ME de se faire une place privilégiée au sein des techniques de détermination de structures 3D de macromolécules biologiques.

Cette procédure possède cependant des points négatifs. De nombreux paramètres restent difficiles à établir ou à contrôler tels que le temps de *blot*, la création d'interfaces air-eau lors de l'interaction des papiers filtres avec la grille, la conductivité thermique de la grille qui influe sur la vitesse de vitrification de l'échantillon et même la façon de tenir la grille avec des pinces dédiées sont tout autant de facteurs qui vont avoir une influence sur la reproductibilité de la congélation (Kasas *et al.*, 2003; Armstrong *et al.*, 2020).

Des techniques alternatives en termes de préparation et vitrification des échantillons existent également et permettent de contourner les problèmes cités ci-dessus. En terme de dépôt d'échantillon, la projection de gouttelettes permet d'éviter les problèmes liés à la création d'interfaces air-eau (Berriman and Unwin, 1994). Elle est réalisée par différentes techniques de pulvérisation permettant la formation de gouttelettes dont le diamètre varie de 0.25 à 25 μm . Elle peut être de type ultrasonique, (Ashtiani *et al.*, 2018; Rubinstein *et al.*, 2019), électrostatique (White *et al.*, 2003) ou « jet d'encre » (Dandey *et al.*, 2020). En plus d'être plus rapide que la technique de blot standard, de l'ordre de la milliseconde contre quelques secondes, elle permet en combinant plusieurs canaux de pulvérisation contenant par exemple, la protéine et son substrat, d'offrir la possibilité d'analyser de potentielles interactions se produisant dans cet intervalle de temps.

La deuxième technique de dépôt d'échantillon réinvente le principe de l'écriture. La grille est placée à une dizaine de micromètre de la pointe du stylet qui est (i) préalablement trempée dans notre échantillon (Ravelli *et al.*, 2020) ou (ii) rempli par ce dernier (Arnold *et al.*, 2017). Ensuite, par effet de capillarité ou en induisant une légère pression à la sortie du stylet, un pont liquide se forme permettant de déposer notre échantillon à la surface de la grille. L'épaisseur de la glace est ainsi contrôlée par le volume que l'on décide de déposer mais également par la vitesse de dépôt et la distance entre la pointe du stylet et la grille. L'un des avantages principaux de cette technique réside dans la possibilité de déposer de multiples échantillons sur un seul et même support (Weissenberger, Henderikx and Peters, 2021).

Enfin, en termes de vitrification et dans un souci d'économie de manipulation et d'automatisation, il est possible de vitrifier son échantillon sur des grilles pré-montées dans leur support « *autogrid* ». Ainsi, à défaut de plonger la grille dans un bain d'éthane liquide, le cryogène est directement pulvérisé de part et d'autre du centre de la grille (Ravelli *et al.*, 2020). Ces ensembles grilles-*autogrids* sont également insérés directement et de manière automatique dans la cartouche qui ira par la suite dans le microscope.

Pour conclure, notre échantillon est donc congelé sur une grille de cryo-ME au sein d'une fine couche d'eau vitreuse transparente aux électrons. Après avoir été insérée dans le microscope, les questions suivantes se posent : comment une image est-elle formée, quelles sont les transformations qu'elles subissent au sein d'un MET disposant d'un système optique imparfait et comment l'échantillon réagit-il au contact d'un faisceau d'électrons ?

2. Formation de l'image en MET, contraste et CTF

En microscopie, la notion de résolution est fondamentale. Elle est définie comme étant la plus petite distance entre deux points d'un objet que l'on peut distinguer sans ambiguïté. Celle-ci dépend en majeure partie de la longueur d'onde de la source lumineuse utilisée. En la diminuant, la résolution maximale atteignable augmente suivant le principe de la limite de diffraction établie par Ernst Abbe en 1873. En MET, l'utilisation d'un faisceau d'électrons accélérés dans une colonne sous vide permet d'atteindre des résolutions théoriques de l'ordre de quelques centièmes d'angström pour des accélérations comprises entre 80 à 300 kV (par exemple à 200 kV la longueur d'onde des électrons est de 0.021 Å). La longueur d'onde n'est donc pas le critère qui va limiter la résolution dans l'imagerie de macromolécules biologiques.

Le faisceau lumineux utilisé en cryo-ME est donc composé d'électrons, qui sont des particules élémentaires présentant une charge négative. En 1924, Louis de Broglie émet le postulat de sa dualité onde-particule : tout élément particulaire dispose de propriétés ondulatoires caractérisées par sa longueur d'onde λ relative à son moment p tels que :

$$\lambda = h / p = h / mv$$

(h : la constante de Planck ; m : la masse et v la vitesse)

En tant que première approximation, cette équation est d'une importance fondamentale. Elle signifie qu'un électron accéléré n'agit pas seulement en tant que particule cinétique, mais également en tant qu'onde. Les ondes d'un faisceau d'électrons qui disposent d'amplitudes et de phases identiques sont considérés comme cohérentes, dans le cas contraire on parlera d'ondes incohérentes. En MET et plus spécifiquement dans la conception des canons à électrons, on cherchera toujours à disposer d'une source lumineuse la plus cohérente possible à une énergie spécifique et dans un vide important afin d'éviter toute perturbation du front d'onde incident. Mais comment le faisceau d'électrons interagit-il avec notre échantillon pour former une image en cryo-ME ?

Le contraste des images obtenues en cryo-ME provient de deux phénomènes. Le premier consiste en l'absorption d'une partie de l'énergie du faisceau incident : c'est ce qu'on appelle le contraste d'amplitude (Figure 20). Malheureusement, en cryo-ME, seule une petite fraction des électrons est absorbée par l'échantillon car les protéines, essentiellement composées d'atomes légers (H, O, C et N, $z \leq 8$), et le milieu dans lequel elles se trouvent, l'eau ($z \leq 8$), n'engendrent que peu de variations d'amplitudes. *A contrario*, les électrons peuvent également être déviés lors de leurs interactions avec l'échantillon, de sorte que le nombre total d'électrons entre le front d'onde incident et le front d'onde diffusé reste identique et sans perte d'énergie : c'est ce qu'on appelle le contraste de phase (Figure 20). Les électrons ayant interagit avec l'échantillon subissent donc une diffusion à des angles différents qui dépend des longueurs de trajet effectuées en traversant l'échantillon. Les variations de phase sont codées dans le front d'onde sortant et rendues visibles en étant converties en variations d'intensités, engendrées par les interactions entre le faisceau d'électron et l'échantillon (Orlova and Saibil, 2011). Ces variations d'intensités sont ultérieurement enregistrées par un système de détection, dont nous aborderons les différents types et caractéristiques ultérieurement.

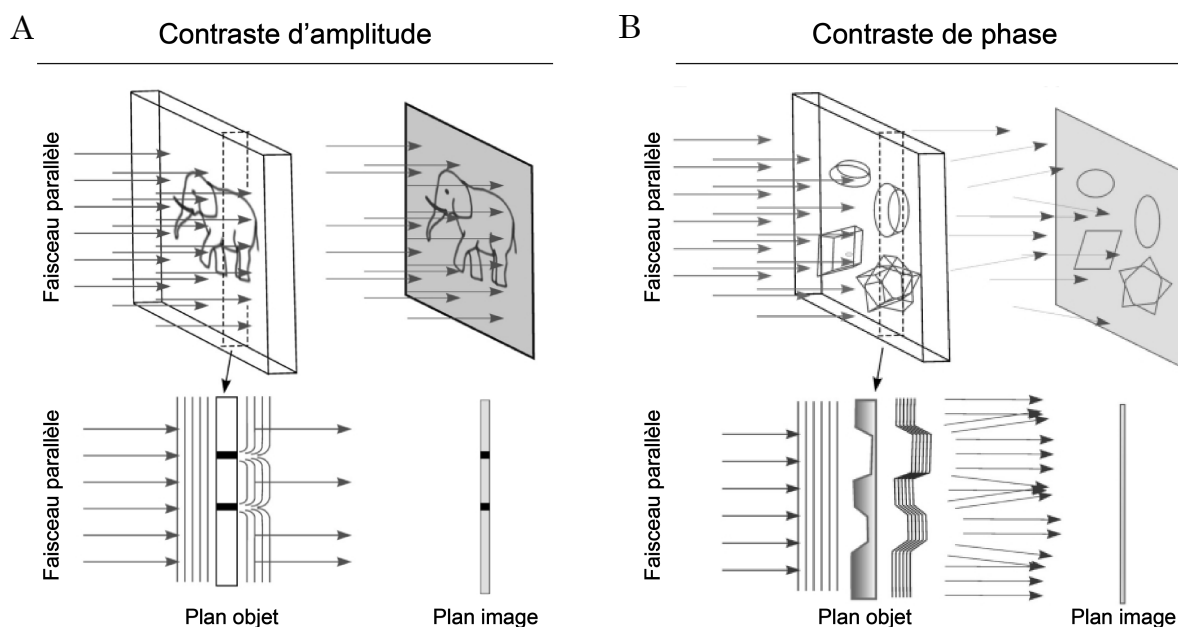


Figure 20. Contraste d'amplitude et de phase en MET. **A**, Notre objet est éclairé par un faisceau parallèle et cohérent. Une partie de l'énergie du faisceau incident est absorbée par l'échantillon modifiant le front d'onde entre le plan objet et le plan image. **B**, Notre objet est éclairé par un faisceau parallèle et cohérent. Les électrons sont déviés lorsqu'ils interagissent avec notre objet, l'intensité n'est pas modifiée mais leurs différences au niveau du plan image sont faibles. Figure adaptée de Orlova and Saibil, 2011.

NB : Dans le cas de la coloration négative, l'utilisation d'atomes lourds comme l'uranium ou le tungstène ($z \geq 74$) permet d'augmenter significativement le contraste d'amplitude des images obtenues ce qui facilite la visualisation de notre protéine. Le désavantage principal de cette technique réside dans le fait que ces atomes ne pénètrent pas notre échantillon mais forme une enveloppe autour de celui-ci, limitant la résolution atteignable à environ 20 \AA .

De plus, si l'on considère notre système optique comme parfait et l'objet focalisé, le front d'onde dans le plan image est (quasi-)équivalent au front d'onde situé juste derrière l'objet. En d'autres termes, la différence de phase entre le front d'onde qui traverse l'échantillon sans interagir avec notre objet, *i.e* le faisceau d'électrons direct, et le front d'onde qui a interagit avec notre objet n'induit aucune différence d'intensité mesurable au niveau du détecteur, équivalent au plan image.

Afin d'améliorer le contraste de nos images, nous sommes obligés de défocaliser, *i.e* de déplacer notre détecteur par rapport à notre plan image ou en

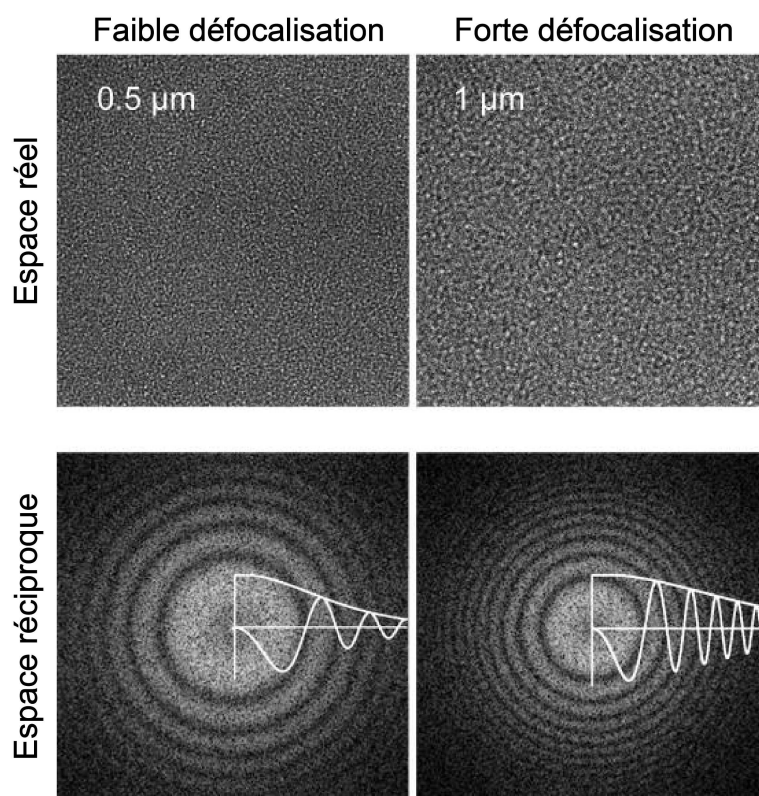


Figure 21. Images de films de carbone et leurs spectres de puissance correspondants collectés à deux défocalisations différentes. A $1 \mu\text{m}$ de défocalisation, les anneaux de Thon sont situés plus près de l'origine et oscillent plus rapidement. Figure adaptée de Orlova and Saibil, 2011.

réalité de déplacer le plan image par rapport au détecteur dont la position est fixe. Cela entraîne un retard de phase supplémentaire et la déformation du front d'onde au niveau du plan focal arrière. L'effet de cette défocalisation entraîne une aberration spécifiquement introduite par l'opérateur afin de visualiser son objet d'intérêt (Figure 21).

Dans l'espace direct, chacune des images obtenues en MET est donc convoluée par une fonction d'étalement du point relative aux différents paramètres d'imagerie utilisés : l'accélération des électrons (généralement comprise entre 120 et 300 kV), la défocalisation induite par l'opérateur et les aberrations des lentilles. La TF de chacune de ces images nous permet de passer de l'espace direct à l'espace réciproque afin de visualiser cette fonction d'étalement du point sous la forme d'une fonction oscillatoire sinusoïdale amortie qui dépend de la fréquence, des amplitudes et des phases de l'image : c'est la fonction de transfert de contraste (CTF) définit telle que :

$$\text{CTF}(\Delta z, s) = \sin\left(2\pi\left[-\frac{s^2\lambda\Delta z}{2} + \frac{\lambda^3 s^4 C_s}{4}\right]\right)$$

(Δz : la défocalisation ; s : la fréquence spatiale ; λ : la longueur d'onde du faisceau d'électrons ; C_s : les aberrations sphériques)

Le spectre de puissance de la CTF est caractérisé par de multiples anneaux concentriques de Thon, dont l'amplitude varie positivement et négativement entre chaque passage à zéro où le signal est nul et donc perdu en espace réel. Pour la CTF, la phase va également changer de 180° à chaque passage par un « zéro ». Sur toutes les plages des basses aux hautes fréquences, la CTF est également modulée par une fonction d'enveloppe, dépendante des conditions d'imagerie et de la nature du faisceau d'électrons, qui induit une diminution progressive de l'intensité du signal aux hautes fréquences (Figure 21) (Erickson and Klug, 1970; Wade, 1992).

En pratique, dans le but de récupérer cette information perdue et de couvrir la majorité des plages de fréquences, il est fondamental de collecter de multiples images à différentes défocalisations. Il sera également nécessaire de corriger

l'inversion des amplitudes présentent entre chaque oscillation et de compenser la diminution des amplitudes aux hautes fréquences (Orlova and Saibil, 2011).

En estimant précisément la CTF de chacune des images obtenues à l'aide de logiciels dédiés (Rohou and Grigorieff, 2015; Zhang, 2016) puis en la corrigeant, la perte d'informations est diminuée et l'on obtient des images dans lesquelles le bruit tout comme l'information des hautes fréquences ont été artificiellement augmentés.

NB : Dans le cas spécifique de l'étude d'objet de taille importante comme des virus entiers, l'assomption d'une seule valeur de défocalisation peut être considérée comme erronée et implique une correction supplémentaire de l'effet de la sphère d'Ewald (Crowther, Huxley and Klug, 1971; Zhu et al., 2018)

3. Dommages liés aux radiations, effets de charge et mouvements

Comme vu précédemment, notre échantillon biologique est congelé dans une fine couche de glace vitreuse d'une épaisseur généralement située autour d'une centaine de nanomètres et idéalement juste légèrement supérieure au diamètre de la particule étudiée. Lorsqu'un faisceau d'électrons accélérés passe à travers un échantillon fin, la majorité des électrons ne sont pas diffusés mais directement transmis sans interactions avec l'échantillon, c'est ce qu'on appelle le faisceau direct. D'autres électrons interagissent avec l'échantillon de manière élastique, *i.e* sans perte d'énergie, et contribuent à la formation de l'image. D'autres enfin interagissent de manière inélastique, *i.e* avec perte d'énergie, et engendrent la détérioration de notre échantillon et *in fine*, de la qualité de nos images (Figure 22).

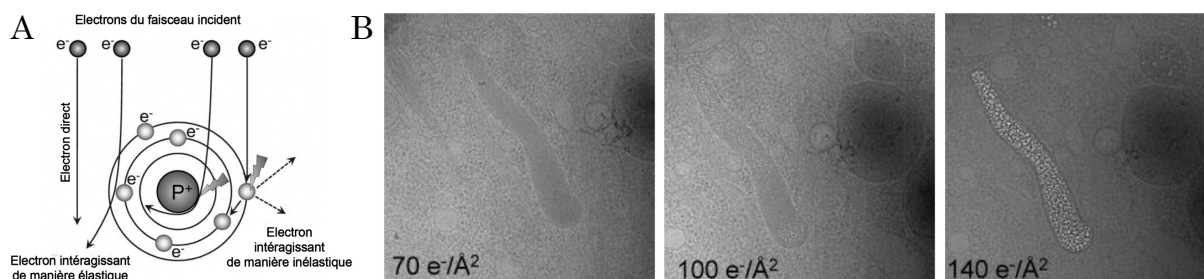


Figure 22. Interactions du faisceau d'électrons incident avec un atome de notre échantillon. **A**, Schéma de la diffusion élastique et inélastique des électrons. La collision des électrons du faisceau avec certains électrons ou le noyau d'un atome entraîne une perte d'énergie (diffusion inélastique), tandis que la déviation par le nuage d'électrons ne modifie pas l'énergie de l'électron incident (diffusion élastique). **B**, Dommages liés aux radiations au cours de l'exposition au faisceau d'électrons. La dose d'électrons est indiquée sur les images. Figure adaptée de Orlova and Saibil, 2011.

La diffusion élastique consiste en l'interaction entre un électron du faisceau incident et le potentiel de Coulomb du noyau d'un atome de l'échantillon. L'électron incident ne perd pas d'énergie lors de ce phénomène et il contribue, lorsque mis en relation avec le faisceau d'électrons non diffusé, à la formation du contraste de phase. Il participe également au contraste d'amplitude car certains électrons très déviés seront arrêtés soit (i) par le diaphragme ou par (ii) les parois de la colonne. La diffusion inélastique consiste quant à elle à l'interaction entre un électron du faisceau et le nuage électronique appartenant à un atome de notre objet. Ainsi,

plus l'échantillon est épais, plus la diffusion inélastique augmente car le faisceau d'électrons entre en contact avec un nombre plus important d'atomes de notre échantillon ou de l'eau vitrifiée dans lequel il se trouve. Cela induit une dégradation des objets avec l'apparition de radicaux libres, la radiolyse de l'eau et entraînera également une diminution du contraste global car le bruit dans l'image finale en sera augmenté (Glaeser, 2016).

Afin de garantir un contraste optimal, l'épaisseur de l'échantillon doit donc être suffisamment faible pour permettre aux électrons de traverser l'échantillon sans occasionner, ou tout du moins limiter, la diffusion inélastique. En analyse de particules isolées par cryo-ME, la limite d'épaisseur de l'échantillon est d'environ 300 nm en utilisant une source d'électrons accélérés à 300 kV. En diminuant l'accélération, cette épaisseur théorique diminue également, en plus de voir la probabilité d'évènements de diffusion secondaire augmenter car les électrons ne disposent plus d'un pouvoir de pénétration suffisant (Russo and Passmore, 2016). En pratique, notre échantillon est détérioré par la diffusion inélastique dès son exposition au faisceau d'électrons et à un rythme supérieur de la contribution de la diffusion élastique à la formation de l'image (Henderson, 1995).

Des modes d'imagerie à faible dose ont donc été développés afin de limiter les effets de la diffusion inélastique et notamment la production de radicaux libres provenant de l'effet ionisant du faisceau. Actuellement, ces modes d'imagerie permettent également la détection individuelle des électrons grâce à de nouvelles caméras dont les caractéristiques seront décrites dans la prochaine section. Malgré l'utilisation de mode d'imagerie à faible dose, généralement contenue sous les $60 e^-/\text{Å}^2$ totaux, l'énergie déposée par le faisceau d'électrons entraîne un effet de charge. Ce dernier se caractérise à différentes échelles tout au long de l'exposition et tout particulièrement au début de celle-ci. Il entraîne en effet une relaxation du support de la grille et des tensions accumulées au niveau de la glace vitrifiée (Figure 23) (Russo and Passmore, 2016). Suivant un ordre de magnitude décroissant nous avons : le mouvement du support qui compose la grille, la flexion de la fine couche de glace vitrifiée, le mouvement brownien de l'eau et enfin les vibrations moléculaires. Tous ces éléments occasionnent des rotations et des translations différentes pour chacune des particules présentes dans la glace et influent

directement sur la qualité des images obtenues et donc sur la qualité de la reconstruction 3D pouvant être recalculée (Naydenova, Jia and Russo, 2020).

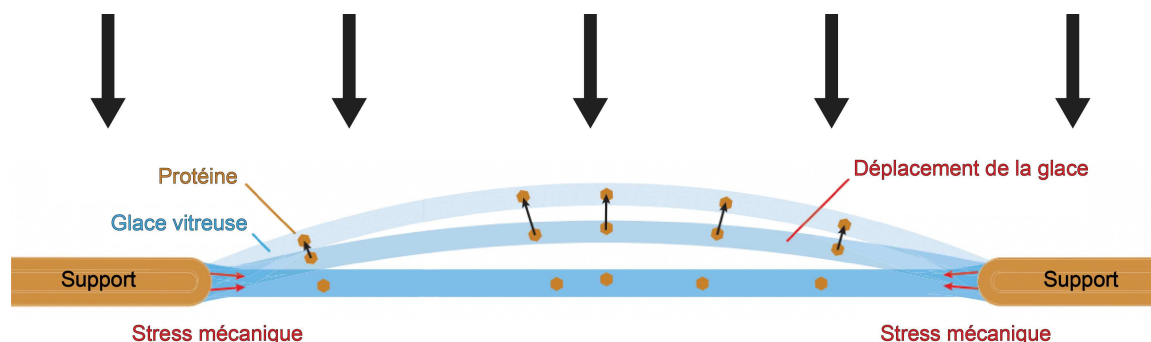


Figure 23. Mouvements induits par le faisceau d'électrons au contact de la grille et de l'échantillon. Le stress mécanique induit par le faisceau (flèches noires) sur le support de la grille et sur la fine couche de glace vitreuse entraîne le mouvement de nos particules et induit un flou dans nos images. Figure adaptée de Naydenova, Jia and Russo, 2020.

Afin de pallier les problèmes précédemment abordés, *i.e* le dommage aux radiations progressif et le mouvement des particules au cours de l'acquisition, différentes solutions ont été développées en termes de matériaux utilisés pour la construction des grilles ou au niveau des logiciels de traitements d'images. L'une d'elle consiste en l'utilisation de support ultra-stable, en or, permettant, par son haut pouvoir de conduction électrique, de limiter ces effets de charge au minimum notamment lors de l'exposition initiale au faisceau d'électrons qui induit une amplitude de mouvement beaucoup plus importante qu'après la relaxation du support de la grille et de la glace (Russo and Passmore, 2014). Ces grilles ont également l'avantage d'être construites entièrement dans le même matériau et permet lors de la congélation de limiter l'effet de dilatation/contraction entre la grille et le support à trous.

Au niveau du traitement de données, il est possible grâce au développement récent des détecteurs directs d'électrons (DDE), que nous aborderons dans la section suivante, de (i) supprimer les premières images de chaque film collecté, permettant ainsi de limiter l'effet de charge initial quantitativement supérieur au mouvement plus localisé et plus léger de fin d'exposition, (ii) de réaligner toutes les images entre elles afin de compenser le mouvement des particules pendant l'exposition mais également (iii) de compenser les dommages liés aux radiations en déterminant un filtre d'exposition permettant de diminuer la contribution des

dernières images qui ne disposent que d'informations de basse fréquence (Li *et al.*, 2013; Scheres, 2014; Grant and Grigorieff, 2015; Zheng *et al.*, 2017). Enfin, ils permettent également en segmentant les images en « sous images » de plus petites dimensions (iv) de déterminer et corriger le mouvement local de sous-groupes de particules au sein de chacune des images du film collecté (Zivanov, Nakane and Scheres, 2019).

Évidemment, le fait de supprimer les premières images des films collectés engendre une perte d'information importante car celles-ci possèdent l'information peu impactée par les dommages d'irradiations.

Récemment, le développement d'un nouveau type de grille en or à maille hexagonale a permis de diminuer l'effet de charge initial à la limite fixée par la pseudo-diffusion des particules dans la glace vitreuse au cours de l'exposition au faisceau d'électrons. Par conséquent, ce nouveau support laisse d'entrevoir une nouvelle avancée dans le domaine de la compréhension des mouvements induit par ces effets de charge lors de l'acquisition et pourrait permettre d'utiliser les images initiales collectées afin d'obtenir des reconstructions 3D à haute résolution ($<2 \text{ \AA}$) (Figure 24) (Naydenova, Jia and Russo, 2020).

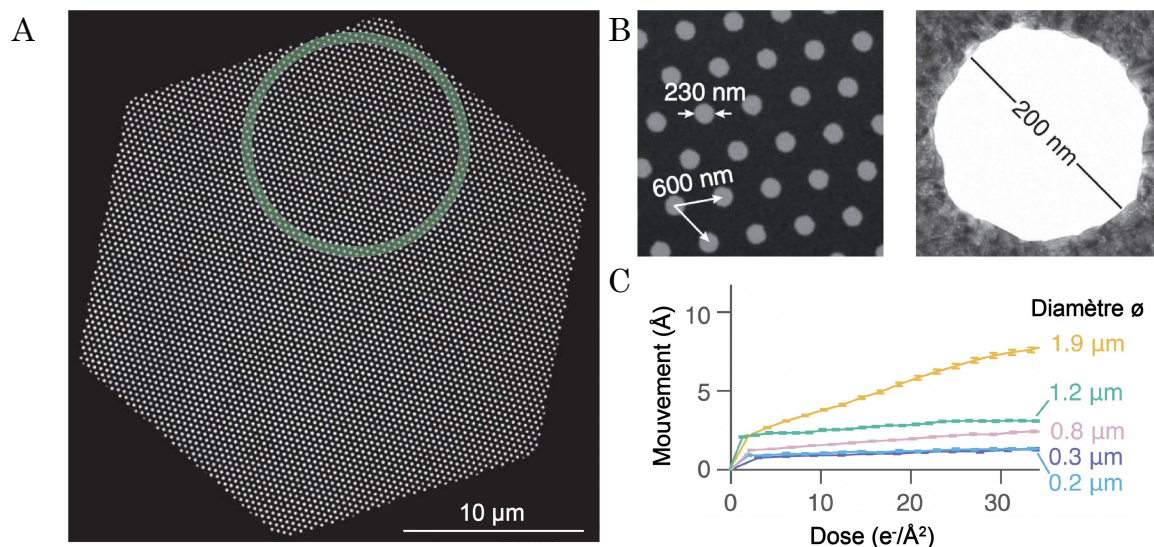


Figure 24. Nouvelle grille en or à maille hexagonale conçue pour la cryo-ME. **A**, Chaque grille de 3 mm de diamètre contient environ 800 hexagones. **B**, Chaque hexagone est subdivisé en plus de 5000 trous de 200 nm de diamètre. **C**, Le mouvement des particules en fonction de la dose reçue au cours de l'exposition est contenu à environ 1 \AA . Figure adaptée de Naydenova, Jia and Russo, 2020.

Des études extensives sont actuellement en cours pour voir l'effet de la géométrie des trous sur ces mouvements initiaux et essayer de les abolir.

4. Les récentes améliorations matérielles et logiciels en cryo-ME

4.1. Au niveau du MET

Les améliorations matérielles au niveau des MET se résument en 4 points principaux. Premièrement, l'arrivée des canons à émission de champs (FEG, X-FEG), qui a permis d'améliorer la cohérence spatiale et la brillance du faisceau d'électrons, *i.e* la dispersion en énergie des électrons est moindre, engendrant une augmentation substantielle du transfert d'informations aux hautes fréquences, en comparaison des précédentes sources utilisées comme le filament de tungstène ou le cristal LaB_6 (Figure 25) (Henning and Adhikari, 2017).

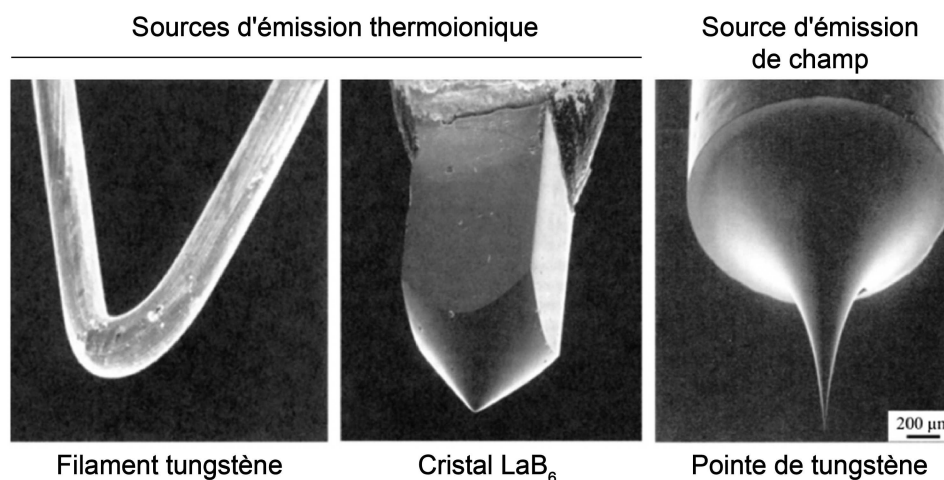


Figure 25. Sources d'émission thermoionique et à effet de champ utilisées en cryo-ME. La pointe FEG est recouverte de zirconium. Figure adaptée de Henning and Adhikari, 2017.

De par la possibilité de diminuer la taille du faisceau, les FEGs ont également permis d'accélérer les acquisitions de données en collectant de multiples trous par décalage du faisceau sans avoir à induire de déplacement physique (Cheng *et al.*, 2018).

Très récemment, un nouveau type de canon a été développé : le *cold*-FEG (CFEG) qui montre une cohérence encore accrue par un facteur 2 en comparaison des FEG/X-FEG classiques, dit de type Schottky (Hamaguchi *et al.*, 2019). Il en résulte un signal plus important aux hautes fréquences et donc une résolution

atteignable supérieure comme en atteste la structure de l'apoferritine obtenue à une résolution de 1.22 Å par Nakane *et al.* en 2020.

Deuxièmement, le développement de nouveaux systèmes de transfert de grilles, de porte-grilles et de *compustage* a permis de limiter les mouvements induits par les changements de position lors des collectes d'images.

Troisièmement, l'amélioration du vide interne au niveau de la colonne a permis de limiter les interférences que peut subir le faisceau d'électrons au contact d'éléments parasites et permettre de travailler plus longtemps avec le même échantillon sans dépôt excessif de contamination de glace à la surface de celui-ci.

Quatrièmement, l'augmentation de l'accélération des électrons a permis de limiter les effets de charges induits par les principes de diffusions précédemment abordés (de Jong and Van Dyck, 1993; Vinothkumar and Henderson, 2016).

NB : Sur ce point, il a été récemment démontré qu'en utilisant une accélération de 100 kV la diffusion élastique est 25% plus importante qu'à une accélération de 300 kV (les électrons ont en effet plus de temps pour interagir avec l'échantillon). De plus, du point de vue des dommages causés par le rayonnement, il a été démontré qu'il est préférable de collecter des images à 100 kV pour des échantillons d'une épaisseur inférieure à 600 Å. Malheureusement, l'élément technologique limitant actuel reste le développement d'un détecteur direct d'électrons dédié à cette accélération (Naydenova et al., 2019; Peet, Henderson and Russo, 2019).

Enfin, un autre développement important concerne la mise au point de lentilles à courant constant. Le courant total qui passe dans les lentilles ne varie pas entre les différents grossissements utilisés ce qui induit un équilibre thermique et une stabilité accrue.

4.2. Au niveau des filtres d'énergie

Afin d'améliorer le contraste de nos images, il est possible d'utiliser un filtre d'énergie permettant d'éliminer les électrons qui interagissent de manière inélastique avec notre échantillon et qui contribuent à la formation du bruit. Deux types de filtres sont utilisés actuellement : les filtres GIF QUANTUM et

CONTINUUM, développés par Gatan qui sont positionnés avant la caméra mais après la colonne (Gubbens *et al.*, 2010). Le filtre Oméga, développé par Jeol, est quant à lui positionné à l'intérieur de la colonne (Tanaka *et al.*, 1999). Récemment la société TFS a également mis sur le marché un nouveau filtre qui semble plus robuste et plus stable que son équivalent de Gatan : le filtre Selectris (et Selectris X).

4.3. Au niveau des logiciels et stratégies de collectes automatiques utilisées

Parallèlement à cela, la possibilité d'acquérir des images automatiquement grâce aux développements de nombreux logiciels tels que Leginon (Carragher *et al.*, 2000), SerialEM (Mastronarde, 2005) (initialement développé pour les collectes de données de cryo-tomographie) ou plus récemment EPU, développé en interne par la compagnie FEI/ThermoFisher, permet de limiter l'effort humain tout en augmentant drastiquement la quantité d'images collectées en comparaison des collectes réalisées manuellement (Schorb *et al.*, 2019). L'augmentation de la quantité de données collectées provient également des nouvelles stratégies de collecte (Cheng *et al.*, 2018).

L'une des avancées majeures consiste à incliner le faisceau *beam tilt/image shift* pour prendre des images dans plusieurs trous plutôt que de déplacer l'échantillon et changer de trou à chaque fois. Cela a pour effet d'augmenter le débit de collecte et permet d'augmenter la résolution des images en induisant une meilleure stabilité mécanique (Figure 26). Le problème principal de cette méthode réside dans l'introduction d'aberrations supplémentaires, comme le *tilt* et le *coma* par exemple, qui induisent une modification du retard de phase des électrons diffusés par rapport aux électrons non diffusés. Afin de corriger ces aberrations, il est possible de calibrer le microscope pour compenser le *tilt* et le *coma* (Wu *et al.*, 2019) et/ou de diviser nos films par groupes optiques, *i.e* en fonction de l'angle d'inclinaison utilisé pour acquérir un film, et d'effectuer ensuite et *in silico* une estimation puis une correction des aberrations optiques induites par l'inclinaison du faisceau (Zivanov *et al.*, 2018; Zivanov, Nakane and Scheres, 2020).

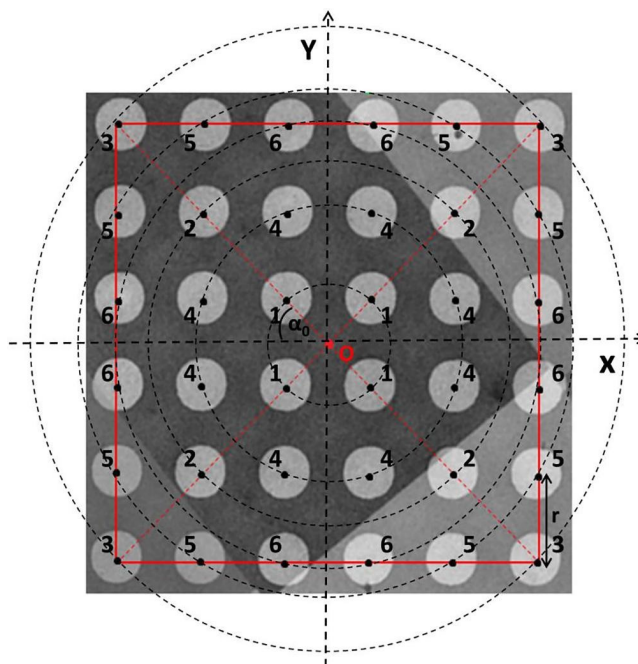


Figure 26. Stratégie de collecte automatisée en utilisant un motif 6×6 sur une grille de maille 1.2/1.3. Le *compustage* a été déplacé avant de lancer la collecte en position O. La collecte des données sur l'ensemble des 36 trous a été effectuée par *beam tilt/image shift*. Les différents groupes optiques sont numérotés de 1 à 6. Figure adaptée de Wu *et al.*, 2019.

4.4. Au niveau des caméras

La quantité d'images collectées a également augmenté en raison du développement de nouveaux détecteurs d'électrons directs (Faruqi *et al.*, 2005; Milazzo *et al.*, 2005; McMullan *et al.*, 2009). Historiquement, les images de MET étaient enregistrées sur des films photographiques d'halogénure d'argent disposant d'un large champ visuel mais nécessitant d'être développés puis numérisés, limitant drastiquement la quantité de données pouvant être collectée, typiquement 50-100 images par jour (Typke *et al.*, 2005). Afin de pallier ce problème, l'utilisation de caméra à dispositif de couplage (CCD) a permis d'obtenir des données numériques directes, augmentant ainsi la vitesse de collecte mais apportant de nouvelles limitations. Le principe des caméras CCD consiste en la conversion de charge des électrons incidents en photons à l'aide d'un écran scintillateur. L'image photonique obtenue est ensuite amplifiée puis transmise à travers un système de fibre optique au capteur CCD (ou à une couche semi-conductrice métal-oxyde CMOS) qui encode le signal numérique. Le problème de ces caméras, en dehors de leur champ visuel restreint, réside dans la perte

d'information générée lors de la conversion du signal électronique en signal photonique dû aux diffusions multiples induites par l'énergie des électrons qui traversent le scintillateur et la fibre optique. Ce principe de détection de signal diminue drastiquement l'efficacité quantique du détecteur (DQE), mesure reflétant la dégradation du rapport signal sur bruit à chaque fréquence spatiale durant la détection, notamment aux hautes fréquences mais également la rapidité avec laquelle l'information est transmise (Figure 27) (Fan and Ellisman, 2000; McMullan, Faruqi and Henderson, 2016; Koning, Koster and Sharp, 2018).

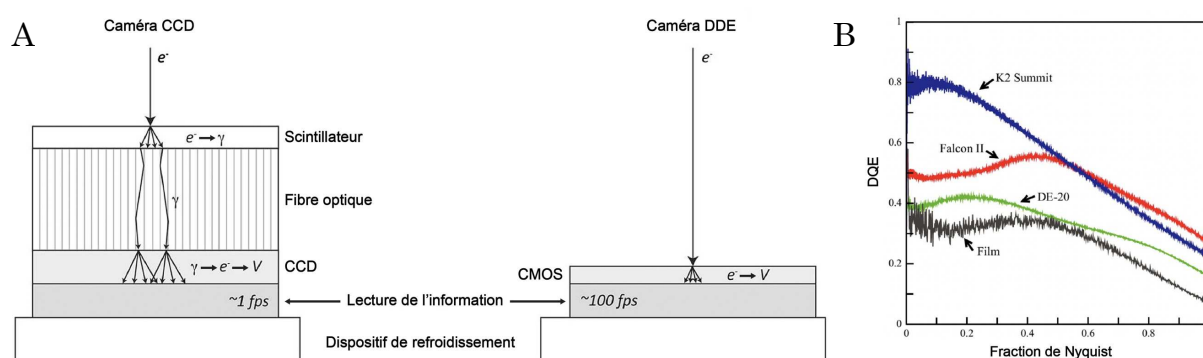


Figure 27. Composition générale des caméras à dispositif de couplage (CCD) et caméras à détection direct d'électrons (DDE) de type CMOS (*complementary metal oxide semiconductor*). **A**, Les caméras CCD détectent les électrons (e^-) via un scintillateur qui convertit les électrons en photons (γ). Cette lumière est transférée via une fibre optique dans une puce CCD qui convertit la lumière en tension sur une matrice de pixels. L'information est lue en quelques secondes et forme une image numérique. En revanche, un DDE CMOS convertit directement les e^- en une tension (V) sur une matrice de pixels lue à grande vitesse. **B**, Comparaison des DQE entre les films, les caméras CCD (DE-20), Falcon II (ThermoFisher) et K2 Summit (Gatan). Figure adaptée à partir de McMullan, Faruqi and Henderson, 2016 et Koning, Koster and Sharp, 2018.

Le développement des détecteurs d'électrons directs (DDE) a permis de pallier ces problèmes. En supprimant le scintillateur et le système de fibres optiques, la charge des électrons incidents est directement transmise à travers une fine couche CMOS couplée à de multiples convertisseurs analogique-numérique. Ce système permet de diminuer les effets de diffusion secondaire des électrons et de localiser chacun des électrons incidents en réduisant sa diffusion au pixel/transistor dont la charge reçue est la plus élevée. On parle d'acquisition en mode de « comptage ». La possibilité de déterminer avec précision chacun de ces événements uniques couplée à l'enregistrement de films à haute fréquence

d'images a largement contribué à l'amélioration du DQE aux hautes fréquences en comparaison des caméras CCD (Faruqi *et al.*, 2005; Milazzo *et al.*, 2005; McMullan *et al.*, 2009). Certains DDE comme la K2, la K3 (Gatan) et la Falcon 4 (ThermoFisher) permettent également « d'abolir » la limite de résolution théorique atteignable pour une taille de pixel donnée (susnommée la fréquence de Nyquist), qui est égale à deux fois la taille de pixel établie par l'opérateur en Å par pixel. Nommée « super-résolution », cette méthode d'imagerie tire parti du fait qu'un seul électron peut être simultanément détecté par plusieurs pixels adjacents. La charge électronique détectée peut donc être attribuée à un quadrant d'un pixel physique multipliant par quatre le nombre de pixels de l'image, par rapport aux pixels du détecteur physique pour la K2 /K3 (Feathers, Spoth and Fromme, 2021) voire par 16 pour la Falcon 4 (Figure 28).



Figure 28. Les différents modes de détection d'électrons. Lorsqu'un électron entre en contact avec la caméra, son signal est diffusé et la charge accumulée divisée entre plusieurs pixels. En mode standard (ou d'intégration), les détecteurs ne suppriment pas le bruit engendré par la diffusion du signal. En mode comptage (ou *counting*), la charge maximale est associée à un pixel. En mode super-résolution, la charge détectée est subdivisée par un facteur 4 dans l'exemple montré. Figure adaptée de <https://www.gatan.com/improving-dqe-counting-and-super-resolution>.

4.5. Conclusion

C'est donc grâce à la compréhension des différents principes physiques relatifs aux électrons, à l'optique et à l'eau acquis au cours du XX^{ème} siècle

qu'aujourd'hui, grâce aux récents développements qui touchent aux METs, aux détecteurs, aux logiciels et aux stratégies de collectes, il est possible de collecter des jeux de données de quantité et de qualité suffisante pour permettre de déterminer des structures de macromolécules biologiques par analyse de particules isolées. En plus de ces avancées majeures, les récents développements d'algorithmes et l'augmentation de la capacité de calcul des ordinateurs nous permettent désormais de déterminer une ou plusieurs reconstructions 3D à haute résolution.

Dans la prochaine section, nous nous intéresserons donc au traitement d'images appliqué à l'analyse de particules uniques en cryo-ME.

5. Le traitement d'images appliqué à l'analyse de particules uniques en cryo-ME

Le traitement d'images appliqué à l'analyse de particules uniques en cryo-ME est divisé en différentes étapes qui conduisent à l'obtention d'une ou plusieurs structures 3D des différents états représentatifs de notre objet d'intérêt. Les sections qui suivent décrivent chacune de ces étapes de manière synthétique en abordant les aspects généraux théoriques et pratiques. Elles abordent également les récents développements logiciels apparus au cours de ces dernières années.

5.1. La correction des mouvements induits par le faisceau d'électron

Une fois notre jeu de données collecté, la première étape consiste en la correction des mouvements induits par le faisceau d'électrons au contact de notre échantillon. Grâce aux nouveaux DDE, nous collectons des films composés de plusieurs images successives. Par corrélation croisée, il est possible d'aligner chacune des TF des images en les translatant à une origine commune afin de corriger les mouvements globaux puis, en subdivisant nos images en sous-parties (classiquement 5x5), de corriger les mouvements locaux de l'échantillon tout au long de l'exposition (Figure 29). Ce processus permet de réaligner chacune des images entre elles à une précision de l'ordre de l'angström. En sommant les images réalignées, on obtient une micrographie dont les mouvements ont été corrigés ce qui restaure l'information aux hautes fréquences précédemment perdue. De plus, afin de limiter les artefacts de réalignement des TF des images lors de corrélations croisées, la valeur des pixels voisins à un pixel mort sont moyennées et la référence de gain du détecteur est également appliquée (Brilot *et al.*, 2012; Li *et al.*, 2013; Zheng *et al.*, 2017).

Parallèlement, il est possible de compenser l'augmentation de la dose reçue par l'échantillon au cours de l'exposition en appliquant un filtre aux séries d'images collectées. Ce dernier permet de prendre en considération les variations d'amplitudes de la CTF au cours de l'exposition afin de donner un poids plus important aux premières images pour lesquelles la dose cumulée d'électrons est

faible et où de ce fait les informations à haute résolution sont présentes (Figure 29). Cette stratégie permet d'éviter de se séparer des dernières images collectées qui contribuent majoritairement à l'information aux basses fréquences, utile lors de l'estimation de la CTF et de l'alignement initial de nos particules. Ceci permet *in fine* d'augmenter le rapport signal sur bruit de nos micrographies (Grant and Grigorieff, 2015; Zheng *et al.*, 2017). Le mouvement induit par le faisceau d'électrons est toujours très important sur les premières images d'un film. Ces mouvements sont très difficiles à corriger et même si ces images contiennent le plus d'informations il est, dans l'état actuel des logiciels de corrections de mouvements, nécessaire de les éliminer. Classiquement nous supprimons la ou les deux premières images du film.

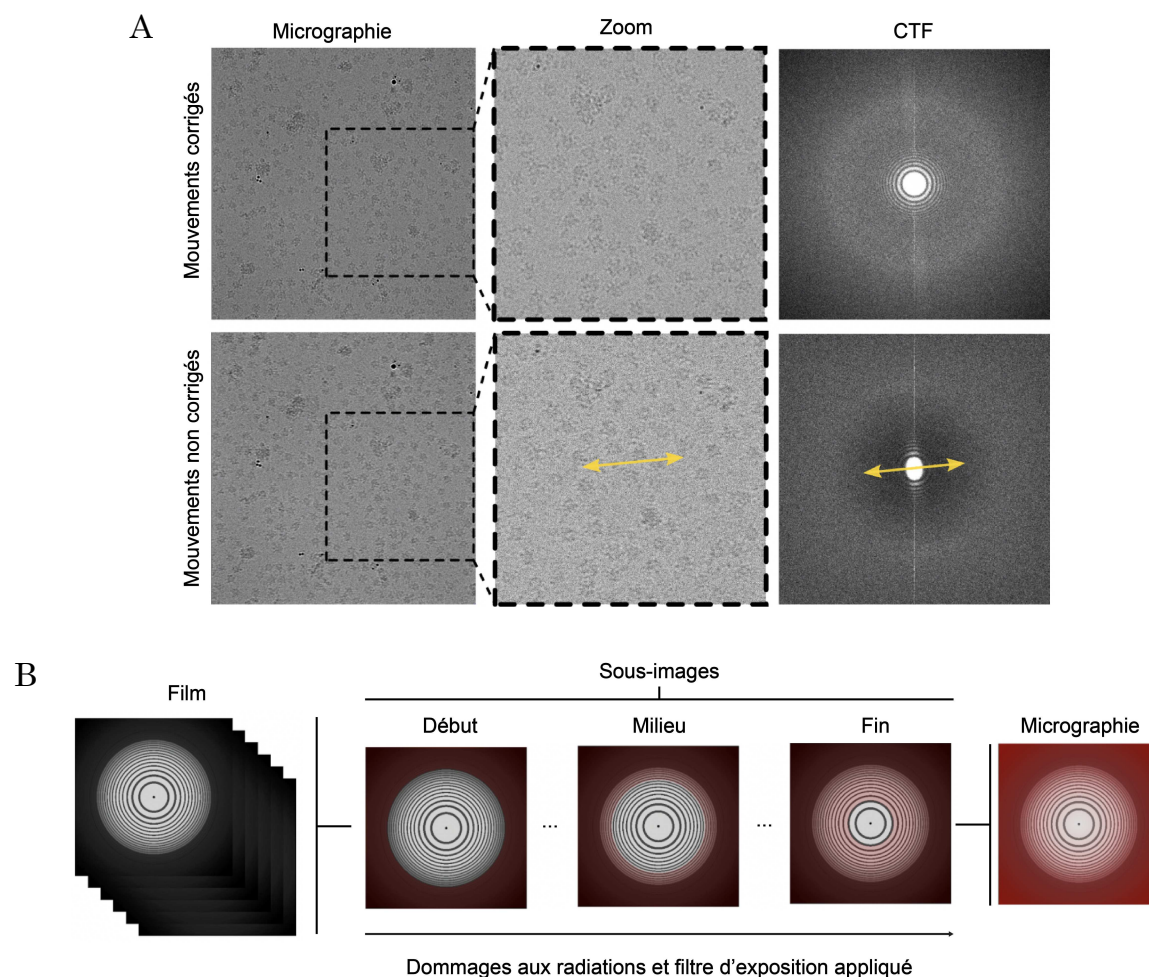


Figure 29. Mouvements induits par le faisceau et filtre d'exposition. **A**, Comparaison entre une micrographie dont les mouvements ont été corrigés ou non. **B**, Un filtre d'exposition est appliqué à nos sous-images. Les sous images de milieu et fin d'exposition présentent une forte réduction des composantes de haute fréquence (ombrage rouge). Les sous-images pondérées en fonction de la dose sont ensuite additionnées dans la micrographie finale.

5.2. L'estimation de la CTF

A partir des films réalignés et moyennés en une micrographie, on procède ensuite à l'estimation de la CTF. La CTF traduit les fluctuations de phases du signal en modulation d'amplitude suivant la fréquence. Les amplitudes sont de ce fait positives, nulles ou négatives suivant les fréquences, induisant la visualisation d'anneaux de Thon dans le spectre de puissance des images. La correction de la CTF a pour objectif de restaurer leur caractère positif (ou nul car si la CTF était nulle à une fréquence aucune restauration n'est possible). Une évaluation précise de la CTF est nécessaire car si cette dernière est mal évaluée, elle entraînera un décalage progressif de la correction de l'amplitude. Les oscillations de la CTF devenant de plus en plus rapprochées vers les hautes fréquences, l'évaluation sera non pertinente à haute fréquence limitant la résolution atteignable de notre future reconstruction 3D. La phase est également corrigée par 180° lorsque cela est nécessaire à cette étape. Comme abordé précédemment, la CTF dépend de l'accélération du faisceau d'électrons, des aberrations optiques du MET (distorsion du au grandissement, comatique, chromatique, astigmatisme, sphérique etc...) et de la défocalisation introduite par l'opérateur (Glaeser *et al.*, 2011).

En pratique, le coefficient des aberrations sphériques est directement fourni par le constructeur et l'accélération du faisceau d'électrons est connue. Il nous reste donc à déterminer la valeur de défocalisation qui, dans un système optique imparfait, dépend également de l'astigmatisme.

L'astigmatisme induit une variation des valeurs de défocalisation au sein de nos micrographies et se caractérise par la déformation de la CTF associée qui n'est plus isotrope. Les anneaux de Thon prennent une forme elliptique caractérisée par deux valeurs de défocalisations, une minimum et une maximum, ainsi qu'un angle qui détermine la direction de la déformation de la CTF. Généralement, l'opérateur corrige l'astigmatisme avant de lancer une collecte ce qui permet de limiter l'écart de défocalisation à une centaine d'angströms. L'astigmatisme peut être induit par des effets de charge et peut donc varier d'un endroit à l'autre d'une grille. Il est donc indispensable de le corriger *in silico* même si une correction a été faite en début de session de microscopie (Figure 30).

Les deux valeurs de défocalisation et l'angle déterminant la direction et la déformation de la CTF sont calculés pour chaque micrographie en effectuant une moyenne radiale de leur spectre de puissance qui est ensuite comparée à une CTF théorique 1D. Les valeurs obtenues sont ensuite affinées en fonction du spectre de puissance 2D de chaque micrographie. Cette méthode d'estimation de défocalisation est également applicable d'une manière analogue à la correction de mouvements locaux précédemment mentionnée : en subdivisant nos micrographies et à condition que le ratio signal sur bruit soit suffisant, il est possible de déterminer les paramètres de défocalisation locale voire même par particule (Zhang, 2016; Punjani *et al.*, 2017; Tegunov and Cramer, 2019).

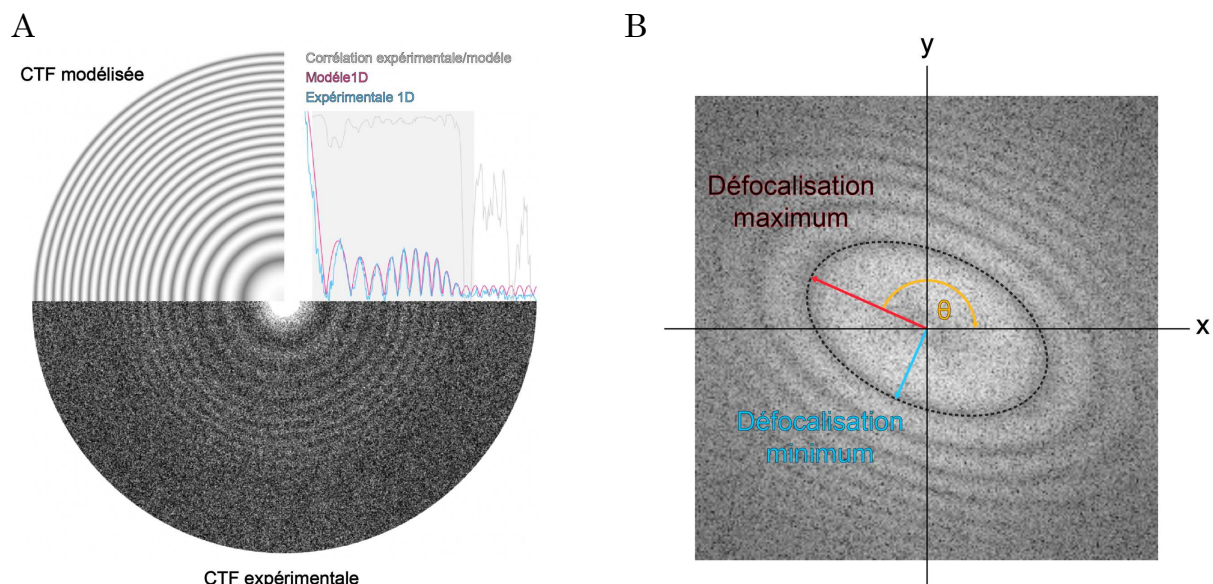


Figure 30. Estimation de la CTF et astigmatisme. **A**, A partir de la CTF expérimentale, une CTF est modélisée en estimant les paramètres de défocalisation. **B**, Les paramètres de défocalisation dépendent de l'astigmatisme, décrit par deux valeurs de défocalisation (minimum et maximum) et un angle θ . Figure adaptée de Tegunov and Cramer, 2019 et Orlova and Saibil, 2011.

5.3. La sélection, l'extraction et la classification 2D des particules

A cette étape, nous disposons de micrographies réalignées dont la CTF a été estimée. Une règle importante à appliquer en cryo-ME consiste à uniquement sélectionner les particules visibles. Si l'œil n'est pas capable de les discerner, que cela soit en raison d'un faible contraste et/ou d'une faible défocalisation, il est peu probable que les logiciels de traitement d'images puissent le faire à notre place et

encore moins d'assigner des orientations et translations précises lors de futures reconstructions ou classifications 3D.

Afin d'aider à la visualisation de nos particules, nous disposons de différents outils. L'un des plus basique consiste à appliquer un filtre passe-bas à nos micrographies qui permet de mieux distinguer les particules par rapport au bruit de fond (Figure 31). Lors de l'acquisition, il est également possible d'utiliser une plaque de phase Volta qui permet d'amplifier le contraste des images en introduisant un décalage de phase supplémentaire. L'utilisation de ce système s'applique majoritairement dans le cadre de l'étude de protéines dont la masse est inférieure à 100 kDa du fait de la complexité technique de l'expérience et de la modification au cours de la collecte du décalage de phase rendant complexe le traitement de l'information (Danev *et al.*, 2014; Khoshouei *et al.*, 2017).

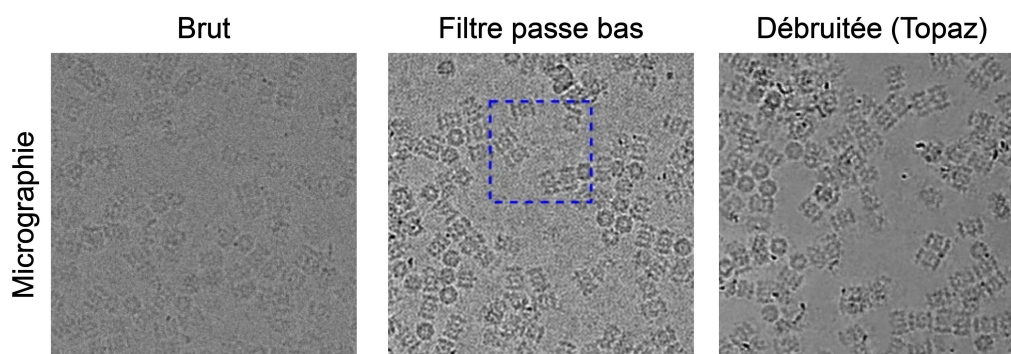


Figure 31. Différentes méthodes permettant de mieux visualiser les particules avant de procéder à leur sélection. Figure adaptée de Bepler *et al.*, 2019.

La sélection des particules est ensuite effectuée par corrélation croisée entre nos micrographies et un ou plusieurs modèles de référence. Il est possible d'utiliser des disques 2D de diamètre proportionnel à la taille de notre objet d'intérêt, des modèles basés sur une reconstruction 3D filtrée à basse résolution (50 Å) et projetée suivant un échantillonnage grossier ou des classes 2D obtenues au préalable. La combinaison de ces différentes méthodes fournit une base solide de visualisation et de sélection des particules même si le principe de corrélation croisée peut souffrir de la sélection de faux positifs comme des contaminations de glace, d'éthane ou les bords des trous des grilles de cryo-ME.

Le récent développement de réseaux neuronaux a permis de mieux visualiser nos particules en débruitant nos micrographies (Figure 31). De plus, en

les entraînant via la sélection manuelle ou automatique de quelques centaines de particules, il est possible de limiter la sélection de faux positifs (Bepler *et al.*, 2019, 2020; Tegunov and Cramer, 2019; Wagner *et al.*, 2019).

Une fois la sélection des particules effectuée, il est nécessaire de les extraire de nos micrographies. Comme vu précédemment, en espace réel nos images sont convoluées par une fonction d'étalement du point (PSF) qui, en espace réciproque, est associée à la multiplication de la TF de notre image par la CTF. En introduisant une défocalisation, l'information de la PSF contenue dans un pixel s'étale sur une large zone. Cet étalement est proportionnel à la défocalisation appliquée : plus on est proche du focus, moins la PSF s'étale (et inversement). Ceci implique que la taille de la boîte utilisée lors de l'extraction de nos particules doit être suffisamment grande pour limiter cette perte d'information, en règle générale de l'ordre de deux fois le diamètre de notre particule (van Heel *et al.*, 2000; Orlova and Saibil, 2011).

Enfin, lors de l'extraction, il est fondamental d'appliquer une échelle de gris commune à l'ensemble de nos particules. C'est ce qu'on appelle la normalisation. Elle est effectuée en utilisant les pixels se situant autour de nos particules afin d'ajuster la valeur du bruit localisé à l'arrière-plan de nos micrographies (Scheres, 2010).

Une fois nos particules extraites et normalisées, il est possible de réaliser des classifications 2D. Cette étape consiste à moyennner les particules extraites qui disposent d'orientations identiques par application de rotations et de translations. Ceci permet d'augmenter le signal sur bruit des particules par moyennation, de séparer les particules qui présentent plus ou moins de détails structuraux et de nettoyer notre jeu de donnée en séparant les vraies particules des faux positifs précédemment extraits. Pour cette thèse, l'approche Bayésienne utilisant le maximum de vraisemblance a été utilisée pour obtenir des classes 2D. Le but de l'optimisation par cette approche consiste en trouver les meilleures classes 2D qui expliquent le mieux les images observées en marginalisant sur l'assignation de la classe (2D), les rotations et les translations en 2D de chacune des images. Itérative,

cette méthode permet progressivement de grouper les particules qui ont des orientations similaires (Figure 32).

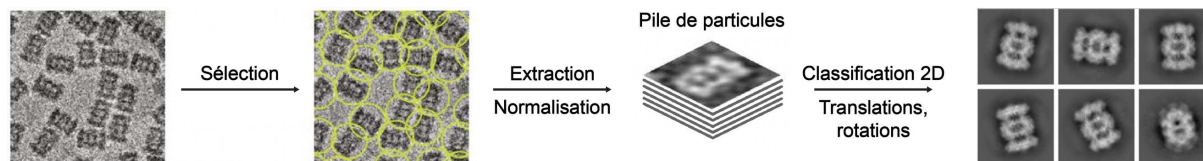


Figure 32. Une fois les mouvements corrigés et la CTF estimée pour chacune des micrographies, les particules sont sélectionnées, extraites puis le niveau de gris est normalisé. Les particules sont ensuite classifiées en 2D afin de supprimer les faux positifs précédemment sélectionnés ou les particules détériorées. Figure adaptée de Tegunov and Cramer, 2019.

5.4. La détermination d'une reconstruction 3D initiale

A l'heure actuelle, la génération d'une reconstruction 3D initiale est une étape relativement aisée et automatique, pour peu qu'une structure analogue à notre objet d'intérêt ne soit déjà disponible sur la banque de données des protéines (PDB). Il n'en n'a malheureusement pas toujours été ainsi... En effet, historiquement, l'une des principales difficultés rencontrées en cryo-ME consistait à déterminer de manière suffisamment précise les trois angles d'Euler (Φ, Θ, Ψ) et les deux translations (x, y) à appliquer à nos particules afin de calculer un modèle initial de notre objet d'intérêt (Figure 33).

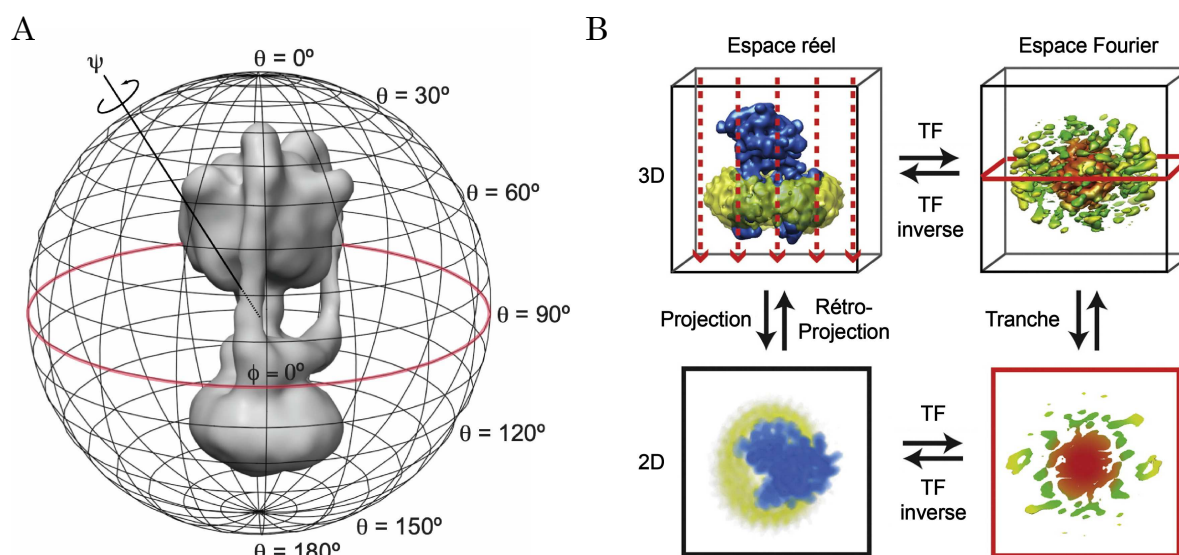


Figure 33. A, Les trois angles d'Euler Φ, Θ, Ψ à déterminer (en plus des deux translations x et y). B, Le théorème de la section centrale. Figure adaptée de Nogales and Scheres, 2015.

Trois méthodes principales étaient utilisées : l'approche des lignes communes (*common lines approach*) en espace réel ou dans l'espace réciproque (Crowther, Huxley and Klug, 1971; Van Heel, 1987) et la méthode d'inclinaison conique aléatoire (*random conical tilt*) (Radermacher *et al.*, 1987). Une fois la reconstruction 3D initiale obtenue, les orientations et les translations de chacune des particules peuvent être affinées itérativement par reprojektion des modèles successifs et affinements des orientations et translations assignées à chaque particule et cela afin d'obtenir une reconstruction 3D à plus haute résolution (Orlova and Saibil, 2011).

La TF d'une projection 2D d'un objet 3D correspond à un plan situé à l'origine appartenant à la TF 3D de ce même objet (théorème de la section centrale) (Figure 33). L'approche des lignes communes en espace réciproque se base sur ce principe. Avec seulement trois projections 2D différentes, ce qui équivaut à l'intersection entre les trois plans correspondants en espace réciproque et la formation de trois lignes communes, chacun des angles reliant ces trois plans peut être déterminés. Ainsi, en déterminant les angles relatifs entre de multiples TF de projections 2D, il est possible de recalculer une reconstruction 3D initiale par TF inverse. Cette approche est particulièrement efficace lorsque l'objet 3D étudié présente un haut degré de symétrie (Crowther, Huxley and Klug, 1971). Le principe des lignes communes en espace réel fait appel au même concept. Pour

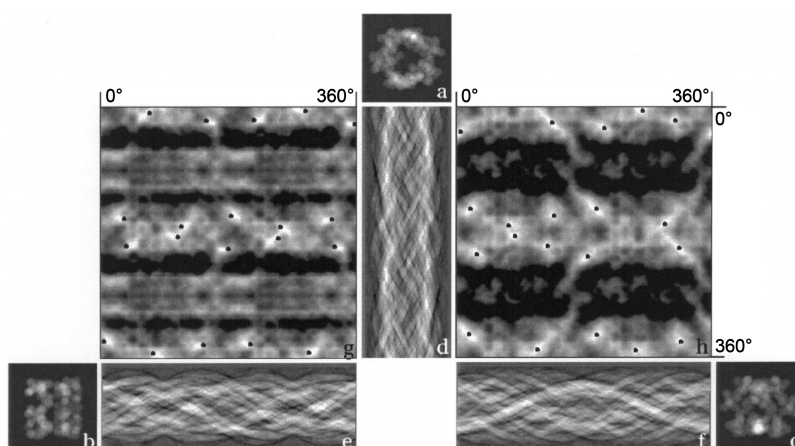


Figure 34. Fonctions de corrélation des sinogrammes. Deux sinogrammes calculés à partir de deux projections 2D d'un même objet 3D ont une ligne commune. Par corrélation de sinogrammes, il est possible de déterminer les angles d'Euler à appliquer aux projections pour recalculer la structure 3D de l'objet. Figure adaptée de Van Heel, 1987.

chaque projection 2D de notre objet, un ensemble de projection 1D est calculé formant des sinogrammes. Par corrélation croisée entre les différents sinogrammes, il est possible de déterminer les angles relatifs aux différentes projections 2D d'un objet 3D (Figure 34) (Van Heel, 1987).

La méthode d'inclinaison conique aléatoire consiste quant à elle à collecter plusieurs paires d'images inclinées à 0° et 60° de tilt (Figure 35). Chaque particule du jeu de données est sélectionnée sur l'image à 0° de tilt et sur l'image à 60° de tilt, la différence d'angle entre les deux particules est de ce fait connue. Les particules étant orientées de façon aléatoires, le principe de cette technique est de classifier en 2D les particules présentant la même orientation sur l'image à 0° en identifiant les rotations dans le plan qui sont nécessaires à leur superposition. Cette information de rotation supplémentaire entre particules à 0° combinée à l'information de 60° d'écart entre les particules permet d'assigner des angles d'Eulers précis aux particules collectées à 60° . Chacun des groupes issus des classes 2D obtenues à 0° tilt permet de recalculer une structure 3D. Comme plusieurs classes 2D sont obtenues à 0° tilt, plusieurs reconstructions initiales 3D en découlent. Chacune souffre d'artéfacts induits par le nombre limité de vues différentes mais par combinaison de toutes les reconstructions 3D obtenues il est cependant possible d'obtenir une reconstruction initiale 3D fiable bien qu'étant limitée en résolution. Pour pallier le manque de complétude de l'information des angles d'Euler, il est possible de collecter les images à -45° et 45° (Radermacher *et al.*, 1987). Cependant, on notera que cette méthode présente des difficultés :

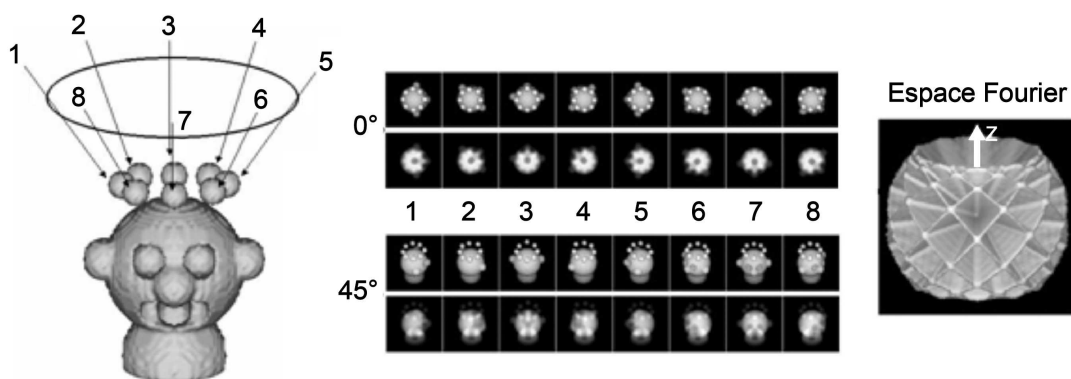


Figure 35. Principe général de la méthode d'inclinaison conique. En collectant des paires d'images à 0° et 45° d'inclinaison, les relations angulaires entre les différentes projections sont connues et permettent de remplir partiellement la TF 3D de notre objet. Figure adaptée du cours de Nicolas Boisset.

souvent effectuées en coloration négative pour avoir suffisamment de contraste, les images à -45 , 45 et 60° présentent des déformations (aplatissements) dus à la coloration et diminuant la qualité de la reconstruction obtenue.

Ces trois méthodes de reconstruction dépendent de l'assignation correcte des angles et des translations appliquées à nos projections 2D afin de générer la reconstruction initiale. Ces erreurs d'estimations peuvent malheureusement converger vers une solution erronée et des minimums locaux entraînant d'importantes complications lors de futurs affinements 3D (Henderson *et al.*, 2012; Henderson, 2013).

Afin de pallier ces problèmes, une nouvelle méthode de détermination de reconstruction 3D *ab initio*, basée sur un algorithme de descente du gradient stochastique (SGD), a été récemment développée et appliquée en cryo-ME afin de déterminer une ou plusieurs reconstructions 3D initiales. En pratique, le SGD va couvrir la totalité des possibilités de reconstructions 3D afin d'éviter de tomber dans des minimums locaux (Figure 36).

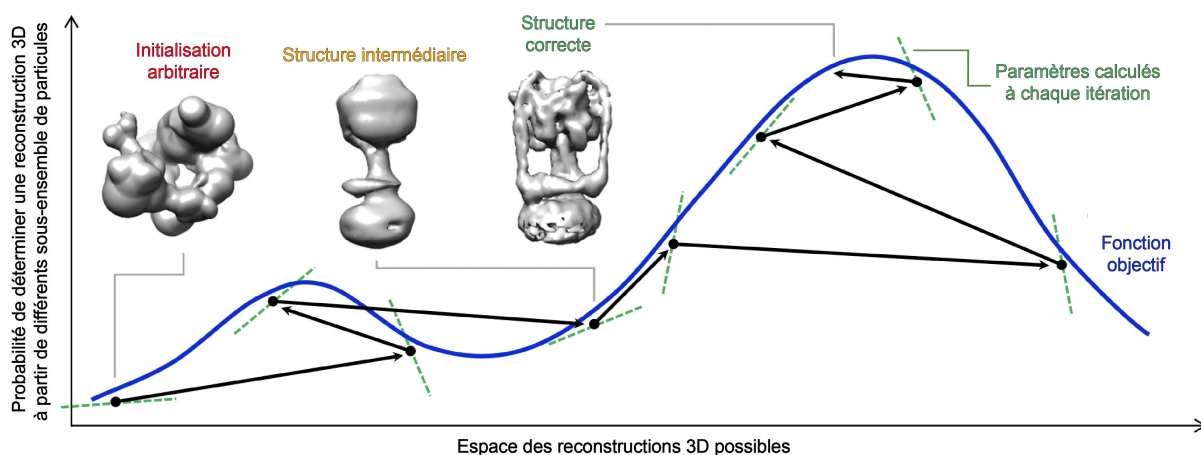


Figure 36. Principe général de détermination d'un modèle initial 3D *ab initio* basé sur un algorithme de descente du gradient stochastique (SGD). Figure adaptée de Punjani *et al.*, 2017.

Le SDG utilise une approche Bayésienne de maximum de vraisemblance. Le but de l'optimisation est de trouver la structure 3D qui explique le mieux les images observées en estimant une variable qui décrit la rotation 3D et la translation 2D de chacune des images. Ceci est décrit mathématiquement par une « fonction objectif ». Le SDG fait une approximation répétitive de cette fonction en

sélectionnant un sous-ensemble de particules à chaque itération et en calculant cette fonction pour ces images. Lors de la première itération, un modèle initial arbitraire est reconstruit à l'aide d'un sous-ensemble de particules dont des paramètres d'initialisation, *i.e* orientations et translations, sont assignés de manière aléatoire. A la 2nde itération, le sous ensemble de particules va évoluer et les variables d'optimisation sont mises à jour basé sur le gradient de la « fonction objectif » déterminé lors de l'itération précédente. Les modèles initiaux ainsi produits sont améliorés de manière itérative et font appels à de multiples sous-ensembles de particules. Ils peuvent ainsi converger vers une ou plusieurs solutions représentative(s) de notre jeu de données (Punjani *et al.*, 2017).

5.5. Le principe de l'affinement 3D d'un modèle initial

Plusieurs logiciels comme SPIDER (Shaikh *et al.*, 2008), EMAN (Tang *et al.*, 2007), FREALIGN (Grigorieff, 2007) ou IMAGIC (van Heel *et al.*, 1996) pour ne citer qu'eux, ont été développés au cours des années 2000 afin d'obtenir une reconstruction 3D affinée à partir d'un modèle initial et de projections 2D.

L'une des principales difficultés consistait à déterminer et améliorer de manière itérative les orientations et les translations de chacune des particules. En effet, les images collectées en cryo-ME sont caractérisées par la présence de bruit sur toutes les plages de fréquences existantes et notamment aux hautes fréquences (où ce bruit représente la majeure partie de l'information). Ce procédé pouvait donc être légèrement biaisé par l'ajustement du bruit aux caractéristiques structurales du modèle 3D initial au fur et à mesure des différentes itérations effectuées. En s'accumulant progressivement au sein de la reconstruction 3D, les artefacts incorporés influent directement sur la qualité de la reconstruction et engendrait une surestimation de sa résolution : c'est ce qu'on appelle l'*overfitting* (Stewart and Grigorieff, 2004; Chen *et al.*, 2013).

Plusieurs solutions ont été développées dans le but d'obtenir une reconstruction 3D dénuée d'artefacts et de déterminer un critère permettant d'estimer une valeur de résolution raisonnable (Henderson *et al.*, 2012). Grâce aux récents développements instrumentaux et d'algorithmes d'affinement 3D, qui basent désormais le calcul des orientations et des translations des particules sur

une approche statistique du maximum de vraisemblance, les reconstructions 3D obtenues sont quasiment exemptes d'*overfitting*. Aujourd'hui, deux logiciels, majoritairement utilisés par la communauté, emploient cette méthode de calcul : RELION (Sjors H.W. Scheres, 2012) et cryoSPARC (Punjani *et al.*, 2017). Parmi ces deux logiciels, RELION a été le premier logiciel à utiliser une approche bayésienne afin de déduire les différents paramètres d'un modèle statistique à partir des données expérimentales. Ainsi, l'étape de l'affinement 3D est définie comme la recherche du modèle mathématique qui a la plus grande probabilité d'être correct en fonction des données observées (Sjors H. W. Scheres, 2012a; Sjors H.W. Scheres, 2012b).

Initialement, afin d'éviter l'introduction d'artefacts structuraux dès les premières itérations d'un affinement 3D, un filtre passe-bas est appliqué à notre modèle initial. Ce dernier ne disposera alors que de caractéristiques structurales de basse résolution suffisantes pour déterminer et améliorer les orientations et les translations de nos particules. Parallèlement, les particules extraites sont aléatoirement divisées en deux groupes et chacun d'eux est utilisé pour affiner deux modèles initiaux distincts de manière totalement indépendante.

La première étape d'un affinement 3D consiste en la projection du modèle initial suivant un échantillonnage grossier d'orientations incluses dans l'unité asymétrique de la particule. Les TF de ces projections de référence sont alors comparées aux TF des images expérimentales afin de déterminer les orientations et les translations propres à chacune des particules (Figure 37). L'approche statistique du maximum de vraisemblance prend son sens à ce niveau-là : l'orientation dans l'espace d'une TF d'image expérimentale n'est pas assignée de manière discrète à une seule projection de référence mais est assignée à plusieurs projections du modèle initial selon des probabilités plus ou moins importantes. Les orientations multiples assignées à toutes les TF des projections expérimentales sont ensuite combinées dans l'espace réciproque afin de reconstruire la TF 3D de notre objet. Par TF inverse, on obtient alors une nouvelle reconstruction pour chacun des deux sous-ensembles de particules. Dans l'espace réciproque et à chaque itération, l'information des deux reconstructions obtenues est alors comparée par *Fourier Shell Correlation (FSC)* (Chen *et al.*, 2013). On appelle cela

le *gold-standard* car les deux reconstructions utilisées pour cette détermination de résolution sont indépendantes. On obtient une valeur de résolution représentative de la reconstruction 3D calculée à un seuil de corrélation égale à 0.143 (Rosenthal and Henderson, 2003).

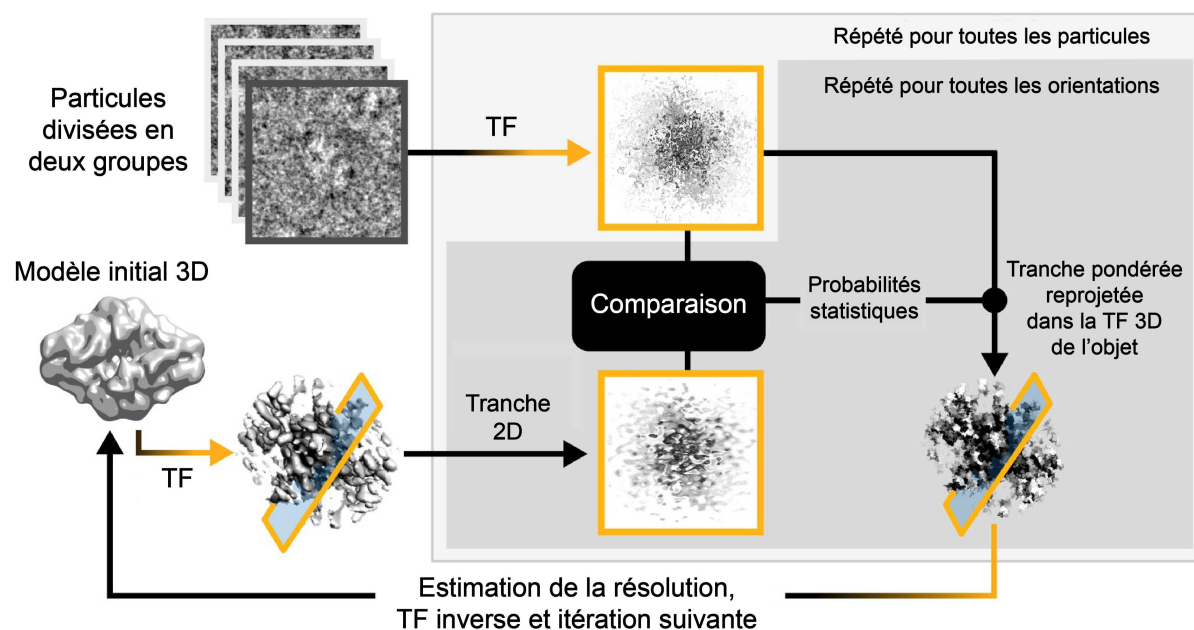


Figure 37. Principe général de l'affinement 3D d'un modèle initial préalablement déterminé. Figure adaptée de Kimanius *et al.*, 2016.

NB : La valeur de ce seuil a été déterminée comme équivalente à une FSC de 0.5 entre une reconstruction calculée à partir de toutes les données et un modèle de référence parfait. C'est une valeur de résolution globale et raisonnable afin d'estimer la qualité d'une reconstruction 3D obtenue par cryo-ME (Rosenthal and Henderson, 2003).

A la prochaine itération, les deux reconstructions seront filtrées à cette résolution et les probabilités préalablement déterminées des orientations et des translations seront recalculées et mises à jour. L'échantillonnage des orientations sera quant à lui automatiquement augmenté durant l'affinement 3D afin d'améliorer la couverture de l'information en espace réciproque (Crowther, DeRosier and Klug, 1970). Itérativement, en incluant peu à peu les informations contenues aux moyennes et hautes fréquences tout en augmentant le taux d'échantillonnage de nos orientations, la qualité de la reconstruction 3D

s'améliorera jusqu'à atteindre la limite de résolution des données. A cet instant, l'affinement des orientations et des translations de nos particules ne sera plus possible car l'information aux hautes fréquences sera équivalente au bruit présent.

Lors de la dernière itération, la comparaison des deux reconstructions 3D affinées de manière indépendante (*half maps*) est effectuée par FSC et la résolution globale de la reconstruction correspond à la valeur reportée lorsque leur corrélation atteint 0.143. Les deux *half-maps* sont alors fusionnées et la reconstruction résultante est filtrée à la résolution précédemment déterminée. Afin de prendre en compte les variations de résolution au sein d'une reconstruction 3D (parties plus rigides vs parties plus flexibles), il est également possible de calculer une multitude de FSCs en scannant notre reconstruction par groupe de voxels. En estimant la résolution locale de cette dernière, il est possible de la filtrer en conséquence afin d'obtenir une reconstruction plus homogène (Cardone, Heymann and Steven, 2013; Kucukelbir, Sigworth and Tagare, 2014).

Enfin, pour caractériser l'effet de l'*overfitting* sur notre reconstruction 3D finale, il est nécessaire de comparer la FSC originale propre à nos données avec la FSC dont les phases des images ont été aléatoirement modifiées aux hautes résolutions. Cette comparaison nous permet de déterminer si l'information aux hautes fréquences contenue dans notre reconstruction est réelle ou biaisée par l'introduction d'artéfacts provenant du bruit ou d'un masque par exemple (Scheres and Chen, 2012; Chen *et al.*, 2013; Rosenthal and Rubinstein, 2015).

5.6. Les différentes options de classification 3D

5.6.1. Comment séparer des états discrets au sein de notre jeu de données ?

En solution, notre protéine d'intérêt purifiée peut être hétérogène, flexible ou subir de multiples changements de conformations afin de remplir différentes fonctions biologiques. En cryo-ME, les protéines observées sont figées car congelées dans une fine couche de glace vitreuse. Il est donc important de séparer les différentes conformations existantes et d'éliminer de potentielles particules détériorées afin de nettoyer notre jeu de données. Dans ce but, il est possible de

réaliser une ou plusieurs étapes de classification 3D dont l'objectif va évidemment dépendre de l'objet étudié et de ce que l'on cherche à classifier. Au cours de cette thèse, l'approche Bayésienne utilisant le maximum de vraisemblance a été utilisée pour obtenir des classes 3D. La théorie derrière son fonctionnement est similaire à ce qui est utilisé lors de classifications 2D, mais cette fois-ci le but est de déterminer différentes reconstructions 3D qui expliquent au mieux les images observées.

A la différence d'un affinement 3D, au cours duquel les données sont initialement divisées en deux et l'échantillonnage des orientations est automatiquement augmenté en fonction de la résolution estimée, la classification 3D permet d'affiner simultanément de multiples structures en divisant aléatoirement le nombre de particules en autant de classes paramétrées par l'utilisateur. De ce fait, l'application de l'approche *gold-standard* abordée dans la section précédente est impossible. Par conséquent, la résolution de chacune des reconstructions résultantes de classification 3D est estimée directement à partir de leur TF 3D pouvant conduire à un problème d'*overfitting* (S.H.W. Scheres, 2016). De plus, au cours de ce processus itératif, l'échantillonnage des orientations est fixe ce qui a pour effet de limiter la résolution atteignable de chacune des classes et le pouvoir de séparation des différentes conformations existantes. Généralement, les premières classifications 3D sont donc réalisées sans appliquer de masque à nos modèles initiaux et l'échantillonnage d'orientations est grossier afin de séparer les différences de conformations majeures affectant un complexe mais également de séparer les particules détériorées ou adjacentes les unes aux d'autres.

Après avoir sélectionné et regroupé des sous-ensembles de particules présentant une conformation similaire, il est possible d'effectuer une classification 3D en utilisant un échantillonnage d'orientation plus fin ou basé sur une recherche d'angle locale, à partir d'orientations déterminées lors d'un précédent affinement 3D, ce qui nous permet de séparer les particules en fonction de variations structurales plus fines.

Malheureusement, il se peut qu'un ou plusieurs domaines d'intérêts présentent une flexibilité supplémentaire nous empêchant d'obtenir sa structure à haute résolution. Afin de focaliser le processus de classification 3D sur ce domaine, il est possible d'appliquer un masque autour de celui-ci dans le but de subdiviser son mouvement en plusieurs classes. Ceci peut dans certains cas permettre de séparer une conformation stable du domaine, qui étant peu représentée au sein du jeu de données entier ne pouvait être clairement visualisée.

A partir d'orientations déterminées via un affinement 3D global, il est également possible de masquer notre domaine d'intérêt, soustraire toute l'information localisée en dehors de celui-ci et recentrer nos particules sur le centre de masse du masque (Figure 38). Dans le cas de l'étude d'un domaine de petite

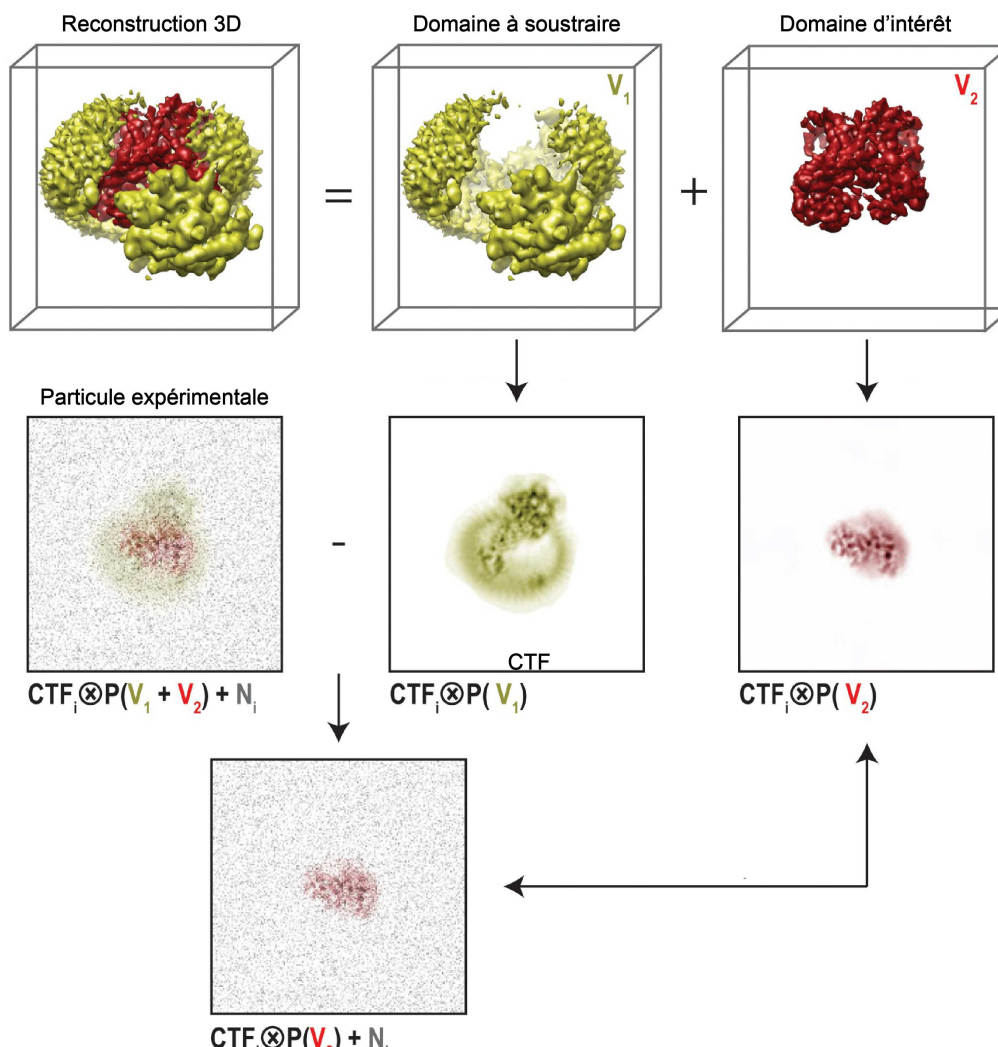


Figure 38. Principe général de la soustraction de signal d'un domaine d'intérêt à partir d'une reconstruction 3D. Figure adaptée de S. H. W. Scheres, 2016

taille, les particules soustraites peuvent être utilisées afin d'effectuer une classification 3D sans alignement ou avec une recherche locale d'angles. Suivant le résultat, il est alors possible de sélectionner un sous-ensemble homogène de particules soustraites afin d'effectuer un affinement 3D dans le but d'améliorer la résolution de notre domaine (S. H. W. Scheres, 2016).

La soustraction couplée à la classification 3D est une combinaison efficace afin de séparer des conformations discrètes d'un ou plusieurs domaines appartenant à notre protéine. Cependant, dans la majorité des cas, la flexibilité d'un domaine est de nature continue ce qui implique la présence d'une multitude de sous populations.

5.6.2. Comment gérer un problème de flexibilité continue ?

Les différentes méthodes de classifications 3D et de soustraction précédemment décrites sont majoritairement utilisées dans le but d'isoler un ou plusieurs sous-ensembles de particules présentes dans un état stable afin de réaliser un affinement 3D permettant d'obtenir une structure à haute résolution dans laquelle il est possible de construire un modèle atomique.

Cependant, il se peut que plusieurs domaines au sein de complexes macromoléculaires présentent un mouvement continu qui peut se révéler critique afin de caractériser la fonction d'une protéine. Dans ce but, plusieurs méthodes ont été récemment développées comme l'affinement 3D multi-domaines (*multi-body refinement*) dans RELION (Nakane *et al.*, 2018) ou l'analyse de la variabilité 3D (*3D variability analysis*) dans cryoSPARC (Punjani and Fleet, 2021). L'affinement 3D multi-domaines est particulièrement efficace en ce qui concerne l'étude de protéine disposant de domaines se déplaçant indépendamment les uns des autres. En soustrayant chacun des domaines à chaque nouvelle itération, il est possible de déterminer et d'améliorer les orientations relatives entre chacun d'eux.

Il est également possible d'effectuer une analyse des différentes composantes principales (*principal component analysis, PCA*) afin de décrire des mouvements 3D continus au sein d'un jeu de données. Le principe général de la PCA consiste à transformer un ensemble d'observations éventuellement corrélées en un ensemble de variables linéairement indépendantes, chacune de ces nouvelles

variables est appelée composante principale. Cette méthode permet de « résumer » les différences/similarités d'un ensemble de particules par détermination d'une matrice de covariance et représentation des vecteurs/valeurs propres (*eigenvectors/ eigenvalues*) correspondant aux différents changements de conformations d'une protéine ou à la fixation d'un ligand par exemple. A la différence des méthodes de classification multi référence et faisant appel au maximum de vraisemblance, qui exigent que le nombre de classes soit fixé avant l'analyse des données, les méthodes basées sur la PCA n'imposent pas de limitation arbitraire initiale sur ce point. En reconstruisant chacun des volumes 3D le long de ces vecteurs propres, il est alors possible de créer plusieurs films dépeignant les mouvements principaux entre les différents domaines d'une protéine (Nakane *et al.*, 2018). L'analyse de la variabilité 3D dans cryoSPARC fait également appel à la PCA et permet en sus de sélectionner les particules appartenant à telle ou telle composante principale afin de procéder à des affinements 3D ultérieurs (Serna, 2019; Punjani and Fleet, 2021).

5.7. Post-processing et sharpening

Les reconstructions 3D obtenues en cryo-ME souffrent d'une diminution constante des amplitudes en fonction de l'augmentation des fréquences. Cet effet est induit par de multiples facteurs inhérents aux propriétés du microscope et du détecteur utilisé mais également des différentes étapes réalisées lors du traitement d'image qui font majoritairement appel au signal des basses et moyennes fréquences. Cette diminution de contraste induite par la somme de ces facteurs peut être visualisée par une représentation de Guinier du logarithme de la moyenne rotationnelle des amplitudes de notre reconstruction 3D en fonction de la résolution (Rosenthal and Henderson, 2003).

Cette représentation se divise en deux zones distinctes. La première est localisée entre les basses et les moyennes fréquences. Elle se caractérise par une diminution rapide des amplitudes qui représente l'information de basse résolution de notre objet d'intérêt ($> 10 \text{ \AA}$). La deuxième zone est localisée entre les moyennes et hautes fréquences. Elle correspond au signal des structures secondaires et des

acides aminés au sein de la protéine et elle est caractérisée par une diminution progressive et linéaire des amplitudes ($< 10 \text{ \AA}$).

Afin de pallier cette diminution d'amplitude tout en évitant d'amplifier le bruit présent aux hautes fréquences, il est possible de déterminer un facteur de température négatif (\AA^2), ou facteur B, à partir d'une courbe de diffusion théorique. En appliquant ce facteur B à notre reconstruction 3D afin de normaliser son signal, la diminution des amplitudes est rendue linéaire sur une plage de résolution plus grande. En comparaison d'une reconstruction 3D non post-traitée, cette normalisation permet la restauration de détails structuraux précédemment cachés par les signaux de basse fréquence tout en évitant l'amplification majeure du bruit de fond (Figure 39) (Rosenthal and Henderson, 2003).

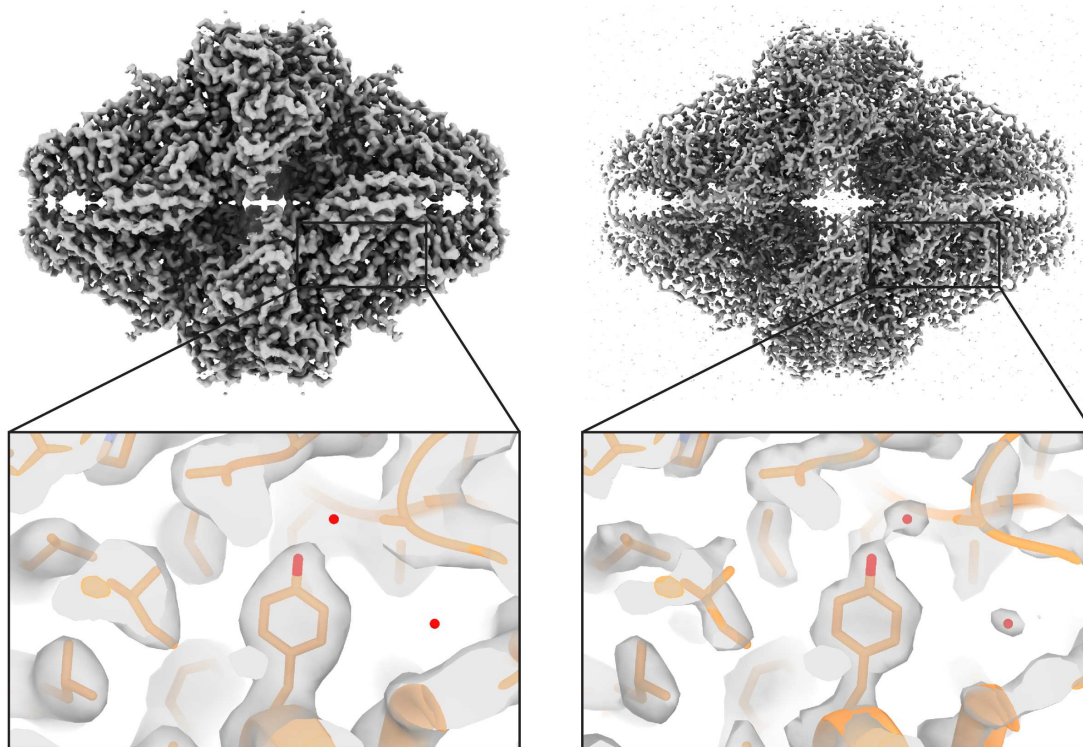


Figure 39. Reconstruction 3D de la β -galactosidase à 2.2 Å de résolution avant *sharpening* (à gauche) et après (à droite). Les détails structuraux précédemment cachés par les signaux de basse fréquence sont désormais visibles à l'image des molécules d'eau (EMDB-2984 ; PDB : 5A1A).

L'application d'un facteur B unique à une reconstruction 3D est valable dans la majorité des cas et notamment dans le cas où notre protéine d'intérêt ne dispose pas d'une flexibilité intrinsèque importante. Dans le cas contraire, il est possible de déterminer et d'appliquer des facteurs B locaux en tenant compte des

différences de résolution locale au sein de notre protéine (Jakobi, Wilmanns and Sachse, 2017; Ramírez-Aportela *et al.*, 2020; Sanchez-Garcia *et al.*, 2020; Kaur *et al.*, 2021).

Une fois notre reconstruction 3D finale obtenue, il est alors possible de construire un modèle 3D de notre protéine en utilisant un logiciel dédié (Emsley and Cowtan, 2004) et de procéder à l'affinement de sa géométrie dans l'espace direct (Afonine *et al.*, 2018).

5.8. Le cas particulier des objets hélicoïdaux

Par rapport au traitement d'images de particules isolées précédemment décrit, les structures hélicoïdales présentent des particularités dans leur traitement. Elles représentent également les premières structures 3D biologiques à avoir été déterminée par ME (De Rosier and Klug, 1968). L'avantage de tels objets hélicoïdaux réside dans leur composition et leur arrangement. Ils disposent de multiples sous-unités identiques et répétées de manière périodique le long d'un axe unique z mais dans des orientations différentes. Afin de déterminer la structure 3D d'un objet hélicoïdal, il est fondamental de retrouver les deux paramètres de symétrie qui le définissent : (i) la translation le long de l'axe z afin de passer d'une sous unité à une autre, appelée *axial rise* (ou p) et (ii) la translation le long de l'axe z afin d'effectuer une rotation de 360° , appelée *pitch* (ou P).

A partir de ces deux paramètres, il est possible de déterminer le nombre de sous-unités par tour (u/t) en divisant la valeur du *pitch* par l'*axial rise* ($P/p = u/t$) et de déterminer l'angle azimutal à appliquer, relatif à l'axe z , afin de passer d'une sous unité u à u_{+1} dans le cadre d'une hélice droite, ou u_{-1} dans le cadre d'une hélice gauche. La valeur de cet angle est appelée *helical twist* ($360^\circ/u/t$). La détermination de ces paramètres hélicoïdaux est effectuée dans l'espace réciproque et se base sur la description mathématique de la TF d'un objet hélicoïdal établie par Cochran, Crick et Vand en 1952, généralisée par Klug, Crick et Wyckoff en 1958 puis Ramachandran et Venkatachalam en 1970.

5.8.1. La transformée de Fourier d'une hélice discontinue

La projection 2D d'une hélice continue correspond à une fonction cosinus de période P et de rayon r . En espace réciproque, sa TF prend la forme d'une croix centrée sur l'origine constituée de *layer lines* perpendiculaires à l'axe de l'hélice z et espacées de $1/P$. Chacune des *layer line* est caractérisée par une fonction de Bessel d'ordre n relative aux harmoniques cylindriques de notre hélice. Plus n augmente et plus l'amplitude maximale de la fonction est éloignée de l'axe méridien de la TF.

Dans espace réel, lorsque l'on multiplie une hélice continue avec une fonction d'interférence caractérisée par une distance périodique égale à p , relative à sa distance en espace réciproque $1/p$, on obtient une hélice discontinue (Figure 40).

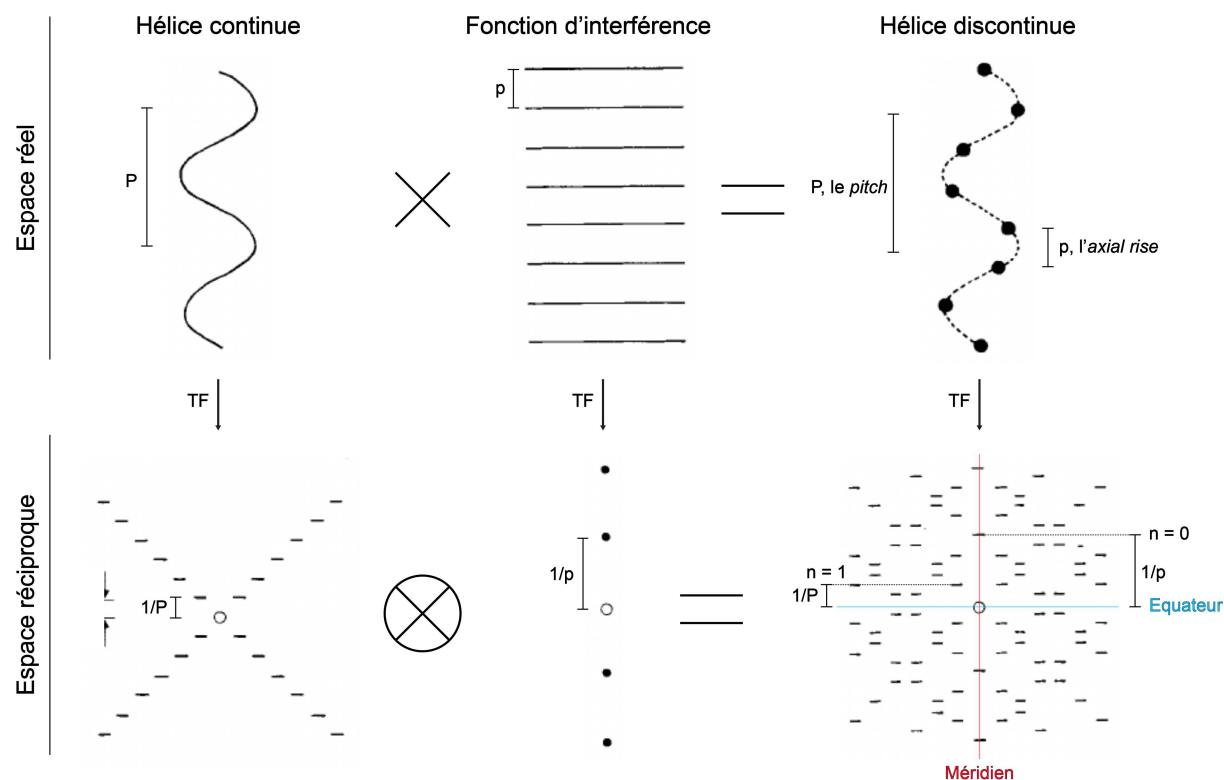


Figure 40. Les différentes fonctions en espace réel qui composent une hélice discontinue et leur TF correspondantes en espace réciproque. Figure adaptée de Stewart, 1988.

La TF d'une hélice discontinue correspond donc à la convolution des TF de l'hélice continue et de la fonction d'interférence. Cette convolution équivaut à la

répétition de la TF de l'hélice continue en chacun des points de la TF de la fonction d'interférence (Figure 40).

A partir de la TF d'une hélice discontinue, il est possible de déterminer chacun des paramètres qui la définit : l'*axial rise* et le *pitch* qui permettent de déduire le nombre de sous-unités par tour ainsi que l'*helical twist* comme décrit précédemment (Figure 40) (Stewart, 1988; Diaz, Rice and Stokes, 2010).

5.8.2. L'indexation en espace réciproque

La méthode principale permettant de déterminer la symétrie d'une hélice consiste en l'indexation des deux paramètres précédemment décrits à partir de la TF de notre objet hélicoïdal : (i) l'*axial rise* qui est égal à la distance $1/p$ entre l'équateur et la *layer line* d'ordre de Bessel $n = 0$ localisée sur le méridien et (ii) le *pitch* qui est égal à la distance $1/P$ entre l'équateur et la *layer line* d'ordre de Bessel $n = 1$ la plus proche du méridien.

En pratique, pour permettre l'indexation de ces paramètres de symétrie, il est nécessaire d'obtenir un spectre de puissance de notre objet biologique hélicoïdal. La première étape consiste en la segmentation de notre hélice afin que l'information structurale de chaque sous-unité soit prise en compte lors du traitement d'images. La segmentation initiale est arbitraire car la symétrie n'est pas connue mais dans l'absolu, elle doit être d'une distance égale ou un multiple de la valeur de l'*axial rise*. Une fois l'extraction effectuée, les segments sont soumis à une étape de classification 2D permettant de les aligner et de les moyennner.

Pour chacune des meilleures classes 2D, la taille initiale des boites est artificiellement augmentée deux à trois fois ce qui a pour effet d'améliorer l'échantillonnage des distances en espace réciproque. Les TF correspondantes à ces classes 2D sont alors additionnées afin de procéder à l'indexation. Si vous détectez une forte intensité périodique localisée sur l'axe méridien, c'est un candidat potentiel pour la valeur de l'*axial rise*... Si vous détectez une forte intensité localisée proche du méridien, c'est un candidat potentiel qui vous donnera la valeur du *pitch*. Ce processus peut être long et laborieux et les paramètres estimés sont généralement peu précis notamment en raison de l'inclinaison des filaments au

sein de la glace ou des différentes ambiguïtés de symétries existantes (Desfosses, 2012; Egelman, 2014).

Récemment (et en lieu et place d'utiliser une règle, une feuille et un crayon), HELIXPLORER a été développé afin d'aider à l'indexation des paramètres hélicoïdaux. En important la somme des spectres de puissance expérimentaux d'une hélice, il est possible de le superposer à un spectre théorique paramétrable afin d'explorer les différentes symétries possibles (<http://rico.ibs.fr/helixplorer/>).

5.8.3. Reconstruction 3D hélicoïdale

L'unique moyen permettant de valider des paramètres hélicoïdaux établis par indexation consiste en l'obtention d'une reconstruction 3D à haute résolution au sein de laquelle les chaînes latérales sont visibles. Tout comme en analyse de particules isolées, plusieurs logiciels comme SPIDER (Frank *et al.*, 1996), SPARX (Behrmann *et al.*, 2012), FREALIX (Rohou and Grigorieff, 2014) ou SPRING (Desfosses *et al.*, 2014) ont été développés afin de recalculer une structure 3D d'un objet biologique hélicoïdal. Aujourd'hui, Relion et cryoSPARC ont également implémentés des algorithmes équivalents mais couplés à une approche statistique du maximum de vraisemblance (He and Scheres, 2017; Punjani *et al.*, 2017).

De manière générale, l'affinement 3D d'une structure hélicoïdale est semblable au principe d'affinement 3D appliqué aux particules isolées à quelques exceptions près. Premièrement, il est possible d'utiliser un cylindre de diamètre équivalent à celui estimé à partir de classes 2D comme reconstruction 3D initiale (A noter cependant que cette méthode ne fonctionne pas toujours dans le cas d'hélices présentant peu de protubérances car un cylindre peut ne pas suffire à une assignation correcte des angles et translations lors de l'affinement). Deuxièmement, en ce qui concerne l'utilisation de segments hélicoïdaux, la division initiale aléatoire du jeu de données est filament-dépendant afin d'éviter d'induire un chevauchement du signal de deux segments voisins entre les affinements 3D de chaque sous-ensemble pouvant conduire à un problème d'*overfitting*. Troisièmement, du fait de la nature hélicoïdale de notre objet supposément rigide et linéaire, des restrictions préalables peuvent être imposées en ce qui concerne les recherches d'angles d'inclinaison et de rotation. Les translations sont également

limitées à la distance d'un *axial rise* dans la direction de l'axe hélicoïdal afin de limiter les problèmes d'*overfitting* lors de l'étape de symétrisation (Figure 41) (He and Scheres, 2017).

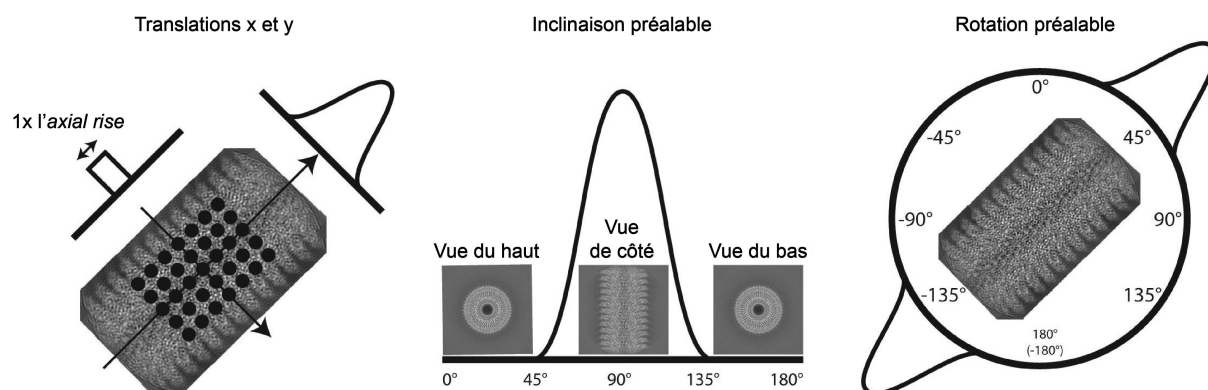


Figure 41. Limites des recherches de translations et d'inclinaison d'une hélice au sein d'une fine couche de glace vitreuse. Figure adaptée de He and Scheres, 2017.

Enfin, dans le but de compenser les erreurs pouvant être réalisées pendant l'indexation, il est possible de donner un intervalle et un incrément pour chaque paramètre de symétrie estimés : l'*axial rise* et l'*helical twist*. Cette recherche locale des paramètres hélicoïdaux est basée sur une méthode de reconstruction hélicoïdale itérative en espace réel (IHRSR) initialement développée par Egelman (Egelman, 2000, 2007) puis intégrée à Relion (He and Scheres, 2017).

Objectifs de thèse

Chez les *Bunyavirales*, les polymérasés viraux (L) sont liées aux régions promotrices 5' et 3' des ARNv induisant la circularisation de chacun des trois segments présents au sein des virus. Le génome libre est quant à lui recouvert par de multiples NPs formant des assemblages hélicoïdaux appelés NCs. Ces complexes L-NC représentent les unités fonctionnelles indispensables aux mécanismes de réplication et de transcription chez les virus à ARN négatif segmenté et sont appelés RNPs. Du point de vue fondamental, leur étude est donc très importante. D'un point de vue plus pratique, les L et les NP sont également des cibles privilégiées dans le développement de molécules antivirales permettant de bloquer le fonctionnement du virus, ou dans la conception d'outils de diagnostic en cas d'infection. Il est donc doublement important de caractériser ces acteurs viraux, biochimiquement et structuralement, afin de mieux comprendre leurs différents mécanismes d'action.

Jusqu'à il y a quelques années, leur caractérisation structurale était restreinte à l'utilisation de la cristallographie aux rayons X, technique qui a contribué (et qui contribue toujours !) à de nombreuses avancées en biologie structurale mais qui partage peu à peu sa place avec la cryo-ME (voir plus haut).

L'objectif global de cette thèse était donc de mettre à profit les avantages qu'offre la cryo-ME afin de caractériser structuralement, puis fonctionnellement à travers un volet biochimique, les différents acteurs de la réplication et de la transcription chez les bunyavirus en utilisant comme modèle d'étude le HTNV et le LACV, appartenant respectivement aux familles *Hantaviridae* et *Peribunyaviridae*.

Le premier objectif de ma thèse était d'exprimer, produire et déterminer (i) la structure de la HTNV-NC recombinante afin de (ii) caractériser les différentes interactions existantes entre les NPs et (iii) de comprendre le mécanisme d'encapsulation du génome viral au sein des HTNV-NCs. Ces travaux ont été réalisés pendant ma 1^{ère} année de thèse et ont abouti à une publication dans le journal *eLife* en Janvier 2019 (Arragain et al., 2019).

Le second objectif s'inscrit dans la continuité du premier. Il consistait en la détermination des structures par cryo-ME de la polymérase entière du HTNV (HTNV-L) et du LACV (LACV-L) en complexe avec les régions promotrices 5' et 3' spécifiques à chacune des deux familles virales. Le but de cette partie était de mettre en lumière les différences et/ou similitudes structurales existantes avec les structures cristallographiques obtenues de la polymérase du virus influenza, de LACV-L₁₋₁₇₅₀ Δ C-terminal résolue en 2015 mais également des récentes structures obtenues par cryo-ME sur les protéines L de la famille des *Arenaviridae* et des *Phenuiviridae*. Les résultats obtenus sur LACV-L en rapport avec ce deuxième objectif, obtenus au cours de ma 2^{ème} et 3^{ème} année de thèse, ont été publiés dans le journal *Nature Communications* en Juillet 2020 (Arragain et al., 2020). Concernant la structure de la HTNV-L, des avancées ont également été obtenues. Non publiées pour le moment, elles sont présentées dans la partie Résultats.

Enfin, le troisième objectif consistait à caractériser l'activité de la LACV-L via la mise en place et l'optimisation de tests d'activités réalisés *in vitro*. En tirant parti d'une meilleure compréhension de l'activité de cette enzyme et des avantages que procure la cryo-ME, il était envisageable de bloquer la LACV-L dans différentes conformation actives dans le but de caractériser structurellement chacune des différentes étapes effectuées par la polymérase virale au cours de la réplication et de la transcription. Ce travail a été réalisé tout au long de ma 3^{ème} et 4^{ème} année de thèse et les résultats obtenus font l'objet d'un troisième article actuellement en cours de soumission.

Chapitre III : Résultats

1. Obtention de la structure haute résolution par cryo-ME de la HTNV-NC

1.1. Introduction

La visualisation de NCs relativement rigides au sein des virions de la famille *Hantaviridae* (Huiskonen *et al.*, 2010; Battisti *et al.*, 2011) ainsi qu'au sein de cellules infectées (Goldsmith *et al.*, 1995) a servi de base à la première question de ma 1^{ère} année de ma thèse : Est-il possible de déterminer la structure à haute résolution de la NC entière du virus HTNV ? Au vu de la forte pathogénicité de ce virus, je n'ai pas cherché à isoler les NCs à partir de ces derniers mais j'ai tenté de les reconstituer à partir de protéines recombinantes.

Comme indiqué en introduction, certaines NPs de sNSVs peuvent se lier à de l'ARN présent au sein de l'organisme d'expression. J'ai donc utilisé cette stratégie pour les HTNV-NPs que j'ai exprimé en cellules d'insectes. La majorité des NPs se sont révélées insolubles à l'exception d'une petite fraction qui parvient à se lier sur des billes de nickel grâce au tag poly-histidine présent en N-terminal de la protéine. A l'aide de la protéase TEV, le tag est alors clivé permettant ainsi de diminuer la concentration en imidazole utilisé pour l'élution de la colonne de nickel et finalement d'obtenir une protéine homogène susceptible d'être observée en ME par coloration négative. L'observation des HTNV-NCs a révélé la présence de fibres rigides qui présentent les mêmes caractéristiques que les NCs natives précédemment observées (Huiskonen *et al.*, 2010; Battisti *et al.*, 2011). L'analyse du ratio d'absorption A_{260}/A_{280} a également indiqué la présence d'ARNs cellulaires dans les NCs, piégés lors de l'expression.

Suite à l'expression et la purification initiale de ces NCs recombinantes, une difficulté demeurait dans le but de déterminer la structure de la HTNV-NC à haute résolution par cryo-ME : obtenir une concentration suffisante de NCs sur les grilles de cryo-ME. Pour atteindre cet objectif, différents tests d'expressions, de lyses et de purifications ont été réalisés afin de maximiser le rendement, la pureté et l'homogénéité de l'échantillon.

En cryo-ME, des grilles Quantifoil de maille 2/1 couvertes d'un film de carbone de 2 nm d'épaisseur ont été initialement utilisées afin d'améliorer la concentration des HTNV-NCs. Un premier jeu de données, collecté sur le MET FEI Polara équipé d'une caméra K2 Summit alors disponible et en fonctionnement sur la plateforme de ME de l'IBS, a permis de déterminer les premiers paramètres de symétrie hélicoïdale par indexation (voir explication de l'indexation dans le Matériel et Méthodes de l'article 1). Malheureusement, dû au faible contraste des micrographies probablement dû au carbone présent sur les grilles ainsi qu'à la faible résolution de la reconstruction 3D obtenue, il n'a pas été possible de valider ces paramètres (Figure 42).

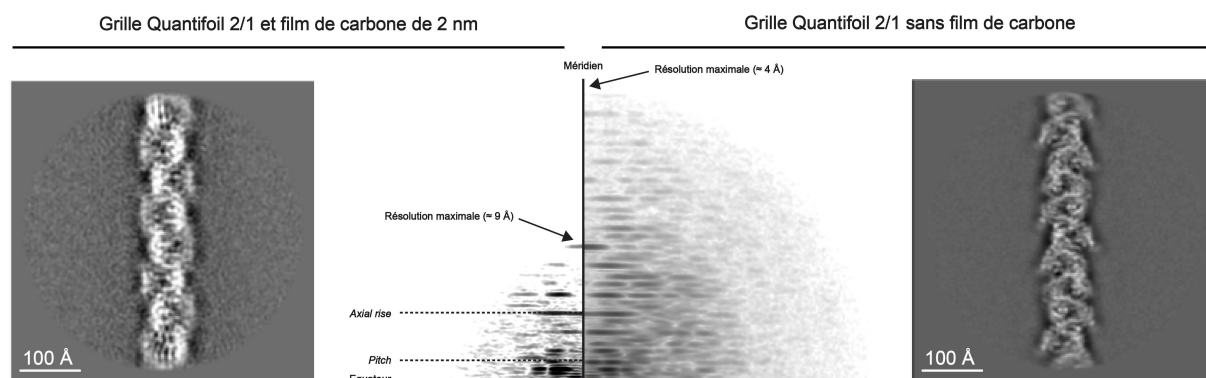


Figure 42. Comparaison des classes 2D et des spectres de puissance obtenus en fonction des grilles utilisées. Le signal maximal en utilisant des grilles couvertes d'un film de carbone de 2 nm est limité à environ 9 Å. Il ne permet pas de valider les paramètres hélicoïdaux déterminés par indexation. En utilisant des grilles sans film de carbone, le signal maximal est de 4 Å. La classe 2D correspondante présente une polarité et des détails de structures secondaires.

Afin d'utiliser des grilles sans film de carbone dans le but d'améliorer le contraste des images collectées, une étape supplémentaire de concentration des HTNV-NCs a été réalisée par ultracentrifugation ce qui m'a permis de les culotter, puis de les resuspendre dans un volume approprié afin d'obtenir une concentration suffisante en cryo-ME. Un second jeu de données a ensuite été collecté sur le MET Polara et une structure à 3.8 Å de résolution globale a été obtenue permettant de valider les paramètres de symétrie précédemment déterminés. Dans le but d'améliorer la résolution de notre reconstruction, un troisième jeu de données a été collecté sur le MET Titan Krios présent à l'ESRF et équipé d'un filtre bio-Quantum et d'une K2 Summit, aboutissant à l'obtention d'une structure à 3.3 Å qui a fait

l'objet d'une première publication que je signe en premier auteur, dans le journal *eLife* en Janvier 2019.

Ce premier article est le fruit des différentes optimisations réalisées au niveau de la préparation de l'échantillon ainsi que de l'apprentissage de l'indexation des spectres de puissance d'objets hélicoïdaux. Étant la première structure haute résolution de NC chez les sNSVs, la structure de la HTNV-NC recombinante apporte des informations importantes sur les différentes interactions qu'entretiennent les différentes NPs entre elles, les résidus impliqués dans la liaison au génome viral et nous a permis d'établir un modèle d'interaction entre la protéine L et les NCs au cours de la synthèse du produit de réplication et de transcription.

1.2. Article n°1 : High resolution cryo-EM structure of the helical RNA-bound Hantaan virus nucleocapsid reveals its assembly mechanisms

High resolution cryo-EM structure of the helical RNA-bound Hantaan virus nucleocapsid reveals its assembly mechanisms

Benoît Arragain¹, Juan Reguera², Ambroise Desfosses¹, Irina Gutsche¹,
Guy Schoehn^{1*}, H  l  ne Malet^{1*}

¹Electron Microscopy and Methods Group, Universit   Grenoble Alpes, CNRS, CEA, Institute for Structural Biology, Grenoble, France; ²Complexes Macromol  culaires Viraux, Aix-Marseille Universit  , CNRS, INSERM, AFMB UMR 7257, Marseille, France

Abstract Negative-strand RNA viruses condense their genome into helical nucleocapsids that constitute essential templates for viral replication and transcription. The intrinsic flexibility of nucleocapsids usually prevents their full-length structural characterisation at high resolution. Here, we describe purification of full-length recombinant metastable helical nucleocapsid of Hantaan virus (*Hantaviridae* family, *Bunyavirales* order) and determine its structure at 3.3   resolution by cryo-electron microscopy. The structure reveals the mechanisms of helical multimerisation via sub-domain exchanges between protomers and highlights nucleotide positions in a continuous positively charged groove compatible with viral genome binding. It uncovers key sites for future structure-based design of antivirals that are currently lacking to counteract life-threatening hantavirus infections. The structure also suggests a model of nucleoprotein-polymerase interaction that would enable replication and transcription solely upon local disruption of the nucleocapsid.

DOI: <https://doi.org/10.7554/eLife.43075.001>

***For correspondence:**

guy.schoehn@ibs.fr (GS);
helene.malet@ibs.fr (HM)

Competing interests: The authors declare that no competing interests exist.

Funding: See page 12

Received: 23 October 2018

Accepted: 10 January 2019

Published: 14 January 2019

Reviewing editor: Edward H Egelman, University of Virginia, United States

   Copyright Arragain et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

Introduction

The *Bunyavirales* order is one of the largest groups of segmented negative-strand RNA viruses (sNSV) that include many pathogenic strains (*Sun et al., 2018*). In particular, the *Hantaviridae* family comprises the virus Hantaan (HTNV) that gives rise to haemorrhagic fevers with renal syndrome and the virus Sin Nombre that is linked to severe pulmonary illnesses with fatality rates up to 40%. Neither treatment nor vaccine is currently available to counteract them.

The *Bunyavirales* genome is usually divided into three RNA segments enwrapped by the viral nucleoproteins (NP). The resulting nucleocapsids (NCs) protect the genome and serve as a replication/transcription template for the viral polymerase (*Reuter and Kr  ger, 2018*). They coat the genomic and anti-genomic RNA during replication but not the mRNA produced by transcription. As they are specific and essential to the viral cycle, NCs constitute an attractive potential target for antiviral drugs.

Nucleoproteins of segmented NSV (sNSV) present a large diversity of folds (*Sun et al., 2018*). Most of the available structures have been determined as rings and monomers that present the advantage of being rigid enough for crystallisation. However, the relevant conformations of assembled NPs in the viral context correspond to flexible helices or pearl-necklaces that encapsulate long RNA segments. Helical crystal structures of La Crosse virus NP (*Peribunyaviridae*, LACV) (*Reguera et al., 2013*) and Crimean Congo Fever Virus NP (*Nairoviridae*, CCFHV) (*Wang et al.,*

eLife digest Rats and mice sometimes transmit hantaviruses, a family of microbes that can cause deadly human diseases. For example, the Hantaan virus leads to haemorrhagic fevers that are potentially fatal. There are no vaccine or even drugs against these infections.

To multiply, viruses must insert their genetic material inside a cell. While the body often detects and destroys foreign genetic information, hantaviruses can still evade our defences. Molecules called nucleoproteins bind to the viral genome, hiding it away in long helices called nucleocapsids. When the virus needs to replicate, an enzyme opens up the nucleocapsid, reads and copies the genetic code, and then closes the helix. Yet, researchers know little about the details of this process, or even the structure of the nucleocapsid.

Here, Arragain et al. use a method called cryo-electron microscopy to examine and piece together the exact 3D structure of the Hantaan virus nucleocapsid. This was possible because the new technique allows scientists to observe biological molecules at an unprecedented, near atomic resolution. The resulting model reveals that the viral genome nests into a groove inside the nucleocapsid. It also shows that specific interactions between nucleoproteins stabilise the helix. Finally, the model helps to provide hypotheses on how the enzyme could read the genome without breaking the capsid.

Mapping out the structure and the interactions of the nucleocapsid is the first step towards finding molecules that could destabilise the helix and neutralise the virus: this could help fight both the Hantaan virus and other members of its deadly family.

DOI: <https://doi.org/10.7554/eLife.43075.002>

2012) have been determined but they correspond to local organisations of pearl-necklace-like native NCs. Influenza double-helical NCs 3D structures have been described at low resolution but display different helical parameters, also reflecting their malleability (Arranz et al., 2012; Moeller et al., 2012). NC flexibility appears as a hallmark in NSVs as NCs from non-segmented NSV such as Ebola or measles virus require C-terminal NP truncations in order to obtain rigid NCs and high-resolution 3D structures (Gutsche et al., 2015; Sugita et al., 2018; Wan et al., 2017). In this context, HTNV-NCs are intriguing as they were shown to be able to form rather rigid helices of 10 nm diameter within viruses (Battisti et al., 2011; Huiskonen et al., 2010) and during cell infection (Goldsmith et al., 1995). We therefore aimed at obtaining their high-resolution 3D structure, identifying the determinants of NP polymerisation and visualising RNA organisation.

Results

Expression of recombinant full-length HTNV-NP in insect cells led to formation of recombinant NCs that have a diameter consistent with native NCs (Figure 1—figure supplement 1). Cryo-EM images collected on a Titan Krios enabled structural determination of HTNV-NC at 3.3 Å resolution (Figure 1A, Video 1, Figure 1—figure supplements 1 and 2, Figure 1—source data 1). HTNV-NC is a left-handed helix with a pitch of 68.03 Å and 3.6 subunits per turn (Figure 1A and Figure 1—figure supplement 1). To derive an atomic model of HTNV-NP, the monomeric crystal structure of HTNV-NP comprising residues 113 to 429 (Olal and Daumke, 2016) was fitted into the EM map, and the N-terminal residues and loops not present in the crystal structures were unambiguously built (Figure 1D). The HTNV-NC model was then iteratively rebuilt and the all-atom model refined using stereochemical restraints.

Although we expressed the full-length NP, residues 1 to 79 are missing in the final map, indicating that they do not follow the helical symmetry. Interestingly, a NP₇₄₋₄₂₉ construct obtained by trypsin limited proteolysis still forms a rigid helix, which means that the N-ter₁₋₇₃ is not necessary for helix stabilisation (Figure 2—figure supplement 1A,B). The N-ter₁₋₇₃ might correspond to a flexibly-linked coiled-coil as previously visualised in the structure of an N-ter construct (Boudko et al., 2007).

The structure of NP₈₀₋₄₂₉ is composed of a core comprising residues 117 to 398 (NP_{core}) that defines two lobes surrounding a positively charged groove (Figure 1B,C, Video 1). The NP_{core} structure is relatively conserved compared to the apo monomeric truncated crystal structure, with a Cα

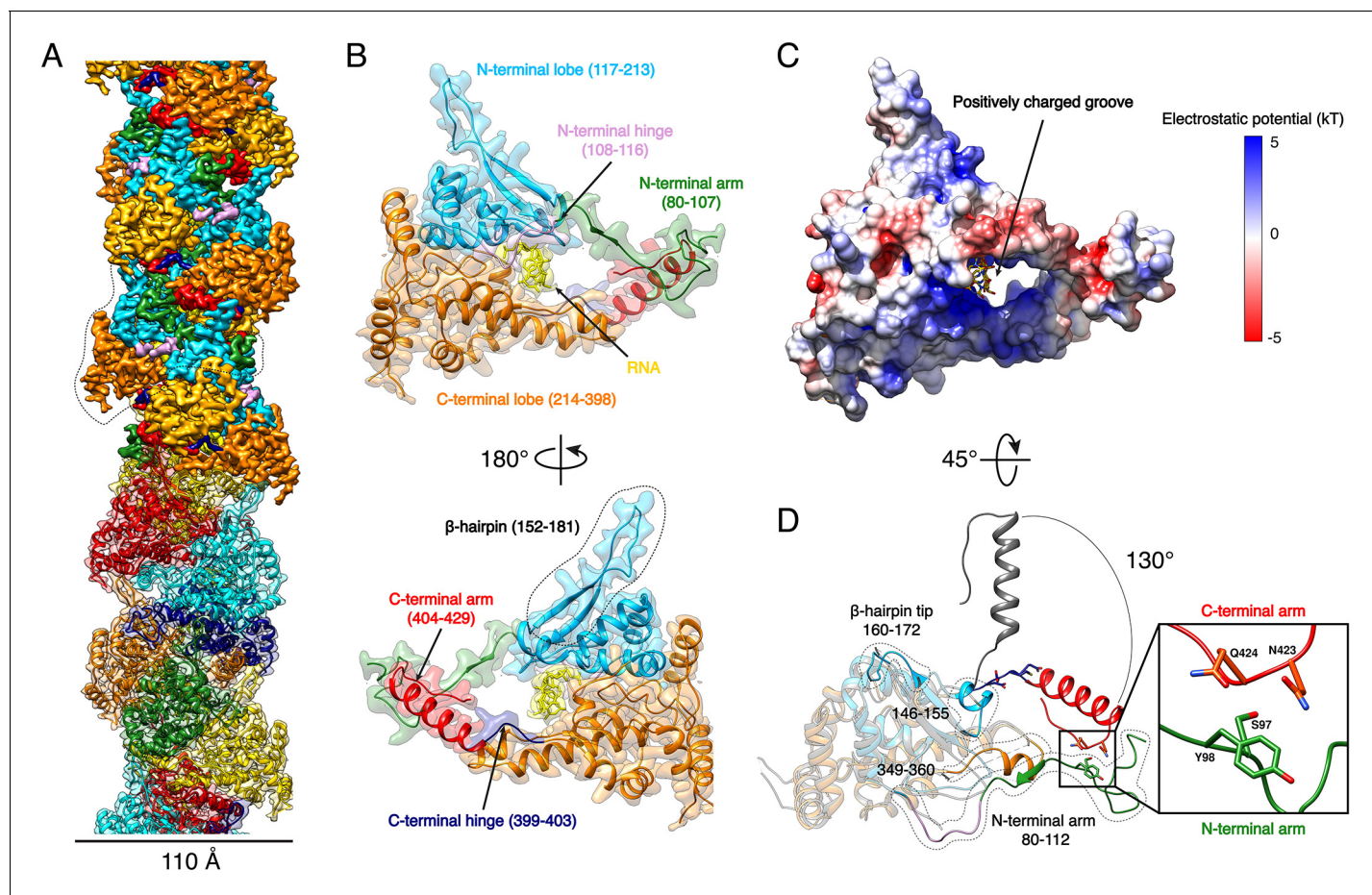


Figure 1. HTNV-NC structure. (A) HTNV-NC helical organisation. The upper part of the cryo-EM map is coloured per NP domain, whereas the lower part is shown as transparent and coloured per NP protomer. The NP surrounded by dotted lines corresponds to **Figure 1B** orientation. (B) Domain organisation of one monomer extracted from the NC helical assembly. Each domain is coloured as in **Figure 1A** upper part. The newly described β -hairpin (residues 152–181) is labelled and surrounded by dotted lines. (C) Electrostatic surface representation of HTNV-NP monomer. (D) Superimposition of monomeric NP_{core} (Olal and Daumke, 2016) (in gray) and NP from the present structure (coloured as in **Figure 1A,B**). The Ct_{arm} rotation is highlighted and newly built elements are labelled, shown as non-transparent and surrounded by dotted lines. A close up view of the Ct_{arm}/Nt_{arm} interaction is shown.

DOI: <https://doi.org/10.7554/eLife.43075.003>

The following source data and figure supplements are available for figure 1:

Source data 1. Cryo-EM data collection, refinement and validation statistics.

DOI: <https://doi.org/10.7554/eLife.43075.007>

Figure supplement 1. Purification, EM analysis and symmetry determination of HTNV-NC.

DOI: <https://doi.org/10.7554/eLife.43075.004>

Figure supplement 2. Resolution of HTNV-NC structure.

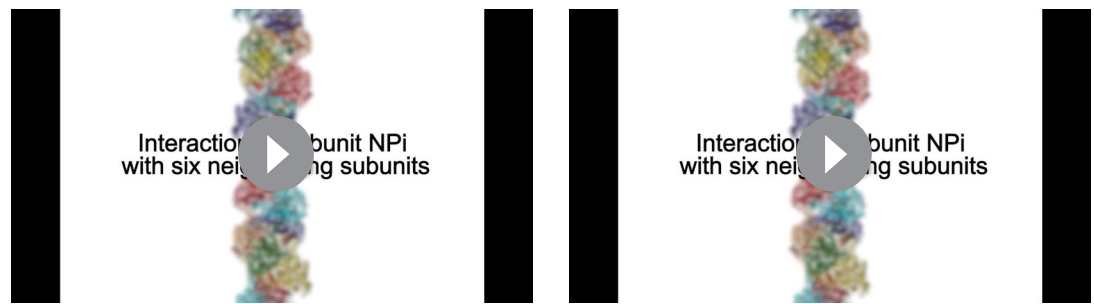
DOI: <https://doi.org/10.7554/eLife.43075.005>

Figure supplement 3. Specific antigenic sites are localised in variable regions of the NC surface.

DOI: <https://doi.org/10.7554/eLife.43075.006>

RMSD of 0.808 Å over 205 atoms (**Figure 1D**). N-terminal and C-terminal arms (Nt_{arm}, Ct_{arm}) are connected to each extremity of NP_{core} by flexible hinges (**Figure 1B, Video 1**). Superimposition of HTNV-NC with the truncated monomeric structure (Olal and Daumke, 2016) shows that the Ct_{arm} undergoes a 130.4° rotation upon multimerisation (**Figure 1D**). As a result, Nt- and Ct_{arm} of the same subunit embrace each other: residues 97–98 of Nt_{arm} contact residues 423–424 of Ct_{arm}, revealing a unique intra-arms interaction specific to HTNV-NC (**Figure 1D, Video 1**).

Recombinant HTNV-NC are rigid and remain stable in a large range of salt, pH and temperature conditions (**Figure 2—figure supplement 1C**). This can be explained by the multiplicity of



Video 1. Overview of HTNV-NC structural organisation. The helical arrangement of HTNV-NC is coloured as in **Figure 1A** upper part. HTNV-NP domain organisation is coloured as in **Figure 1B**. Ct_{arm}/Nt_{arm} interaction is highlighted.

DOI: <https://doi.org/10.7554/eLife.43075.008>

Video 2. Interaction of subunit NP_i with six neighbouring subunits. Subunit NP_i (in green) interacts: (i) with subunit NP_{i-1} (in yellow) via its Nt_{arm}, (ii) with subunit NP_{i+1} (in orange) via its Ct_{arm}, with subunit NP_{i+2} (in dark blue) and subunit NP_{i+3} (in cyan) via its β-hairpin tip. Due to helical symmetry, NP_{i-3} β-hairpin tip (in light cyan) also interacts with NP_i (same interaction as NP_{i-1}-NP_{i+3}). Consistently, NP_{i-2} β-hairpin tip (in red) also interacts with NP_i (same interaction as NP_{i-1}-NP_{i+2}).

DOI: <https://doi.org/10.7554/eLife.43075.011>

interactions between protomers, each NP interacting with six other subunits (**Figure 2A, Video 2**). Successive subunit interactions rely on exchange of their Nt_{arm}, and Ct_{arm} that make intimate contacts with the core domain of neighbouring protomers (**Figure 2A,B,C**), resulting in a buried area of 2704 Å² at each NP-NP interface. The NP_i Nt_{arm} forms an elongated structure that binds to residues 155–160, 177–181, 189–192 and 136–140 of the NP_{i-1} subunit (**Figure 2B, Video 2**). Since constructs lacking the Nt_{arm} remain monomeric (**Olal and Daumke, 2016**), the identified contacts appear to be essential for multimerisation. The NP_i amphipatic Ct_{arm} binds in a hydrophobic pocket of the protomer NP_{i+1} comprising residues 334–346 and 378–394 (**Figure 2C, Video 2**). This agrees with the results of double hybrid experiments (**Kaukinen et al., 2004; Yoshimatsu et al., 2003**) which suggested that interaction of C-terminal helices of neighbouring protomers are critical to oligomerisation. Another key actor of multimerisation is the β-hairpin₁₅₂₋₁₈₁ that protrudes towards the exterior of the NP-core (**Figure 1B**). The two β-strands of the hairpin (residues 155–160 and 177–181) interact with the Nt_{arm} residues 101–103 from the following subunit to form a 3-stranded β-sheet (**Figure 2A,B, Video 2**). As shown by pull-down assays of mutants L102A, V104A, this structure has a main contribution in NC stabilisation (**Guo et al., 2016**). In addition, the β-hairpin₁₅₂₋₁₈₁ tip (residues 162–175) acts as a clamp and seals the C-terminus-mediated contacts between NP_{i+2} (residues 409 to 419) and NP_{i+3} (residues 385–390), thereby buttressing and rigidifying the metastable helical form of the HTNV-NC (**Figure 2A,D, Video 2**).

HTNV recombinant NP was expressed in the absence of viral RNA, but after purification the optical density (OD) at 260–280 nm was measured to be around 1.0, which strongly suggests the binding of insect cell RNAs during expression. Consistently, three nucleotides with a partial occupancy can be visualised for each NP (**Figure 3A**). Conserved residues R197, R314, R368, R146 and K153 interact via salt bridges with RNA phosphates, in line with the versatility of RNA sequence to be incorporated by NP. In addition, F361 stabilises a stacking interaction network between the nucleotide bases. The nucleotides bind in a continuous positively charged groove oriented towards the interior of the NC. Density corresponding to additional nucleotides is visible in this groove, but present at low occupancy, suggesting the lower affinity of NC-RNA binding in these regions. Interestingly, the observed nucleotides occupancy is in accordance with electromobility shift assays of Sin Nombre NP mutants (**Guo et al., 2016**): indeed, in this work, residues directly interacting with the three visible RNA nucleotides were identified as the ones displaying the highest affinity with RNA. Altogether, these results reveal that HTNV-NC structure is compatible with binding of long viral RNA, thereby reinforcing the biological relevance of this structure (**Figure 3B,C**).

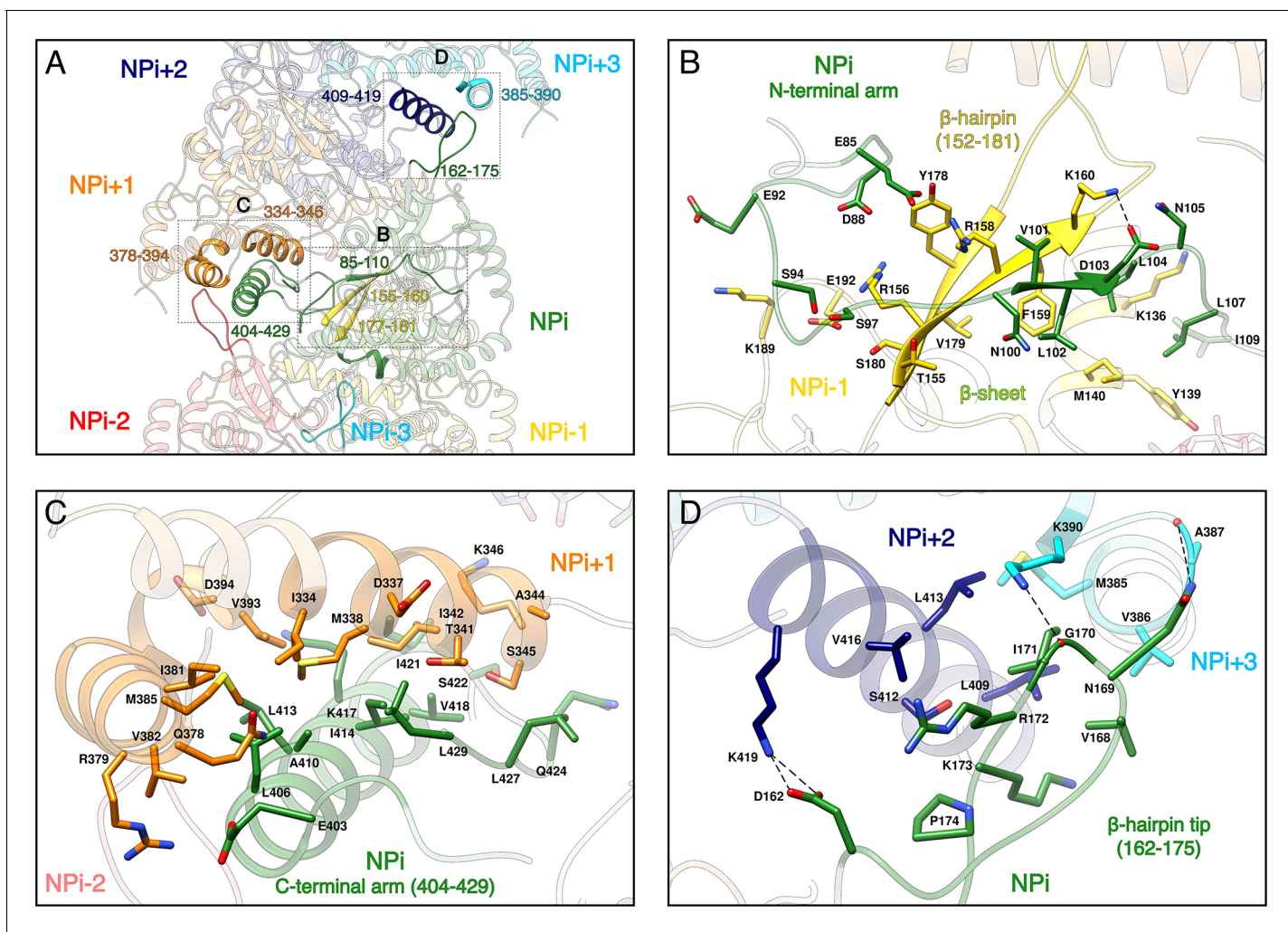


Figure 2. NP-NP interactions. (A) General view. The filament orientation corresponds to RNA direction from 5' to 3'. Each NP protomer is coloured differently. Interacting regions of NP_i with subunits ranging from NP_{i-3} to NP_{i+3} are shown as non-transparent. Positions of **Figure 2B,C,D** close-up views are indicated with dotted lines. (B) NP_i Nt_{arm} binding site in NP_{i-1}. Residues 101–103 from NP_i Nt_{arm} form a β -sheet with β -strands 155–160 and 177–181 from NP_{i-1}. Hydrogen bonds are shown as black dotted lines. (C) NP_i Ct_{arm} binding site in NP_{i+1}. (D) Binding of NP_i β -hairpin tip (residues 162–175) on NP_{i+2} (residues 409–419) and NP_{i+3} (residues 385–390). NP_i I171 is plugged into a hydrophobic pocket formed by L409 and L413 of the NP_{i+2} C-terminal helix and M385 and K390 sidechains of the NP_{i+3} C-terminal lobe. Hydrogen bonds between the residue pairs N169-A387 and G170-K390 from NP_i and NP_{i+3} respectively, further stabilise this interaction contributing to the rigidification of HTNV-NC.

DOI: <https://doi.org/10.7554/eLife.43075.009>

The following figure supplement is available for figure 2:

Figure supplement 1. Limited proteolysis and stability of recombinant HTNV-NCs.

DOI: <https://doi.org/10.7554/eLife.43075.010>

Discussion

Structural determination of a full-length RNA-bound helical NSV NC at 3.3 Å resolution is particularly valuable because the usual intrinsic flexibility of NSV full-length NCs prevents their high-resolution analysis. It fits together the pieces of the jigsaw accumulated over several years of biochemistry and structural analysis on Hantavirus NP. HTNV recombinant NC structure indeed displays 3.6 subunits per turn and is thus consistent with observations of HTNV-NP trimers made by several groups (*Alfadhli et al., 2001; Kaukinen et al., 2004*). The present structure is also compatible with the proposed model (*Kaukinen et al., 2001*) which suggested that NPs first trimerise around the viral RNA and then gradually assemble to form longer multimers. The role of Nt_{arm} and Ct_{arm} exchange between successive subunits, identified by double-hybrid and pull-down experiments (*Guo et al.,*

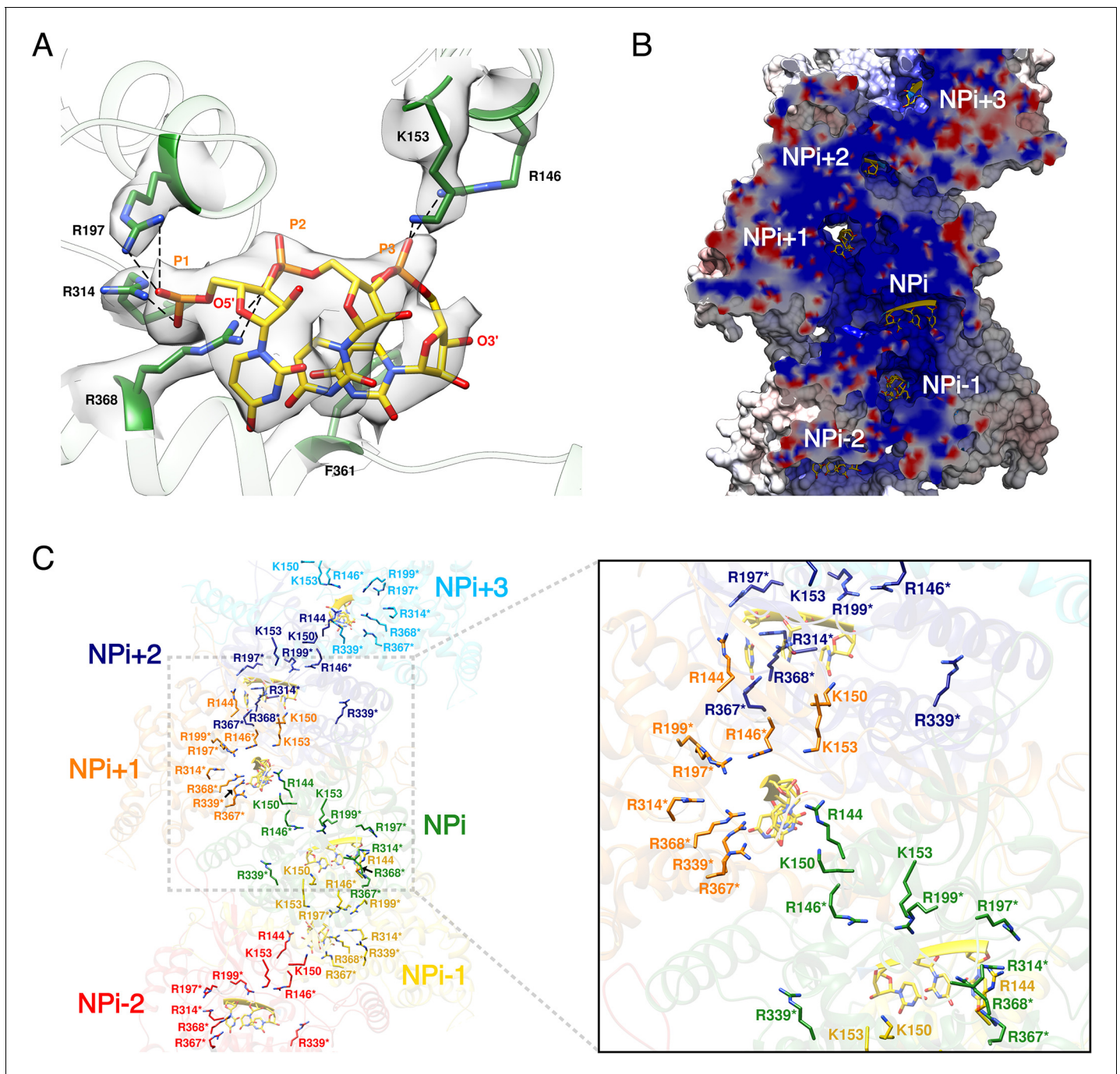


Figure 3. HTNV-NC RNA binding site. (A) RNA binding mode of 3 nucleotides. EM density of RNA and RNA-binding residues are displayed in transparent grey. Hydrogen bonds are shown as dotted lines. (B) Cut-away view of NC electrostatic potential showing the continuous positively charged groove. RNA nucleotides are displayed. (C) RNA-binding residues are shown as sticks and coloured per subunit. RNA-binding residues defined in *Guo et al. (2016)*, namely R146*, R197*, R199*, R314*, R339*, R367*, R368* are labelled with stars, while RNA-binding residues identified in the present structure, namely R144, K150, K153 are shown without stars.

DOI: <https://doi.org/10.7554/eLife.43075.012>

The following figure supplement is available for figure 3:

Figure supplement 1. Comparison of LACV and HTNV-NP/NC, RNA-binding model.

DOI: <https://doi.org/10.7554/eLife.43075.013>

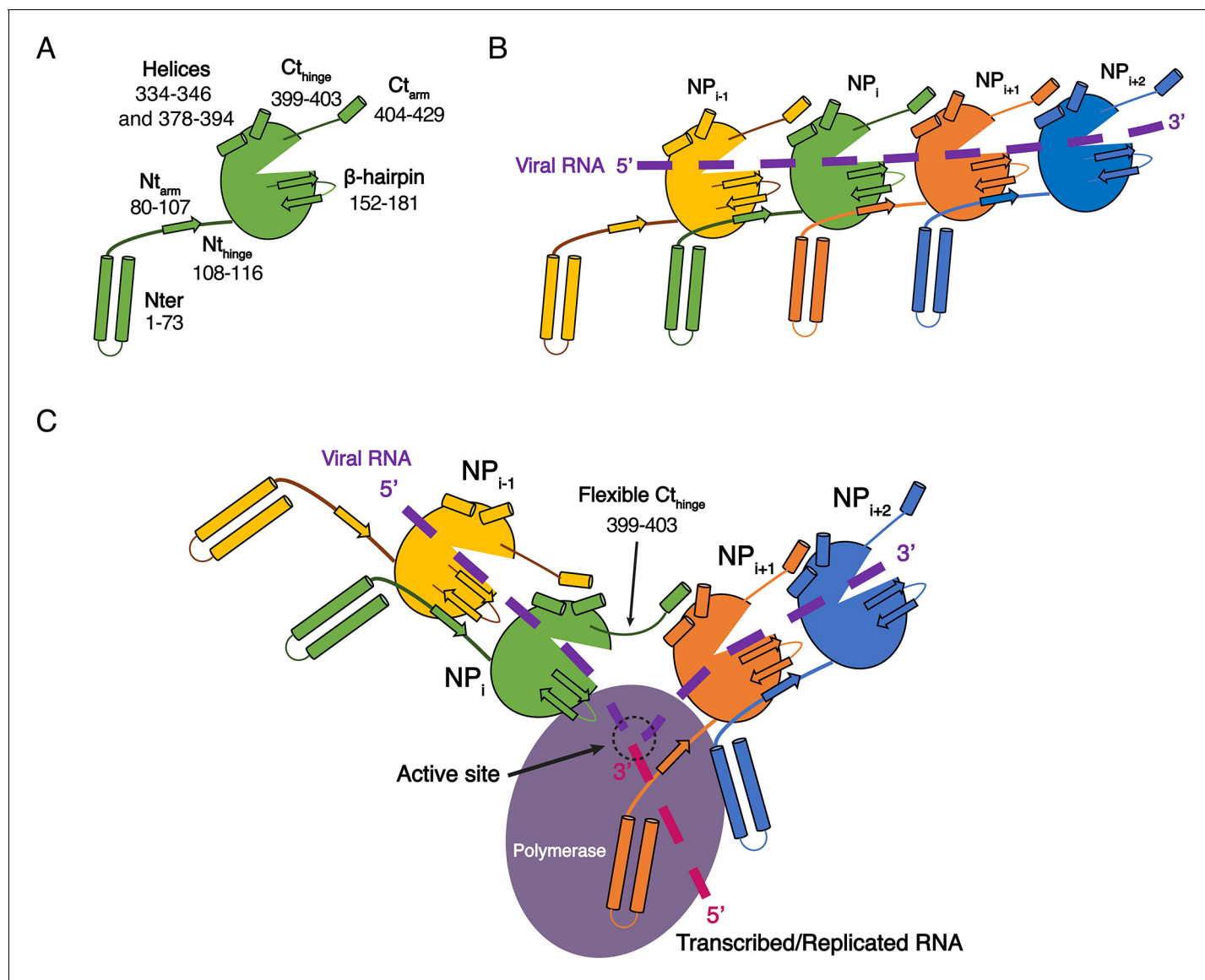


Figure 4. Model of HTNV replication and transcription. (A) Schematics representation of HTNV-NP. Major secondary structures involved in interprotomer interactions are represented as arrows for β -strand and cylinders for α -helices. RNA binding cavity is represented as a clipped part from the NP_{core} region (green oval). (B) Schematic representation of RNA binding and NP-NP interactions. RNA is shown as a dotted purple line. Main NP-NP interactions between adjacent subunits are indicated. For clarity, NP_i interactions with NP_{i-2}, NP_{i-3}, NP_{i+2} and NP_{i+3} are absent from the schematic representation. (C) Replication working hypothesis model which is inspired from [Gerlach et al. \(2015\)](#). The polymerase is shown in purple and the newly transcribed/replicated RNA is indicated as a pink dotted line. The model proposes that the polymerase binds to the flexible N-terminal₁₋₇₃ region during replication/transcription in order to move along the NC. This would be reminiscent to P and L binding to flexible sNSV C-terminal region. Binding of the polymerase to the N-terminal₁₋₇₃ region could destabilise the adjacent Nt_{arm} binding. This localised disruption of NP-NP interaction would create a local opening of the NC enabling transient access of the polymerase to few RNA nucleotides for replication/transcription. Such an opening could be possible without disturbing the whole NC as NP to NP contacts are driven not only by the Nt_{arm} but also by the Ct_{arm}. The Ct_{arm} interaction is likely to remain intact even with a local opening of the NC as the C-terminal hinge allows the Ct_{arm} to undergo large rotation.

DOI: <https://doi.org/10.7554/eLife.43075.014>

2016; [Kaukinen et al., 2004](#); [Yoshimatsu et al., 2003](#)), is in line with HTNV recombinant NC structure. These observations, together with the visualisation of a continuous positively charged groove, strongly suggest that the present structure is biologically pertinent. HTNV recombinant NCs are thus likely to be similar to helical NCs observed within viral particles, although the fact that the latter are less straight implies that at least the β -hairpin₁₅₂₋₁₈₁ might change conformation in the viral context,

thereby enabling more flexibility (Battisti et al., 2011; Huiskonen et al., 2010). Other conformations of NCs are in addition likely to exist because the 5' and 3' end of HTNV viral RNA are known to bind to the viral polymerase, implying that the NC must somehow be circularised (Garcin et al., 1995). Accordingly, flexible pearl-necklaces are also visible within virions, in infected cells and in NC extracted from viral particles (Battisti et al., 2011; Goldsmith et al., 1995; Guo et al., 2016; Huiskonen et al., 2010). Combination of helical NCs observed here and flexible pearl-necklaces thus represent relevant genome-encapsidating conformations.

HTNV and LACV NP_{core} share a common fold (Guo et al., 2016; Olal and Daumke, 2016) enabling their structural superimposition (Figure 3—figure supplement 1A) and comparison of their RNA binding mode. This reveals that the three nucleotides present in HTNV-NC adopt similar conformations as the nucleotides 5, 6 and 7 of LACV-NP (Figure 3—figure supplement 1B) (Reguera et al., 2013). Key residues involved in the binding of HTNV-NC nucleotides, namely R197, R367 and F361 are conserved in LACV-NP (R94, R183 and Y177). Superposition of RNA-bound LACV-NP monomers on HTNV-NC shows that LACV RNA fits reasonably well into the HTNV-NC positively charged groove (Figure 3—figure supplement 1C). However, the proposed HTNV-NC RNA path is slightly shorter, suggesting that each HTNV-NP can contain between 8 and 10 nucleotides (Figure 3—figure supplement 1C). Modelling of RNA binding in HTNV-NC based on RNA position in LACV-NP (see Materials and methods) indicates that RNA phosphates and riboses interact with residues R146, K150, K153, R197, R314, R339, R367 and R368, while bases are surrounded by residues 113–116, 143–151, 182–188 and 217–222 (Figure 3—figure supplement 1D).

While the structures of HTNV and LACV NPs share a common fold for their cores, they however differ in their N-terminal organisation, HTNV Nt_{arm} being significantly longer and linked to the unresolved N-terminal₁₋₇₃ region (Figure 3—figure supplement 1A). The N-terminal₁₋₇₃ region may bring the polymerase in close contact to NP as suggested (Cheng et al., 2014), and thus play a role similar to the intrinsically disordered phosphoproteins and NP C-terminal regions in nsNSV (Ivanov et al., 2011). One may therefore hypothesise that during genome reading by the polymerase, the Nt_{arm} of one HTNV-NP could detach from its adjacent subunit, thereby affecting the conformation of the next NP β-hairpin and removing the seal over the two following NPs. This would provoke a local disruption of the metastable NC and provide RNA access to the polymerase, enabling replication and transcription (Figure 4). This opens new routes for future experiments, such as cryo-EM analysis of *in vivo* reconstituted mini-ribonucleoprotein particles (RNP) or cryo-electron tomography of viral RNPs, that would decipher the exact mode of interaction between NPs, polymerase and RNA during RNA replication/transcription. The rigidity of HTNV-NC and the absence of disordered phosphoprotein in HTNV should facilitate this study compared to the more complex NP/phosphoprotein/polymerase complex of nsNSV, and thus represents a unique opportunity to unravel significant properties of NSV replication.

In addition to these fundamental aspects, the HTNV-NC structure reveals the position of specific antigenic sites on variable regions of the NC surface (Tischler et al., 2008) thus providing a structure-based rationale for diagnosis (Figure 1—figure supplement 3). It may also stimulate the design of antivirals, because it defines key regions involved in NP oligomerisation and RNA encapsidation.

Materials and methods

Key resources table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
recombinant DNA reagent	HTNV-NP	Genearth		Synthetic gene
Peptide, recombinant protein	HTNV-NP	This article	UniProtKB-P05133	HTNV-NP was obtained by expression in insect cells of the synthetic gene HTNV-NP mentioned above
Software, algorithm	Motioncor2	doi: 10.1038/nmeth.2472		http://msg.ucsf.edu/em/software/motioncor2.html

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Software, algorithm	Gctf	doi: 10.1016/j.jsb.2015.11.003	RRID:SCR_016500	https://www.mrc-lmb.cam.ac.uk/kzhang/
Software, algorithm	Relion2.1 and Relion3	doi: 10.1016/j.jsb.2012.09.006 and doi: 10.1101/421123	RRID:SCR_016274	https://www2.mrc-lmb.cam.ac.uk/relion/index.php?title=Main_Page
Software, algorithm	Coot	doi: 10.1107/S0907444910007493	RRID:SCR_014222	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
Software, algorithm	RCrane	doi: 10.1107/S0907444912018549		https://pylelab.org/software/rcrane-readme
Software, algorithm	PHENIX	doi: 10.1107/S0907444909052925	RRID:SCR_014224	https://www.phenix-online.org/
Software, algorithm	Chimera	doi: 10.1002/jcc.20084	RRID:SCR_004097	https://www.cgl.ucsf.edu/chimera/
Software, algorithm	Haddock	doi: 10.1038/nprot.2010.32 and 10.1016/j.jmb.2015.09.01		https://haddock.science.uu.nl/

Cloning, Expression and Purification

Sequence-optimised synthetic DNA encoding a N-terminal his-tag, a TEV protease recognition site and the HTNV-NP (NCBI accession code NC_005218) was synthesised (Geneart) and cloned into a pFastBac1 vector between NdeI and NotI restriction sites.

The HTNV-NP-expressing baculovirus was generated via the standard Bac-to-Bac method (Invitrogen). For large scale expression, Sf21 cells at 0.5×10^6 cells/mL concentration were infected by adding 0.1% of virus. Expression was stopped 72 hr after the day of proliferation arrest.

The cells were disrupted by sonication 3 min (10 s ON, 20 s OFF, 40% amplitude) on ice in lysis buffer (20 mM Tris-HCl pH 8, 300 mM NaCl, 10 mM MgCl₂, 2 mM β-mercaptoethanol and 20 mM Imidazole) with EDTA free protease inhibitor complex (Roche). After lysate centrifugation at 20,000 rpm during 45 min at 4°C, protein from the soluble fraction was loaded on a nickel column, washed with 10 volumes of lysis buffer, 10 volumes lysis buffer supplemented with 50 mM Imidazole and eluted with 5 volumes of lysis buffer supplemented with 500 mM Imidazole. The eluted protein was cleaved with TEV protease in a 20:1 w/w ratio overnight at 4°C in dialysis against lysis buffer resulting in an almost complete cleavage. A second nickel column step was performed to remove unwanted material. The resulting protein was concentrated by ultracentrifugation using Optima XE SW55Ti rotor (Beckman Coulter) and 0.8 mL Ultra Clear tube during 2 hr at 45,000 rpm, 4°C. Concentrated HTNV-NC present in the 20 µl bottom fraction of each tube was gently resuspended. The 260/280 nm absorbance ratio was measured to be around one at the end of the purification indicating presence of nucleic acid.

Three biologically independent batches of large-scale expression and fifteen biologically independent batches of purifications were performed and gave reproducible results.

Limited proteolysis by trypsin and N-terminal sequencing

Limited proteolysis of HTNV-NC was performed at 20°C in lysis buffer with a 3:2 w/w ratio of HTNV-NP/trypsin. Proteolysis was stopped by the addition of denaturing SDS-PAGE loading dye and incubation at 95°C for 5 min. Digestion was visualised in 10% acrylamide SDS-Page gels. For N-terminal sequencing, proteins were transferred on PVDF membrane previously incubated in 10 mM CAPS pH 11, 10% methanol buffer. PVDF membrane was stained using 0.1% Coomassie Brilliant Blue R-250, 40% methanol, 1% acetic acid buffer until bands were visible, washed using 50% methanol and dried. Visible bands were cut and subjected to N-terminal sequencing. Amino acid sequence determination based on Edman degradation was performed using an Applied Biosystems gas-phase sequencer model 492 (s/n: 9510287J). Phenylthiohydantoin amino acid derivatives generated at each sequence cycle were identified and quantitated on-line with an Applied Biosystems Model

140C HPLC system using the data analysis system for protein sequencing from Applied Biosystems (software Procise PC v2.1).

Electron microscopy

For negative stain EM grid preparation, 4 μl of sample was applied between mica and carbon layer and stained using sodium silicotungstate (SST) 2%. After removing the mica part, the grid was deposited on the carbon layer and dried at room temperature. Micrographs were collected at 2.5 μm defocus using a FEI Tecnai F20 operated at 200 kV on 4 k*4 k CETA FEI CCD camera.

For cryo-EM grid preparation, Quantifoil grids 400 mesh 2/1 were glow-discharged at 30 mA for 1 min. 3 μl of sample were applied on the resulting glow-discharged grids and excess solution was blotted during 2.5 s force seven with a Vitrobot Mark IV (FEI) and the grid frozen in liquid ethane. Two biologically independent datasets, arising from different expression and purification batches, were collected on two high-end cryo-electron microscopes and gave consistent structures. The first cryo-EM dataset was collected on a FEI Polara F30 microscope operated at 300 kV equipped with a K2 summit GATAN direct electron detector and resulted in a 3.8 \AA resolution structure. The second dataset, that gave rise to the 3.3 \AA resolution structure presented in the present article, was collected on a FEI Titan Krios operated at 300 kV equipped with a Gatan Bioquantum LS/967 energy filter coupled to a Gatan K2 direct electron detector camera. For this second dataset, automated data collection was performed using EPU FEI software. Zero-loss micrographs were recorded at a 46,860x magnification giving pixel size of 1.067 \AA with defocus ranging from 0.8 to 3.5 μm . In total, 4328 movies with 28 frames per movie were collected with a total dose of 40 $\text{e}^-/\text{\AA}^2$.

Image processing

Micrographs were initially selected based on visual quality inspection. Movie drift correction was performed in Motioncor2 (Li et al., 2013) excluding the two first frames. CTF parameters were determined in Gctf (Zhang, 2016) (RRID:SCR_016500). All subsequent processing steps were performed in Relion2.1 and Relion3 software (Scheres, 2012; Zivanov et al., 2018) (RRID:SCR_016274). Straight HTNV-NC were manually picked and computationally cut with an inter-box distance of 38 \AA along the helical axis into overlapping boxes of 400*400 pixels, resulting in 168,709 extracted segments. 2D classification was used to eliminate bad quality filaments. The best 2D classes were aligned and padded in a square of 1200*1200 pixels². The individual 2D Power Spectra (PS) of each best class were averaged and used for a first estimation of the helical parameters using Fourier Bessel indexing (Cochran et al., 1952). An estimation of the repeat c was first inferred from the layer lines regularly spaced at multiples of $\sim 335.6 \text{\AA}^{-1}$ (Figure 1—figure supplement 1E). Several strong meridional layer lines regularly spaced at multiples of 18.64\AA^{-1} indicated the axial rise p between subunits. According to both c estimation and p , pitch P was inferred to be around 67.12\AA^{-1} as a strong layer line with a first intensity near the meridian can be seen at $l = 5$. Therefore, the structure repeats after $u = 18$ subunits ($18 (u) = 335.6 (c)/18.64 (p)$), in $t = 5$ turns, resulting in a number of units per turn u/t of ~ 3.6 .

3D helical reconstruction was performed using the 111,248 selected segments from the best 2D classes. A 130 \AA diameter featureless cylinder was used as an initial model. Helical symmetry search was performed between $-100 \pm 4^\circ$ for the helical twist ($360^\circ/u/t$) and $18.64 \pm 0.5 \text{\AA}$ for the helical rise. Helical twist and rise after refinement were respectively -99.97° and 18.96 \AA .

For the final 3D helical reconstruction, poorly aligned segments with an angular tilt $>\pm 20^\circ$ were discarded resulting in 105,665 segments. Local CTF-determination was calculated for each segment using Gctf (Zhang, 2016) and finally CTF-refinement was performed using Relion 3.0¹⁸. The 10 last frames of each motion corrected micrographs were removed resulting in a total dose of $\sim 22 \text{e}^-/\text{\AA}^2$. The latest 3D map filtered at 15 \AA resolution was used as initial model and helical symmetry search further refined during a last refinement. The resulting final reconstruction displays refined helical parameters of -99.95° for the twist and 18.87 \AA for the rise. Post-processing, done using a B-factor of -103\AA^2 , resulted in a 3.3 \AA resolution reconstruction using the FSC 0.143 cutoff criteria (Figure 1—figure supplement 2A). Local resolution variations were estimated in Relion (Figure 1—figure supplement 2B).

Observation of nucleotides with partial occupancy suggests that only a fraction of HTNV-NC segments analysed contains RNA. We thus attempted to perform both 2D and 3D classifications in

order to separate RNA-bound and apo HTNV-NC segments using several strategies including usage of different masks, subtraction/absence of subtraction of protein density, symmetry release. However, due to the low molecular weight of RNA compared the protein (9 RNA nucleotides bound per HTNV-NP would correspond to 3 kDa of RNA versus 50 kDa of protein), these classifications did not succeed in separating the apo and RNA-bound states.

Model building and refinement

The monomeric crystal structure of HTNV-NP comprising residues 113 to 429 (PDB code: 5FSG) was initially fitted into the EM density as two separate rigid bodies containing residues 113–398 and 399–429. Loops comprising residues 146–155 and 349–360, β -hairpin 160–172 and the N-terminal arm residues 80 to 112 - previously not present in the crystal structure - were manually built in Coot (*Emsley et al., 2010*). The three RNA nucleotides visible in the density map were manually built and adjusted using Coot (*Emsley et al., 2010*) (RRID:SCR_014222) and RCrane (*Keating and Pyle, 2012*). NP monomeric model was symmetrised according to the helical parameters to form a filament model of 7 subunits. Fiber model was then iteratively rebuilt and all-atom refined using stereochemical and NCS restraints within PHENIX (*Adams et al., 2010*) (RRID:SCR_014224).

RNA modelling based on LACV NP RNA

In order to model HTNV RNA, two HTNV-NP were extracted from HTNV-NC, together with their respective three nucleotides (called 'HTNV-nts'). Five, six or seven LACV-nts were used to link the three HTNV-nts, giving rise to hybrid models containing between eight and ten nucleotides per NP. They were regularised in Coot, minimised in Chimera (*Pettersen et al., 2004*) (RRID:SCR_004097) and were then subjected to modelling in Haddock (*de Vries et al., 2010; van Zundert et al., 2016*). During the modelling procedure, the interaction between HTNV-NP and HTNV-nucleotides was maintained in the same conformation as in the EM map using unambiguous distance restraints. In addition, Haddock ambiguous interaction restraints were defined: residues involved in interactions (*Guo et al., 2016*) were considered as 'active', while all solvent accessible surface neighbours of active residues were defined as 'passive'. The initial starting orientations were not randomised. Rigid body minimisation, semi-flexible simulated annealing and flexible explicit solvent refinement was performed. For each starting RNA, the resulting models were very similar and gathered in one single cluster. The model with the lowest energy and the lowest Haddock score was thus considered as being the modelling result.

Accession numbers

The atomic coordinates and the cryo-EM map have respectively been deposited in the PDB and EMDB under the accession codes 6I2N and EMD-0333.

Acknowledgements

We thank Stephen Cusack for providing the synthetic gene and enabling JR and HM to initiate this project while being post-doctoral researchers in his lab. We thank Karine Huard, Angélique Fraudeau, Delphine Guilligay, Pascal Fender and Alice Aubert for technical advices on expression and purification; Jean-Pierre Andrieu for N-terminal sequencing; Maria Bacia and Michael Hons for help with cryo-EM data collection; Leandro Estrozi for advices on helical symmetry determination and discussion on image processing. This work used the platforms of the Grenoble Instruct-ERIC Center (ISBG: UMS 3518 CNRS-CEA-UGA-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). The electron microscope facility is supported by the Rhône-Alpes Auvergne Region, the Fondation Recherche Medicale (FRM), the fonds FEDER and the GIS-Infrastructures en Biologie Sante et Agronomie (IBISA). We acknowledge the European Synchrotron Radiation Facility for provision of microscope time on CM01. We thank all platform staff that enabled us to perform these analyses.

Additional information

Funding

Funder	Grant reference number	Author
Université Grenoble Alpes	G7H-IRS17H50	Hélène Malet

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Benoît Arragain, Data curation, Validation, Investigation, Visualization, Methodology, Writing—original draft, Writing—review and editing; Juan Reguera, Ambroise Desfosses, Data curation, Investigation, Methodology, Writing—review and editing; Irina Gutsche, Resources, Data curation, Methodology, Writing—review and editing; Guy Schoehn, Conceptualization, Resources, Data curation, Supervision, Funding acquisition, Validation, Methodology, Project administration, Writing—review and editing; Hélène Malet, Conceptualization, Data curation, Supervision, Funding acquisition, Validation, Investigation, Visualization, Methodology, Writing—original draft, Project administration, Writing—review and editing

Author ORCIDs

Benoît Arragain  <http://orcid.org/0000-0002-5593-4682>

Juan Reguera  <http://orcid.org/0000-0003-4977-7948>

Ambroise Desfosses  <http://orcid.org/0000-0002-6525-5042>

Irina Gutsche  <http://orcid.org/0000-0002-1908-3921>

Guy Schoehn  <http://orcid.org/0000-0002-1459-3201>

Hélène Malet  <https://orcid.org/0000-0002-2834-7386>

Decision letter and Author response

Decision letter <https://doi.org/10.7554/eLife.43075.023>

Author response <https://doi.org/10.7554/eLife.43075.024>

Additional files

Supplementary files

- Transparent reporting form

DOI: <https://doi.org/10.7554/eLife.43075.015>

Data availability

The cryo-EM has been deposited in EMDB under the accession code EMD-0333. The atomic coordinates have been deposited in PDB under the accession code 6I2N.

The following datasets were generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Arragain B, Reguera J, Desfosses A, Gutsche I	2018	Data from: High resolution cryo-EM structure of the helical RNA-bound Hantaan virus nucleocapsid reveals its assembly mechanisms	https://www.rcsb.org/structure/6I2N	Protein Data Bank, 6I2N
Arragain B, Reguera J, Desfosses A, Gutsche I	2018	Data from: High resolution cryo-EM structure of the helical RNA-bound Hantaan virus nucleocapsid reveals its assembly mechanisms	https://www.ebi.ac.uk/pdbe/entry/emdb/EMD-0333	Electron Microscopy Data Bank, EMD-0333

References

- Adams PD**, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography* **66**:213–221. DOI: <https://doi.org/10.1107/S0907444909052925>, PMID: 20124702
- Alfadhli A**, Love Z, Arvidson B, Seeds J, Willey J, Barklis E. 2001. Hantavirus nucleocapsid protein oligomerization. *Journal of Virology* **75**:2019–2023. DOI: <https://doi.org/10.1128/JVI.75.4.2019-2023.2001>, PMID: 11160704
- Arranz R**, Coloma R, Chichón FJ, Conesa JJ, Carrascosa JL, Valpuesta JM, Ortín J, Martín-Benito J. 2012. The structure of native influenza virion ribonucleoproteins. *Science* **338**:1634–1637. DOI: <https://doi.org/10.1126/science.1228172>, PMID: 23180776
- Battisti AJ**, Chu YK, Chipman PR, Kaufmann B, Jonsson CB, Rossmann MG. 2011. Structural studies of hantaan virus. *Journal of Virology* **85**:835–841. DOI: <https://doi.org/10.1128/JVI.01847-10>, PMID: 21068243
- Boudko SP**, Kuhn RJ, Rossmann MG. 2007. The coiled-coil domain structure of the sin nombre virus nucleocapsid protein. *Journal of Molecular Biology* **366**:1538–1544. DOI: <https://doi.org/10.1016/j.jmb.2006.12.046>, PMID: 17222867
- Cheng E**, Wang Z, Mir MA. 2014. Interaction between hantavirus nucleocapsid protein (N) and RNA-dependent RNA polymerase (RdRp) mutants reveals the requirement of an N-RdRp interaction for viral RNA synthesis. *Journal of Virology* **88**:8706–8712. DOI: <https://doi.org/10.1128/JVI.00405-14>, PMID: 24850733
- Cochran W**, Crick FH, Vand V. 1952. The structure of synthetic polypeptides. I. the transform of atoms on a helix. *Acta Crystallographica* **5**:581–586. DOI: <https://doi.org/10.1107/S0365110X52001635>
- de Vries SJ**, van Dijk M, Bonvin AM. 2010. The HADDOCK web server for data-driven biomolecular docking. *Nature Protocols* **5**:883–897. DOI: <https://doi.org/10.1038/nprot.2010.32>, PMID: 20431534
- Emsley P**, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of coot. *Acta Crystallographica. Section D, Biological Crystallography* **66**:486–501. DOI: <https://doi.org/10.1107/S0907444910007493>, PMID: 20383002
- Garcin D**, Lezzi M, Dobbs M, Elliott RM, Schmaljohn C, Kang CY, Kolakofsky D. 1995. The 5' ends of hantaan virus (Bunyaviridae) RNAs suggest a prime-and-realign mechanism for the initiation of RNA synthesis. *Journal of Virology* **69**:5754–5762. PMID: 7637020
- Gerlach P**, Malet H, Cusack S, Reguera J. 2015. Structural insights into bunyavirus replication and its regulation by the vRNA promoter. *Cell* **161**:1267–1279. DOI: <https://doi.org/10.1016/j.cell.2015.05.006>, PMID: 26004069
- Goldsmith CS**, Elliott LH, Peters CJ, Zaki SR. 1995. Ultrastructural characteristics of Sin Nombre virus, causative agent of hantavirus pulmonary syndrome. *Archives of Virology* **140**:2107–2122. DOI: <https://doi.org/10.1007/BF01323234>, PMID: 8572935
- Guo Y**, Wang W, Sun Y, Ma C, Wang X, Wang X, Liu P, Shen S, Li B, Lin J, Deng F, Wang H, Lou Z. 2016. Crystal structure of the core region of hantavirus nucleocapsid protein reveals the mechanism for ribonucleoprotein complex formation. *Journal of Virology* **90**:1048–1061. DOI: <https://doi.org/10.1128/JVI.02523-15>, PMID: 26559827
- Gutsche I**, Desfosses A, Effantin G, Ling WL, Haupt M, Ruigrok RW, Sachse C, Schoehn G. 2015. Structural virology. Near-atomic cryo-EM structure of the helical measles virus nucleocapsid. *Science* **348**:704–707. DOI: <https://doi.org/10.1126/science.aaa5137>, PMID: 25883315
- Huiskonen JT**, Hepojoki J, Laurinmäki P, Vaeheri A, Lankinen H, Butcher SJ, Grünwald K. 2010. Electron cryotomography of tula hantavirus suggests a unique assembly paradigm for enveloped viruses. *Journal of Virology* **84**:4889–4897. DOI: <https://doi.org/10.1128/JVI.00057-10>, PMID: 20219926
- Ivanov I**, Yabukarski F, Ruigrok RW, Jamin M. 2011. Structural insights into the rhabdovirus transcription/replication complex. *Virus Research* **162**:126–137. DOI: <https://doi.org/10.1016/j.virusres.2011.09.025>, PMID: 21963663
- Kaukinen P**, Koistinen V, Vapalahti O, Vaeheri A, Plyusnin A. 2001. Interaction between molecules of hantavirus nucleocapsid protein. *Journal of General Virology* **82**:1845–1853. DOI: <https://doi.org/10.1099/0022-1317-82-8-1845>, PMID: 11457990
- Kaukinen P**, Kumar V, Tulimäki K, Engelhardt P, Vaeheri A, Plyusnin A. 2004. Oligomerization of hantavirus N protein: c-terminal alpha-helices interact to form a shared hydrophobic space. *Journal of Virology* **78**:13669–13677. DOI: <https://doi.org/10.1128/JVI.78.24.13669-13677.2004>, PMID: 15564476
- Keating KS**, Pyle AM. 2012. RCrane: semi-automated RNA model building. *Acta Crystallographica. Section D, Biological Crystallography* **68**:985–995. DOI: <https://doi.org/10.1107/S0907444912018549>, PMID: 22868764
- Li X**, Mooney P, Zheng S, Booth CR, Braunfeld MB, Gubbens S, Agard DA, Cheng Y. 2013. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* **10**:584–590. DOI: <https://doi.org/10.1038/nmeth.2472>, PMID: 23644547
- Moeller A**, Kirchdoerfer RN, Potter CS, Carragher B, Wilson IA. 2012. Organization of the influenza virus replication machinery. *Science* **338**:1631–1634. DOI: <https://doi.org/10.1126/science.1227270>, PMID: 23180774
- Olal D**, Daumke O. 2016. Structure of the hantavirus nucleoprotein provides insights into the mechanism of RNA encapsidation. *Cell Reports* **14**:2092–2099. DOI: <https://doi.org/10.1016/j.celrep.2016.02.005>, PMID: 26923588

- Pettersen EF**, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**:1605–1612. DOI: <https://doi.org/10.1002/jcc.20084>, PMID: 15264254
- Reguera J**, Malet H, Weber F, Cusack S. 2013. Structural basis for encapsidation of genomic RNA by la crosse orthobunyavirus nucleoprotein. *PNAS* **110**:7246–7251. DOI: <https://doi.org/10.1073/pnas.1302298110>, PMID: 23589854
- Reuter M**, Krüger DH. 2018. The nucleocapsid protein of hantaviruses: much more than a genome-wrapping protein. *Virus Genes* **54**:5–16. DOI: <https://doi.org/10.1007/s11262-017-1522-3>, PMID: 29159494
- Scheres SH**. 2012. RELION: implementation of a bayesian approach to cryo-EM structure determination. *Journal of Structural Biology* **180**:519–530. DOI: <https://doi.org/10.1016/j.jsb.2012.09.006>, PMID: 23000701
- Sugita Y**, Matsunami H, Kawaoka Y, Noda T, Wolf M. 2018. Cryo-EM structure of the ebola virus nucleoprotein–RNA complex at 3.6 Å resolution. *Nature* **1**. DOI: <https://doi.org/10.1038/s41586-018-0630-0>
- Sun Y**, Li J, Gao GF, Tien P, Liu W. 2018. Bunyavirales ribonucleoproteins: the viral replication and transcription machinery. *Critical Reviews in Microbiology* **44**:522–540. DOI: <https://doi.org/10.1080/1040841X.2018.1446901>, PMID: 29516765
- Tischler ND**, Roseblatt M, Valenzuela PD. 2008. Characterization of cross-reactive and serotype-specific epitopes on the nucleocapsid proteins of hantaviruses. *Virus Research* **135**:1–9. DOI: <https://doi.org/10.1016/j.virusres.2008.01.013>, PMID: 18342973
- van Zundert GCP**, Rodrigues J, Trellet M, Schmitz C, Kastrius PL, Karaca E, Melquiond ASJ, van Dijk M, de Vries SJ, Bonvin A. 2016. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of Molecular Biology* **428**:720–725. DOI: <https://doi.org/10.1016/j.jmb.2015.09.014>, PMID: 26410586
- Wan W**, Kolesnikova L, Clarke M, Koehler A, Noda T, Becker S, Briggs JAG. 2017. Structure and assembly of the ebola virus nucleocapsid. *Nature* **551**:394–397. DOI: <https://doi.org/10.1038/nature24490>, PMID: 29144446
- Wang Y**, Dutta S, Karlberg H, Devignot S, Weber F, Hao Q, Tan YJ, Mirazimi A, Kotaka M. 2012. Structure of Crimean-Congo hemorrhagic fever virus nucleoprotein: superhelical homo-oligomers and the role of caspase-3 cleavage. *Journal of Virology* **86**:12294–12303. DOI: <https://doi.org/10.1128/JVI.01627-12>, PMID: 22951837
- Yoshimatsu K**, Lee BH, Araki K, Morimatsu M, Ogino M, Ebihara H, Arikawa J. 2003. The multimerization of hantavirus nucleocapsid protein depends on type-specific epitopes. *Journal of Virology* **77**:943–952. DOI: <https://doi.org/10.1128/JVI.77.2.943-952.2003>, PMID: 12502810
- Zhang K**. 2016. Gctf: real-time CTF determination and correction. *Journal of Structural Biology* **193**:1–12. DOI: <https://doi.org/10.1016/j.jsb.2015.11.003>, PMID: 26592709
- Zivanov J**, Nakane T, Forsberg BO, Kimanius D, Hagen WJ, Lindahl E, Scheres SH. 2018. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**:e42166. DOI: <https://doi.org/10.7554/eLife.42166>, PMID: 30412051



Figures and figure supplements

High resolution cryo-EM structure of the helical RNA-bound Hantaan virus nucleocapsid reveals its assembly mechanisms

Benoît Arragain *et al*

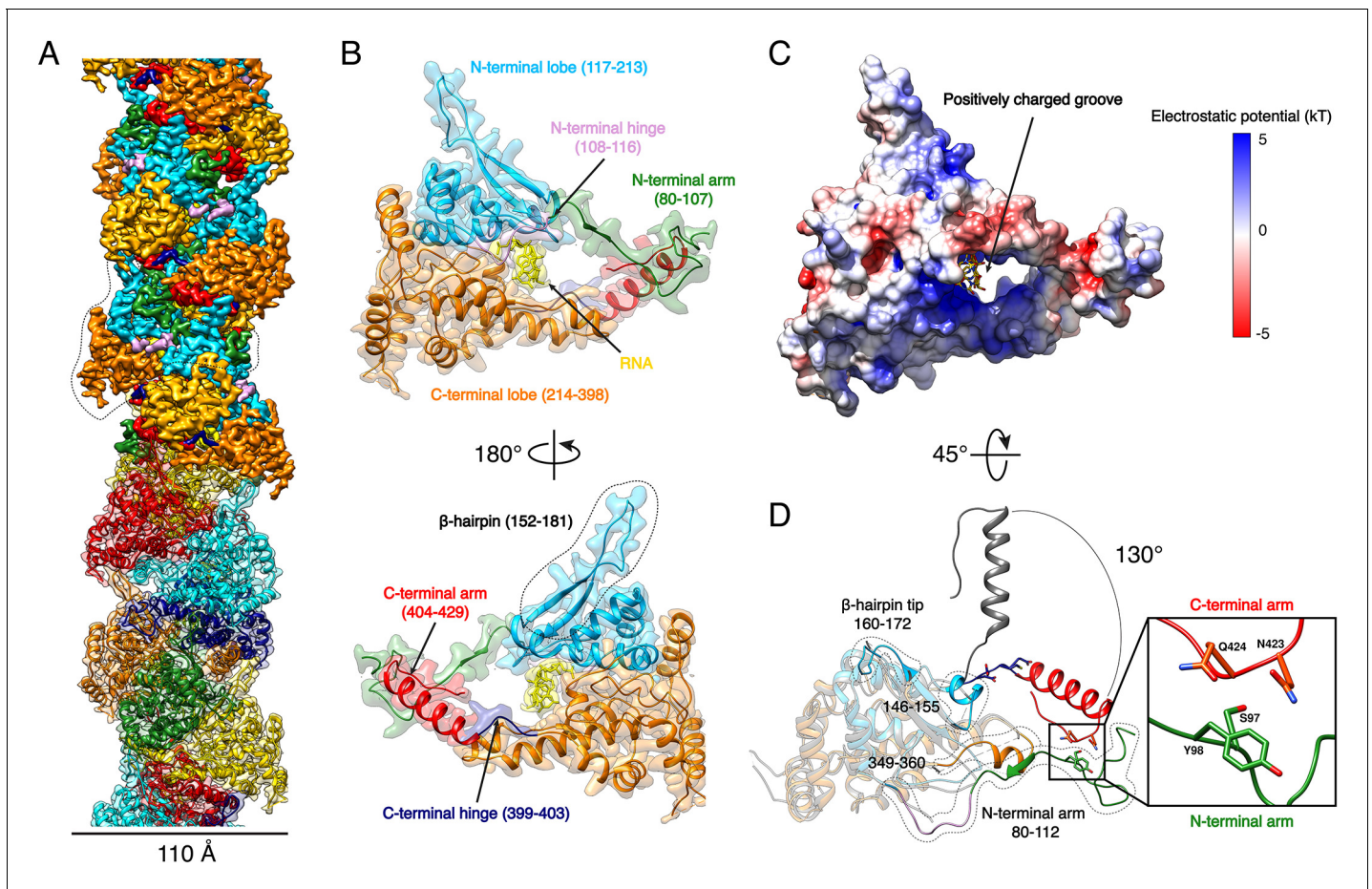


Figure 1. HTNV-NC structure. (A) HTNV-NC helical organisation. The upper part of the cryo-EM map is coloured per NP domain, whereas the lower part is shown as transparent and coloured per NP protomer. The NP surrounded by dotted lines corresponds to **Figure 1B** orientation. (B) Domain organisation of one monomer extracted from the NC helical assembly. Each domain is coloured as in **Figure 1A** upper part. The newly described β -hairpin (residues 152–181) is labelled and surrounded by dotted lines. (C) Electrostatic surface representation of HTNV-NP monomer. (D) Superimposition of monomeric NP_{core} (Olal and Daumke, 2016) (in gray) and NP from the present structure (coloured as in **Figure 1A,B**). The Ct_{arm} rotation is highlighted and newly built elements are labelled, shown as non-transparent and surrounded by dotted lines. A close up view of the Ct_{arm}/Nt_{arm} interaction is shown.

DOI: <https://doi.org/10.7554/eLife.43075.003>

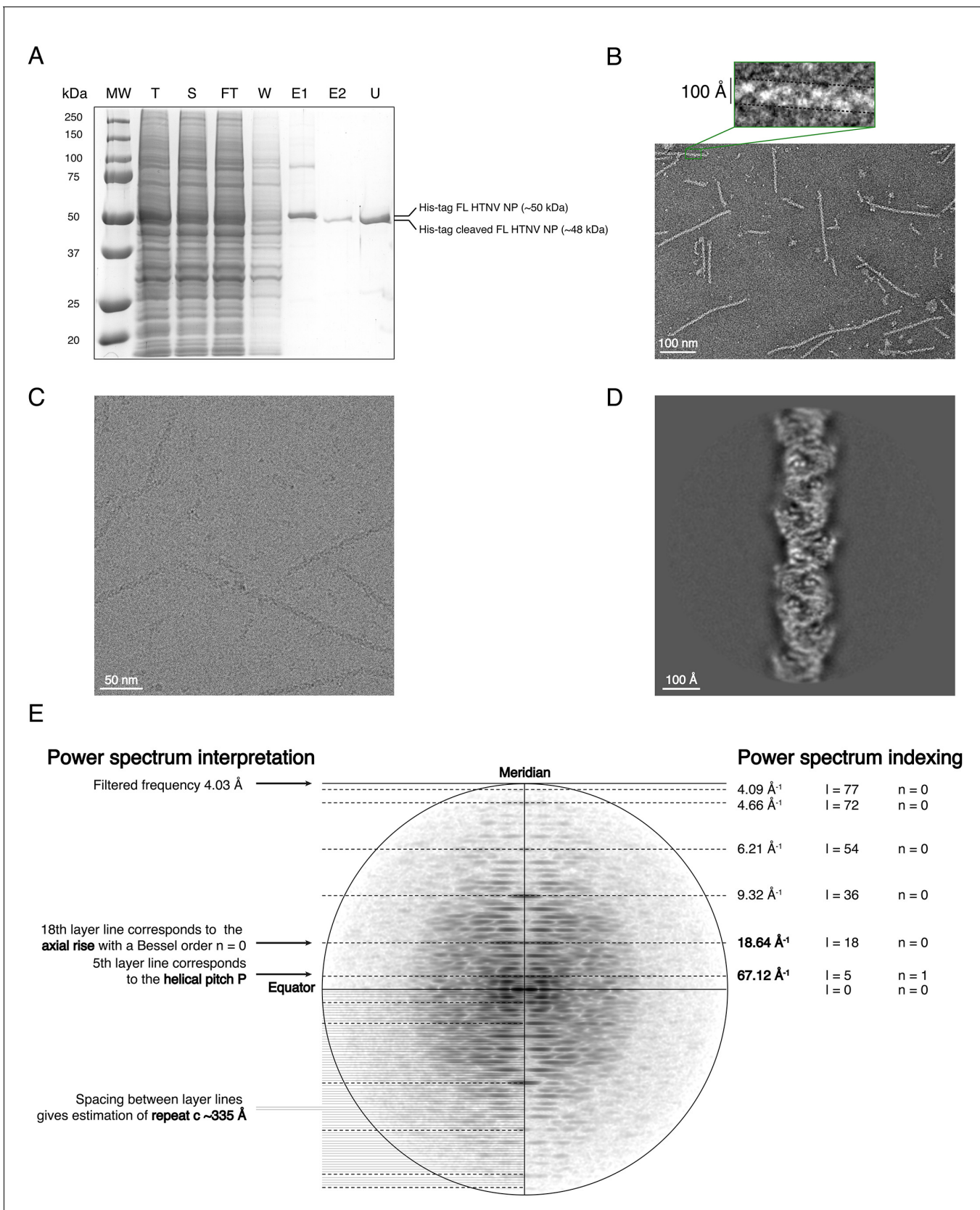


Figure 1—figure supplement 1. Purification, EM analysis and symmetry determination of HTNV-NC. (A) SDS-PAGE gel of HTNV-NC purification. T: total fraction, S: soluble fraction, FT: flow through of the 1st Ni-NTA purification, W: wash of the 1st Ni-NTA purification, E1: elution of the 1st Ni-NTA purification. Figure 1—figure supplement 1 continued on next page

Figure 1—figure supplement 1 continued

purification, E2: supernatant after TEV cleavage and 2nd Ni-NTA purification, U: concentrated protein after ultra-centrifugation. (B,C) Negative stain (B) and cryo electron micrographs (C) of HTNV-NC. Scale bar is indicated. (D) 2D class average. (E) 2D class average power spectrum, indexing and interpretation enabling symmetry determination.

DOI: <https://doi.org/10.7554/eLife.43075.004>

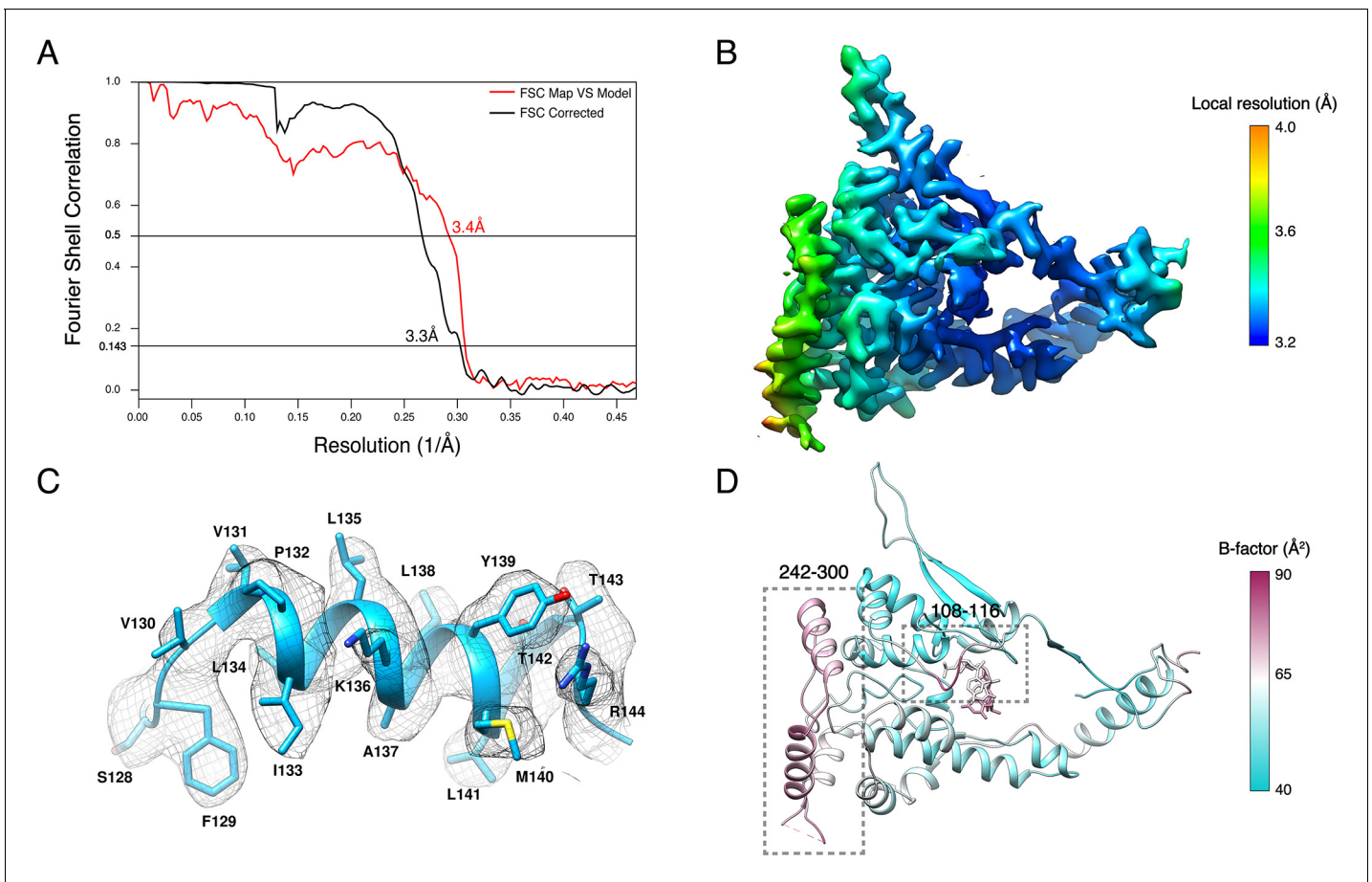


Figure 1—figure supplement 2. Resolution of HTNV-NC structure. (A) The gold-standard Fourier shell correlation (FSC) of masked map indicates a resolution of 3.3 Å with the FSC = 0.143 criteria. The FSC between the map and the model is 3.4 Å according to FSC = 0.5 criterion. (B) Local resolution. (C) Zoom-up view of a representative α -helix that shows clear side chains densities illustrating the quality of the map. (D) The B-factor analysis strongly correlates with the local resolution analysis (B) and shows that the N-terminal hinge and the solvent-exposed region containing residues 242–300 are flexible.

DOI: <https://doi.org/10.7554/eLife.43075.005>

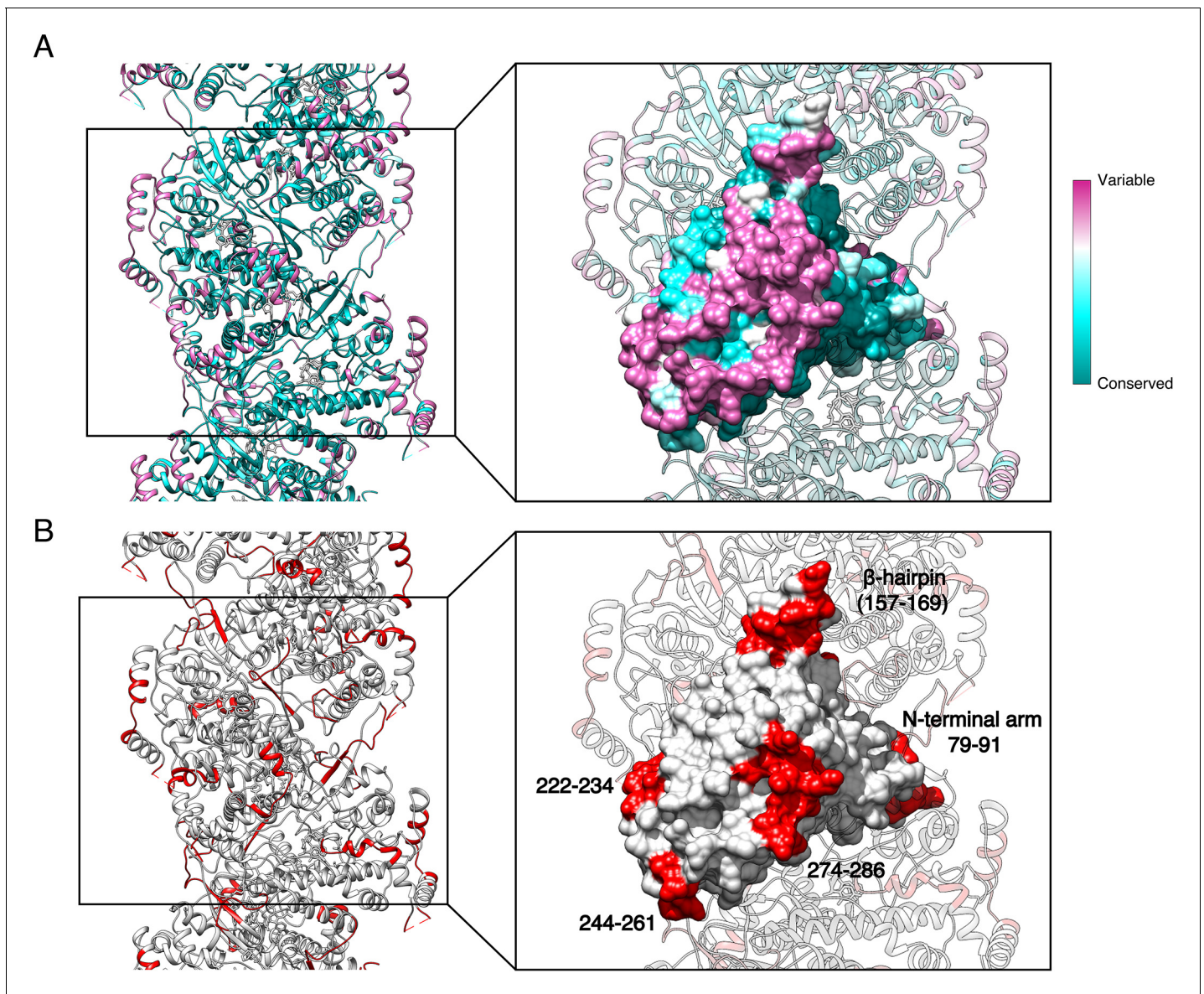


Figure 1—figure supplement 3. Specific antigenic sites are localised in variable regions of the NC surface. (A) Surface conservation mapping of HTNV-NC. Regions localised towards NC interior, in particular regions involved in RNA-binding and NP-NP oligomerisation are conserved, whereas HTNV-NC exposed regions are more variable. (B) Epitopes involved in serotype-specific and cross-reactive recognition of HTNV-NC determined in (Tischler et al., 2008) are shown in red. They correspond to surface-exposed regions of HTNV-NC corroborating their necessary accessibility.

DOI: <https://doi.org/10.7554/eLife.43075.006>

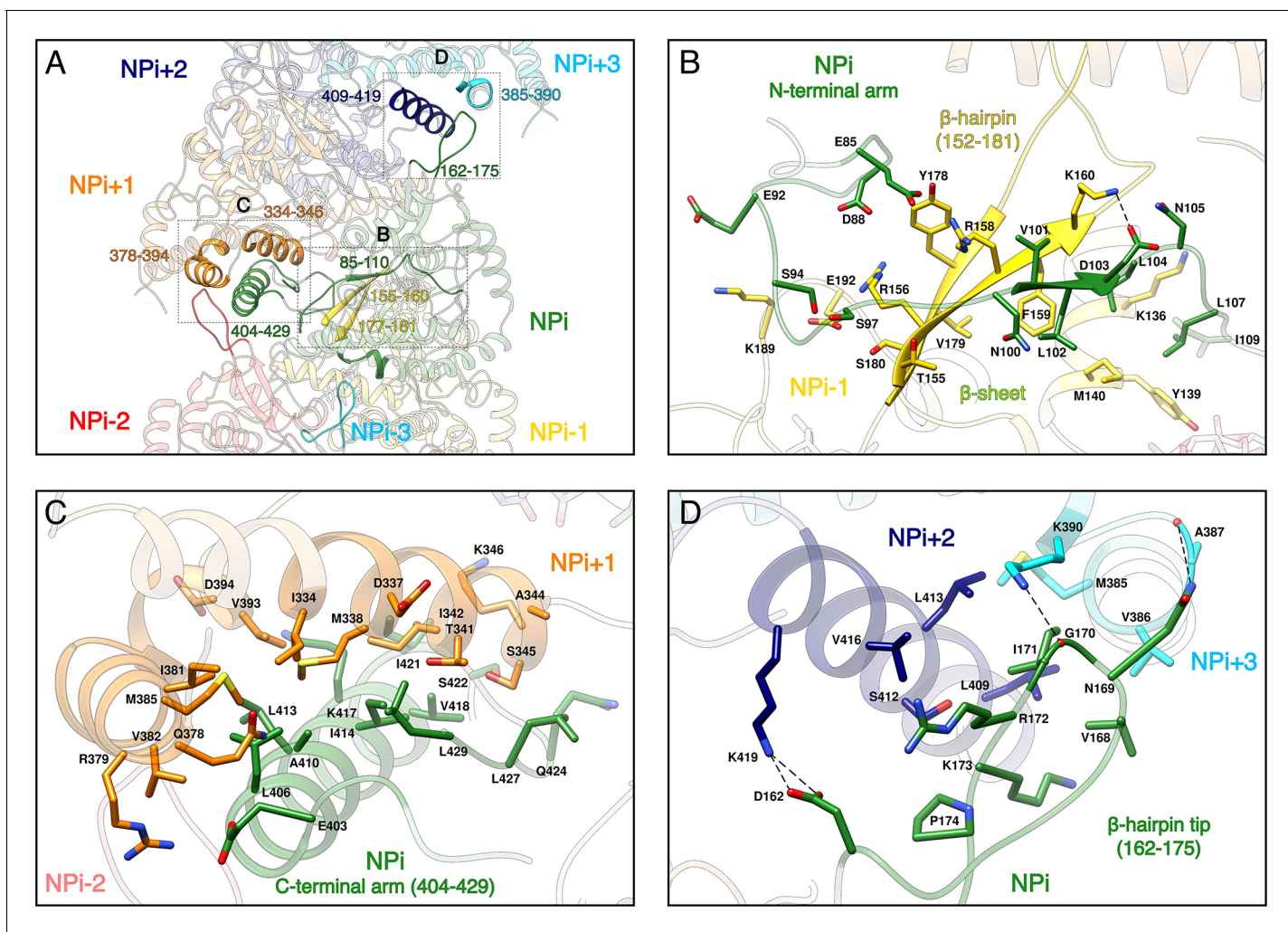


Figure 2. NP-NP interactions. (A) General view. The filament orientation corresponds to RNA direction from 5' to 3'. Each NP protomer is coloured differently. Interacting regions of NP_i with subunits ranging from NP_{i-3} to NP_{i+3} are shown as non-transparent. Positions of **Figure 2B,C,D** close-up views are indicated with dotted lines. (B) NP_i Nt_{arm} binding site in NP_{i-1}. Residues 101–103 from NP_i Nt_{arm} form a β-sheet with β-strands 155–160 and 177–181 from NP_{i-1}. Hydrogen bonds are shown as black dotted lines. (C) NP_i Ct_{arm} binding site in NP_{i+1}. (D) Binding of NP_i β-hairpin tip (residues 162–175) on NP_{i+2} (residues 409–419) and NP_{i+3} (residues 385–390). NP_i I171 is plugged into a hydrophobic pocket formed by L409 and L413 of the NP_{i+2} C-terminal helix and M385 and K390 sidechains of the NP_{i+3} C-terminal lobe. Hydrogen bonds between the residue pairs N169-A387 and G170-K390 from NP_i and NP_{i+3} respectively, further stabilise this interaction contributing to the rigidification of HTNV-NC.

DOI: <https://doi.org/10.7554/eLife.43075.009>

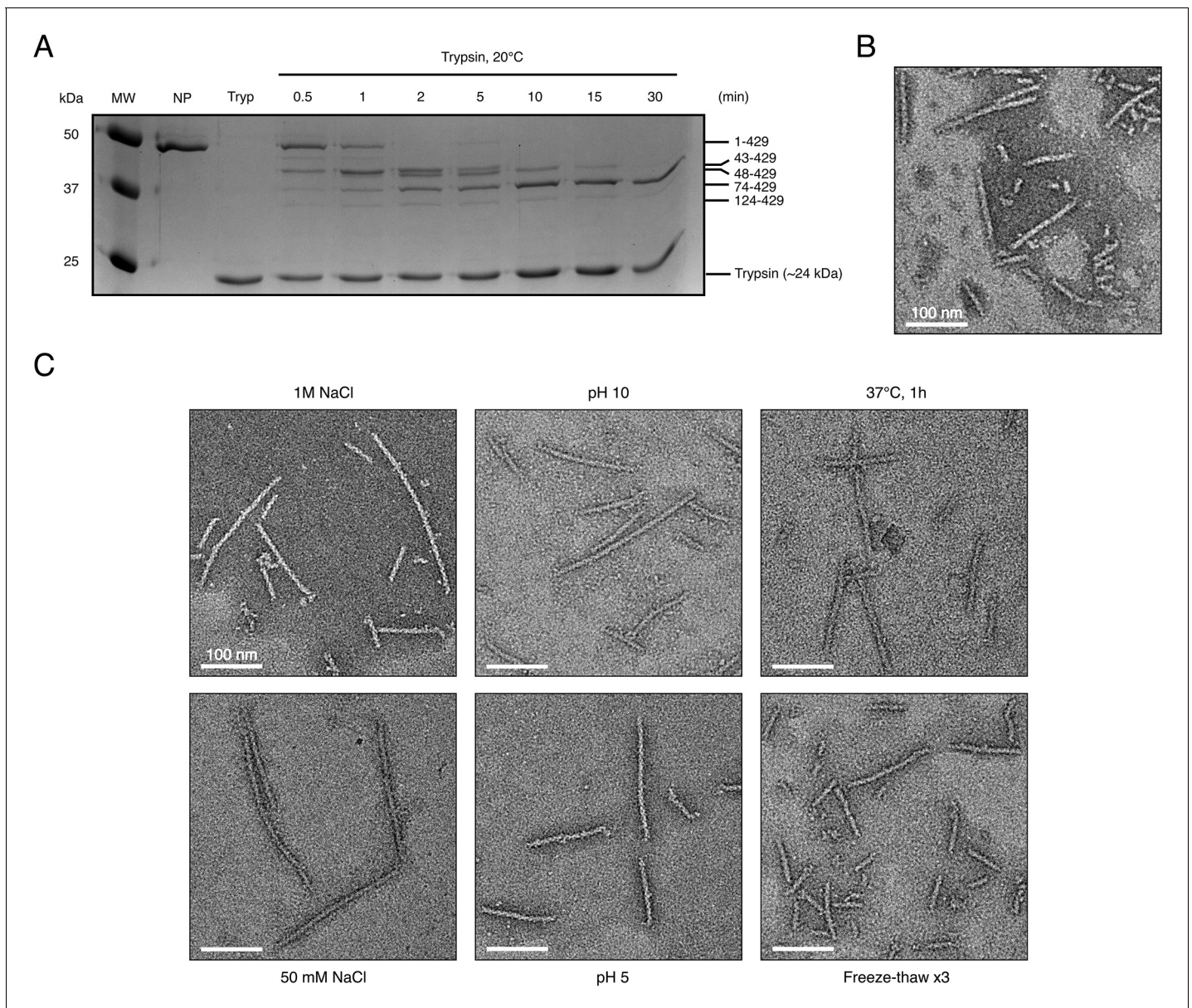


Figure 2—figure supplement 1. Limited proteolysis and stability of recombinant HTNV-NCs. (A) SDS-Page analysis of trypsin digested HTNV-NC. MW represents the molecular weight markers. NC were incubated with trypsin (Tryp) at 20°C for different duration from 0 (NP) to 30 min. After 30 min, a stable construct containing residues 74 to 429 is obtained. Presence of intermediate limited proteolysis fragments containing residues 43–429 and 48–429 strongly suggests that the N-terminal residues 1–73 are not disordered and might correspond to a coiled-coil as observed by *Boudko et al. (2007)*. (B) Negative stain EM micrograph of truncated construct NP₇₄₋₄₂₉ displays NC similar to the ones of full-length NP. (C) Negative stain micrographs of full-length HTNV-NCs incubated in different pHs and salt concentration as indicated. NC fibers stay stable in harsh conditions. Consecutive freeze-thaw breaks NC in small rigid parts. Scale bars are indicated.

DOI: <https://doi.org/10.7554/eLife.43075.010>

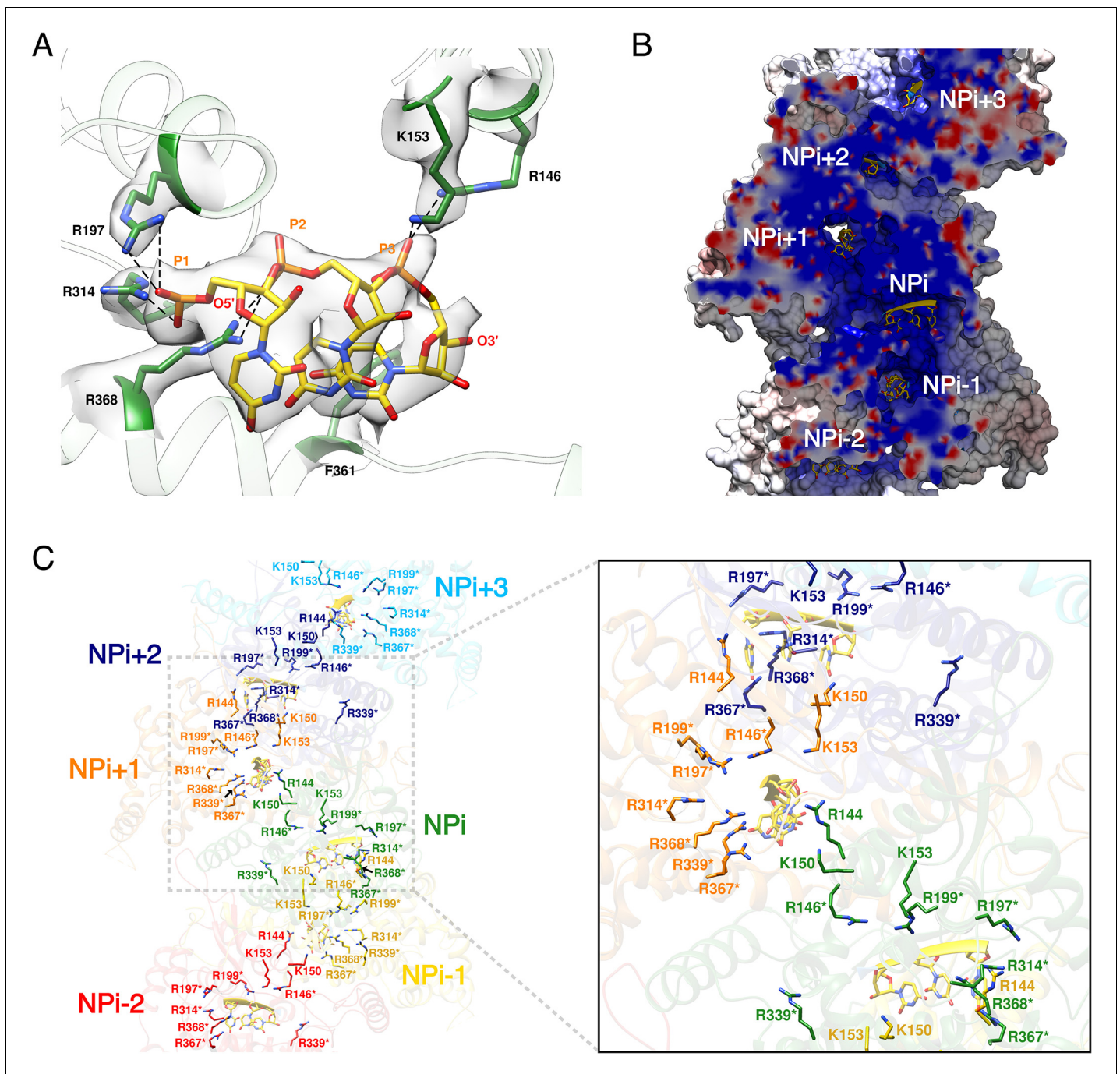


Figure 3. HTNV-NC RNA binding site. (A) RNA binding mode of 3 nucleotides. EM density of RNA and RNA-binding residues are displayed in transparent grey. Hydrogen bonds are shown as dotted lines. (B) Cut-away view of NC electrostatic potential showing the continuous positively-charged groove. RNA nucleotides are displayed. (C) RNA-binding residues are shown as sticks and coloured per subunit. RNA-binding residues defined in *Guo et al. (2016)*, namely R146*, R197*, R199*, R314*, R339*, R367*, R368* are labelled with stars, while RNA-binding residues identified in the present structure, namely R144, K150, K153 are shown without stars.

DOI: <https://doi.org/10.7554/eLife.43075.012>

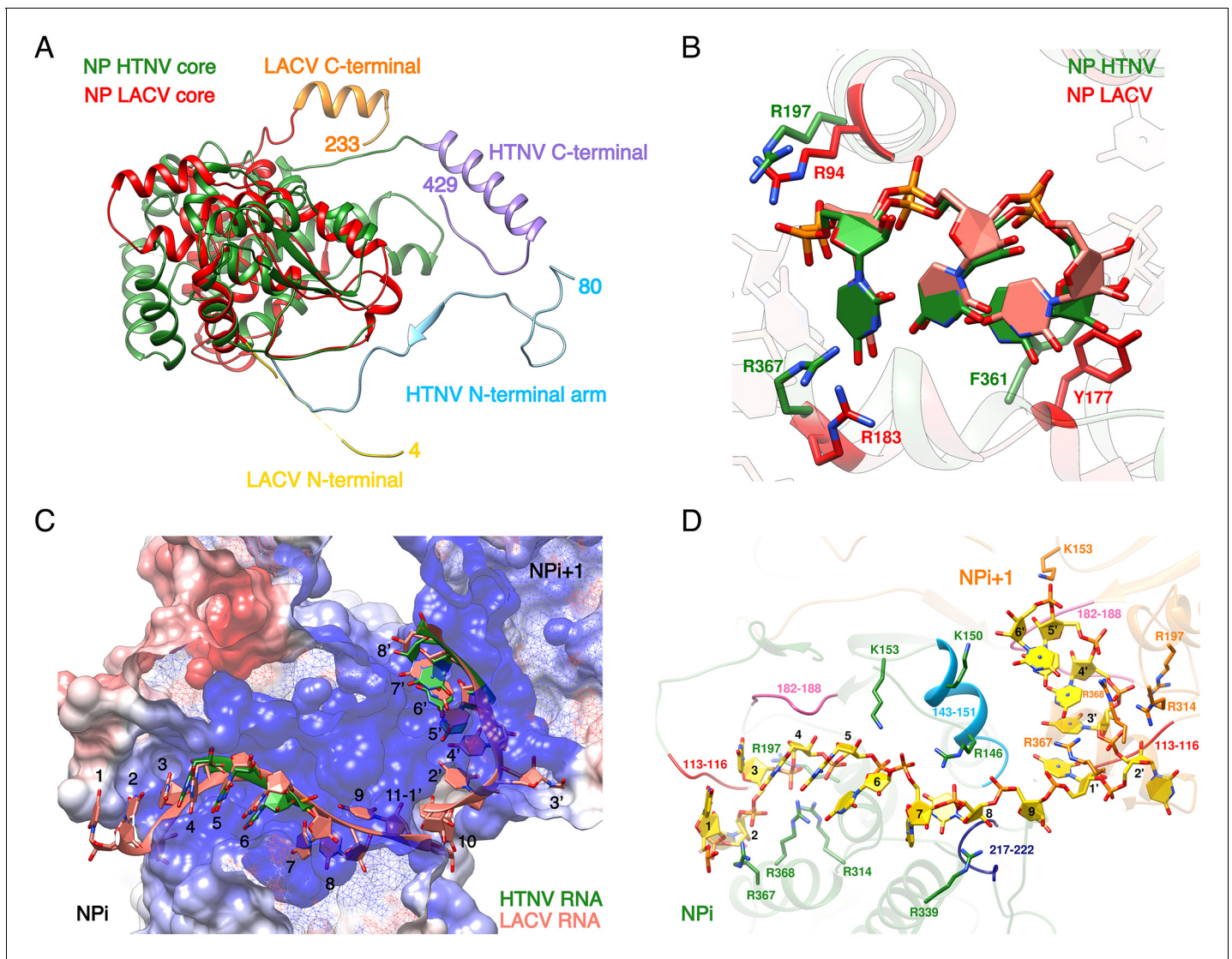


Figure 3—figure supplement 1. Comparison of LACV and HTNV-NP/NC, RNA-binding model. (A) Superimposition of HTNV and LACV NP_{cores}. HTNV and LACV Nt_{arms} and Ct_{arms} contain the same secondary structures but LACV arms are significantly shorter and differ in orientation. (B) HTNV-NP-bound nucleotides adopt a similar conformation as LACV-NP-bound nucleotides 5 to 7. The nucleotide binding pocket is conserved in HTNV-NP and LACV-NP. (C) Cut-away view of HTNV-NC electrostatic potential. Visible HTNV nucleotides are shown in green. LACV nucleotides position based on superimposition of LACV and HTNV NP_{cores} are shown in salmon. LACV RNA fits reasonably well into HTNV-NC RNA path, however LACV nucleotide 11 from NP_i clashes with LACV nucleotide one from NP_{i+1} showing that each HTNV-NP binds less than 11 nucleotides. LACV nucleotides numbering is indicated. (D) Model of HTNV-NC bound to nine nucleotides. Nucleotides numbering from 1 to 9 is indicated. Residues interacting with riboses and phosphates are shown as sticks. Regions interacting with bases are coloured and specifically labelled.

DOI: <https://doi.org/10.7554/eLife.43075.013>

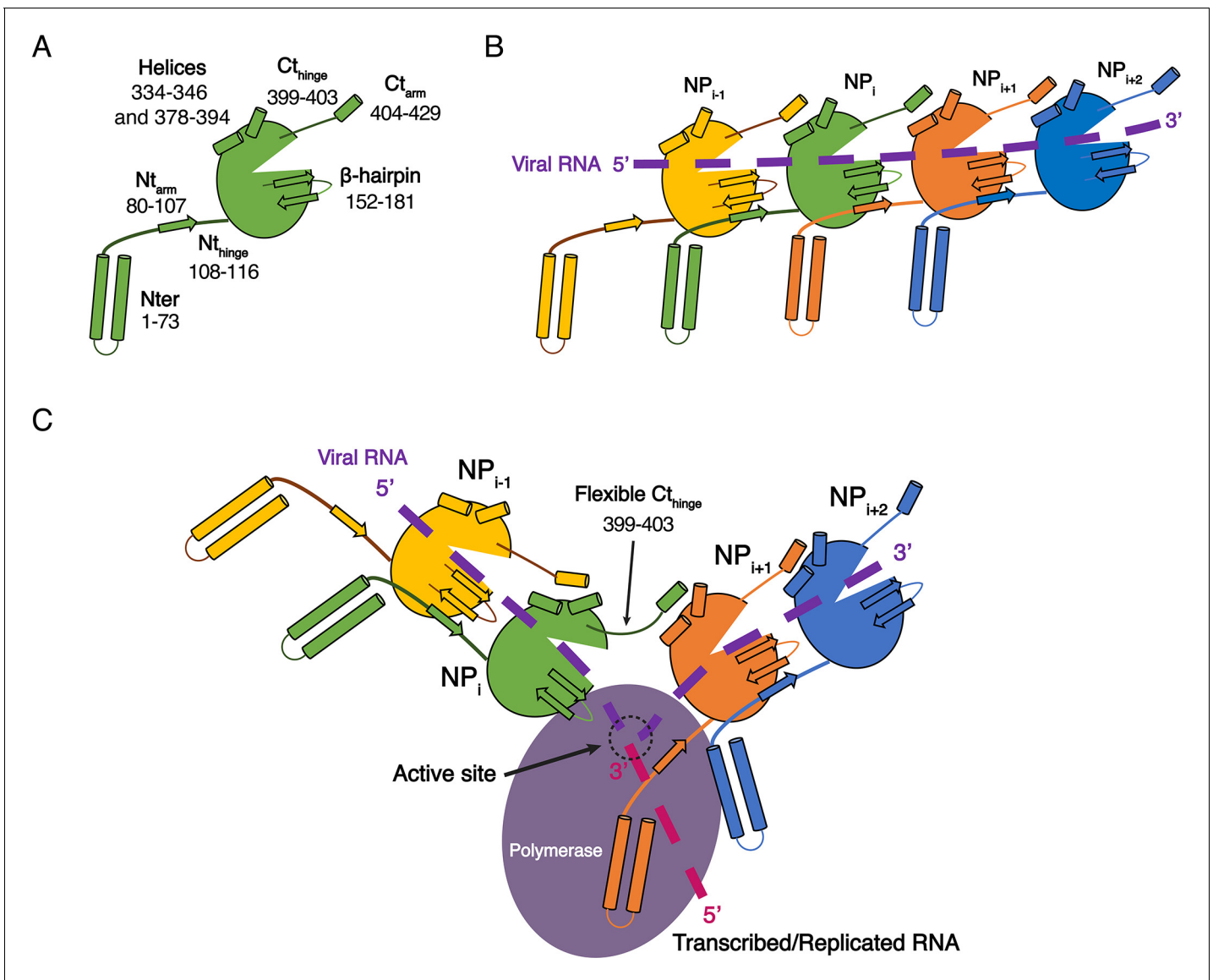


Figure 4. Model of HTNV replication and transcription. (A) Schematics representation of HTNV-NP. Major secondary structures involved in interprotomer interactions are represented as arrows for β -strand and cylinders for α -helices. RNA binding cavity is represented as a clipped part from the NP_{core} region (green oval). (B) Schematic representation of RNA binding and NP-NP interactions. RNA is shown as a dotted purple line. Main NP-NP interactions between adjacent subunits are indicated. For clarity, NP_i interactions with NP_{i-2}, NP_{i-3}, NP_{i+2} and NP_{i+3} are absent from the schematic representation. (C) Replication working hypothesis model which is inspired from [Gerlach et al. \(2015\)](#). The polymerase is shown in purple and the newly transcribed/replicated RNA is indicated as a pink dotted line. The model proposes that the polymerase binds to the flexible N-terminal₁₋₇₃ region during replication/transcription in order to move along the NC. This would be reminiscent to P and L binding to flexible sNSV C-terminal region. Binding of the polymerase to the N-terminal₁₋₇₃ region could destabilise the adjacent Nt_{arm} binding. This localised disruption of NP-NP interaction would create a local opening of the NC enabling transient access of the polymerase to few RNA nucleotides for replication/transcription. Such an opening could be possible without disturbing the whole NC as NP to NP contacts are driven not only by the Nt_{arm} but also by the Ct_{arm}. The Ct_{arm} interaction is likely to remain intact even with a local opening of the NC as the C-terminal hinge allows the Ct_{arm} to undergo large rotation.

DOI: <https://doi.org/10.7554/eLife.43075.014>

1.3. Potentiel mécanisme de déstabilisation de la HTNV-NC

Les deux avantages principaux dont j'ai tiré parti afin de résoudre la structure de la HTNV-NC sont leur importante rigidité et leur grande stabilité (Figure 43).

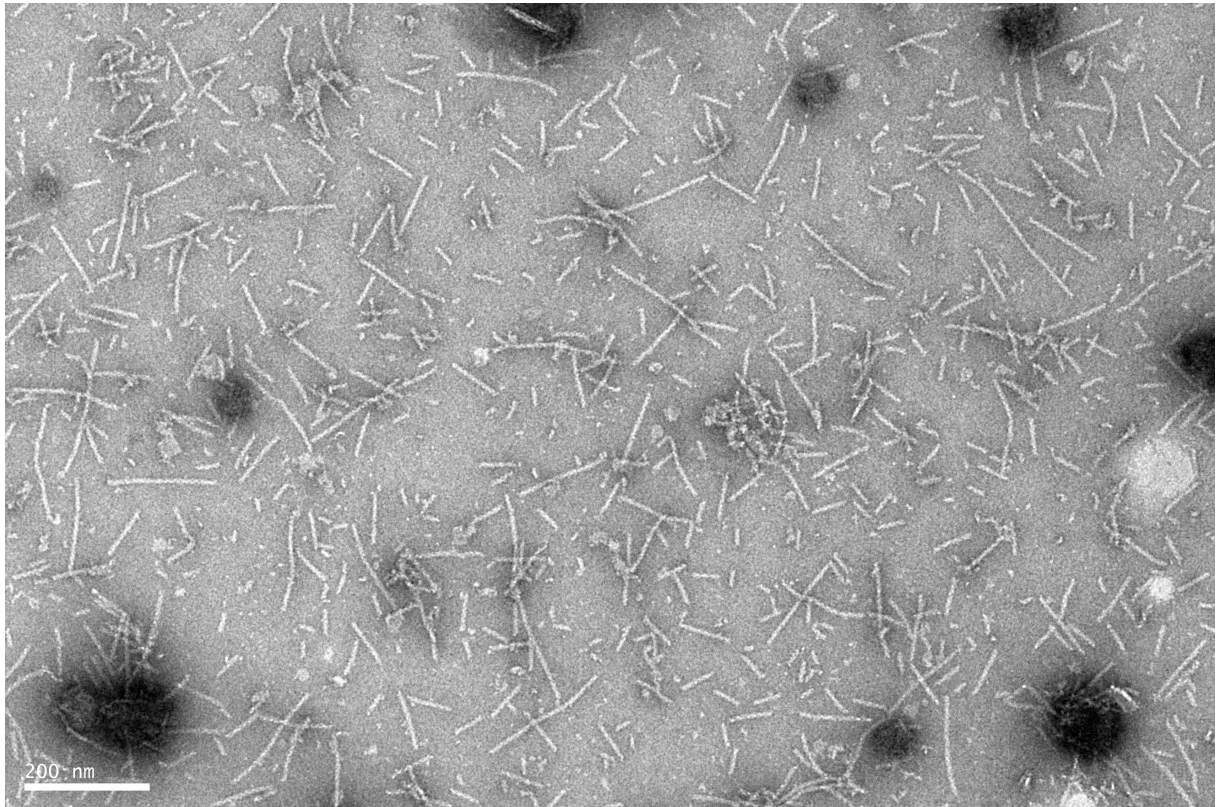


Figure 43. La HTNV-NC observée par coloration négative en ME. Les NCs sont stables et rigides. (Barre d'échelle = 200 nm).

In vivo, ces assemblages hélicoïdaux doivent cependant prendre différentes conformations afin de permettre (i) la circularisation des NCs nécessaire pour lier la polymérase aux extrémités 5' et 3' du génome et (ii) l'empaquetage des NCs dans de nouveaux virions (Huiskonen *et al.*, 2010) qui ont une taille physique ne permettant l'encapsidation que d'une seule NC rigide. De plus, dans le cadre de l'unité fonctionnelle du virus, les RNPs, la protéine L doit également progresser le long de la NC afin de procéder à la synthèse du produit de réplication et de transcription tout en évitant d'exposer le génome viral à l'environnement extérieur. Ces éléments laissent donc supposer qu'une ou plusieurs déstabilisations au sein des interactions présentes entre les NPs survien(nen)t, mais comment ?

Pour répondre à cette question, j'ai cherché à comprendre quels étaient les interactions susceptibles de déstabiliser les interactions entre les NPs sans complètement abolir la formation des NCs. Alors que les interactions entre les NPs successives (NP_i avec NP_{i+1} et NP_{i-1}), via leurs extrémités N- et C-terminales, semblent être indispensables au maintien de l'organisation globale des NCs, les interactions de « moyennes distances » entre NP_{i-3}/NP_i , NP_{i-2}/NP_i , NP_i/NP_{i+2} et NP_i/NP_{i+3} pourraient être impliquées dans le maintien de la rigidité de l'hélice sans pour autant être indispensables au maintien du complexe. En analysant la structure des HTNV-NCs recombinantes, je suis parvenu à l'hypothèse que la perturbation de l'interaction entre les NP_{i+2}/NP_{i+3} et la pointe de l'épingle à cheveux β (162-175) de la NP_i pourrait rendre la NC plus flexible sans la détruire. Afin de vérifier cette hypothèse, une nouvelle construction de la NP a été clonée. Les résidus 165 à 174 ont été supprimés et remplacés par une glycine (HTNV-NP $\Delta\beta$ -hairpin₁₆₅₋₁₇₄). La construction résultante a ensuite été exprimée puis purifiée en suivant le même protocole précédemment décrit dans Arragain *et al.*, 2019. J'ai réalisé ces expériences avec une étudiante en M1 que j'ai co-encadré, Camille Manière.

L'observation par coloration négative et cryo-ME de la HTNV-NC $\Delta\beta$ -hairpin₁₆₅₋₁₇₄ a permis de conclure deux choses. Premièrement, la délétion des résidus de l'épingle à cheveux β n'empêche pas la formation des NCs lors de son expression. Deuxièmement, les NCs formées à partir de cette construction de NP disposent d'une flexibilité beaucoup plus importante que les HTNV-NCs originales (Figure 44).

Ces observations laissent donc supposer qu'une potentielle déstabilisation *in vivo* de l'interaction entre l'épingle à cheveux β de NP_i et NP_{i+2}/NP_{i+3} est possible. Elle pourrait être induite par la HTNV-L qui, lors de la lecture de l'ARN, entrainerait la déstabilisation successive des différentes épingles à cheveux β de chacune des NPs. Ceci pourrait permettre à la polymérase d'accéder de façon transitoire à l'ARN afin de le répliquer ou de le transcrire. Ceci demeure bien-sûr une hypothèse à confirmer/infirmier dans le cadre de l'étude des RNPs.

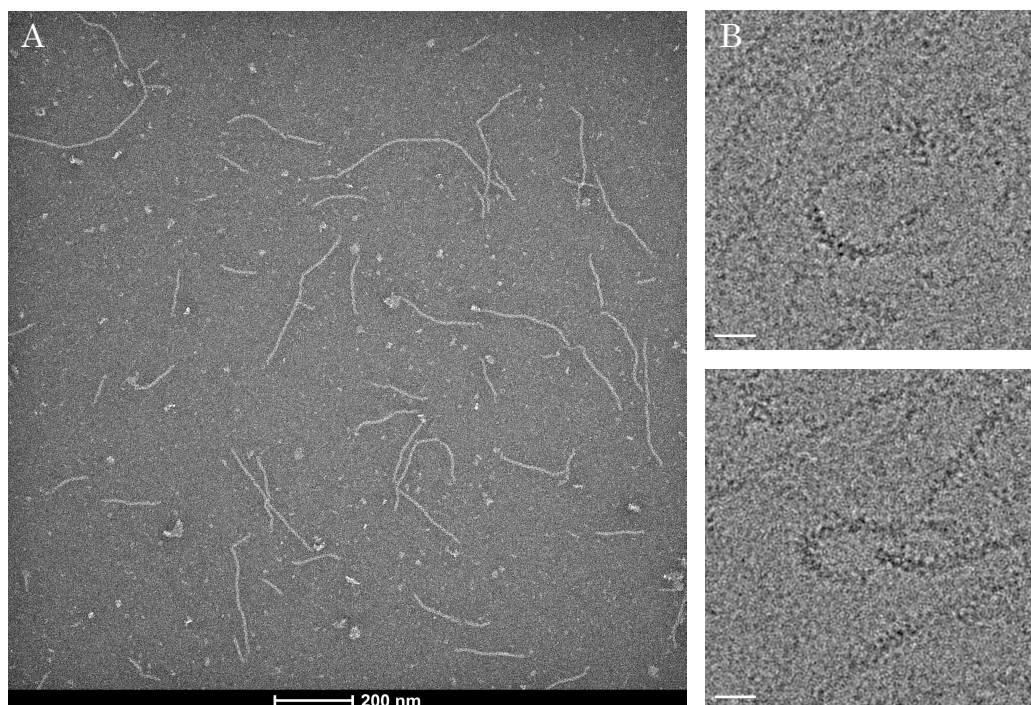


Figure 44. Flexibilité de la HTNV-NC $\Delta\beta$ -hairpin₁₆₅₋₁₇₄ observée par coloration négative et cryo-ME. **A**, Image de coloration négative. **B**, Image de cryo-ME. (Barre d'échelle = 200 Å).

Cette expérience m'a montré qu'afin de comprendre le fonctionnement de la réplication et de la transcription, il était nécessaire de s'attaquer à la caractérisation structurale et fonctionnelle de la protéine L, l'enzyme centrale catalysant ces deux mécanismes. Initialement, je me suis donc intéressé à l'analyse structurale par cryo-ME de la HTNV-L en complexe avec les extrémités 5' et 3' de l'ARN_V.

1.4. Optimisation de l'expression, de la purification et de la congélation de la HTNV-L

Afin de procéder à l'analyse structurale de la HTNV-L, la protéine L a tout d'abord été exprimée en cellules d'insectes puis purifiée à partir d'un protocole préexistant pour la LACV-L (Gerlach et al., 2015). Le plasmide pFastBac contenant la séquence de la HTNV-L (UniProt – P23456) a été fourni par le Dr Juan Reguera et le Dr Stephen Cusack.

1.4.1. Expression en cellules d'insectes

Après transformation des cellules DH10 par choc thermique, un criblage bleu/blanc a été réalisé pour identifier les cellules qui ont intégré le plasmide d'intérêt au sein du génome baculoviral. La transfection du génome dans des cellules Sf21 a été réalisée avec l'agent de transfection X-tremeGene (Sigmaaldrich). La production de protéine, relative à la production de *Yellow Fluorescent Protein* (YFP), a été évaluée par mesure de fluorescence à 529 nm. Les expressions ont été réalisées en utilisant des cellules Hi5 pour la HTNV-L. Les cellules ont été infectées avec 0.1% de virus à une concentration de 0.5×10^6 cellules/mL. La concentration des cellules et la mesure de fluorescence YFP ont été effectuées de manière quotidienne. L'expression a été arrêtée 72h après le jour d'arrêt de la prolifération (DPA). Les cellules ont été récoltées par centrifugation de 15 min à 1000 g et 4°C, congelées dans de l'azote liquide et conservés à -80°C.

1.4.2. Purification de la HTNV-L

La lyse cellulaire a été réalisée par sonication pendant 3 min 30 sec (10 sec ON, 20 sec OFF, amplitude 50 %) sur glace dans un tampon de lyse (50 mM Tris-HCl pH 8, 300 mM NaCl, 20 mM Imidazole, 2 mM TCEP, 10 % glycérol) avec des inhibiteurs de protéase sans EDTA.

Le lysat a été centrifugé à 20 000 RPM pendant 45 min à 4°C. Les protéines solubles ont été sujettes à une précipitation au sulfate d'ammonium (NH_4SO_4 à 0.5 mg/ml), puis centrifugées à 30 000 RPM pendant 45 min, à 4°C. Le surnageant a été éliminé et les protéines ont été resuspendues dans le même tampon/volume de lyse initial et centrifugées à nouveau en utilisant les conditions précédemment décrites.

La HTNV-L, présente dans le surnageant, a ensuite été purifiée par chromatographie d'affinité Ni-NTA après une étape de lavage en utilisant un tampon de lyse avec une concentration finale d'imidazole de 50 mM et enfin, éluée en utilisant le tampon de lyse avec une concentration finale d'imidazole de 400 mM.

Les différentes fractions de la HTNV-L ont été regroupées et dialysées pendant 1h à 4°C dans un tampon « héparine » (50 mM Tris-HCl pH 8, 250 mM NaCl, 2 mM TCEP, 10% de glycérol). Les protéines ont ensuite été chargées sur une colonne héparine et éluées en utilisant 50 mM Tris-HCl pH 8.5, 500 mM NaCl, 10 mM TCEP, 5% de glycérol. Les fractions contenant le plus de protéines ont été directement chargées sur une colonne de chromatographie d'exclusion de taille S200 (SEC), sans étape de concentration, dans un tampon final contenant 20 mM Tris-HCl pH 8, 300 mM NaCl, 2 mM TCEP.

La pureté et l'homogénéité de l'échantillon ont ensuite été caractérisées par gel SDS-PAGE et par coloration négative. La protéine s'est finalement révélée être de qualité suffisante pour passer à l'optimisation des conditions de congélation dans le but de déterminer une structure à haute résolution par cryo-ME (Figure 45).

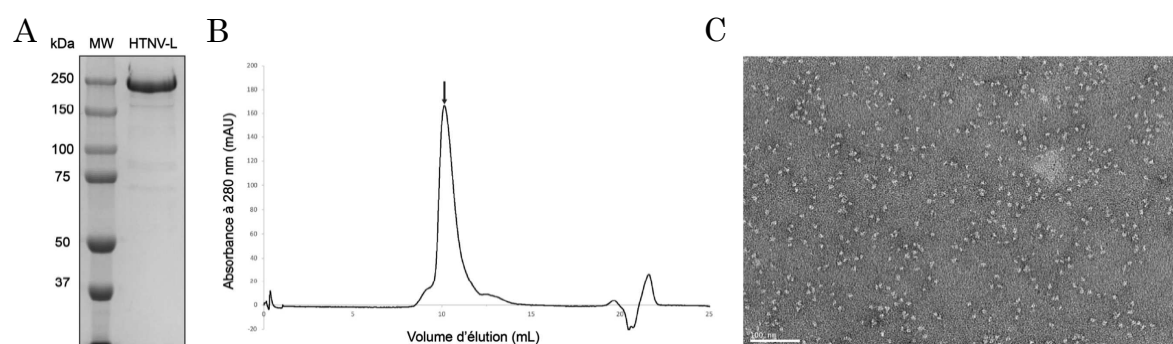


Figure 45. Optimisation de l'expression et de la purification de la HTNV-L. **A**, Gel SDS-PAGE 6% après une étape de chromatographie d'exclusion de taille (SEC). **B**, L'absorbance à 280 nm (mAU) en fonction du volume d'élué. La flèche correspond à la fraction déposée sur gel et à la micrographie de coloration négative présentée en **C**. **C**, La HTNV-L observée par coloration négative. (Barre d'échelle = 100 nm).

Les premières images obtenues par cryo-ME ont cependant montré que la HTNV-L avait tendance à s'agréger, phénomène potentiellement causé par la ségrégation des protéines à l'interface air-eau lors de l'élimination de l'excès de liquide qui précède la congélation de l'échantillon (D'Imprima *et al.*, 2019). De plus, les analyses d'images effectuées ont également mis en évidence la présence d'une orientation préférentielle importante (Figure 46) (Lyumkis, 2019). Afin de pallier ces problèmes, de nombreux essais d'optimisations ont été réalisés concernant les compositions des tampons, les conditions de congélation et les différents types de grilles utilisées.

Au niveau de la stabilisation de la protéine, de nombreuses compositions de tampons (concentrations en sels, mise en présence des extrémités 5' et 3' de l'ARNv, stabilisant à faible concentration comme le glycérol < 1%) ont été testées via des expériences de *Thermal Shift Assay* ou via des kits de criblage (*RUBIC buffer screen* et *RUBIC additive screen*) sur la plateforme HTX de l'EMBL (données non présentées). Afin de limiter les interactions à l'interface air-eau lors de la congélation, l'ajout de détergents (CHAPSO, NP-40) a également été testé. Malheureusement, aucun de ces tests n'a permis d'améliorer la qualité de l'échantillon en cryo-ME (Figure 46).

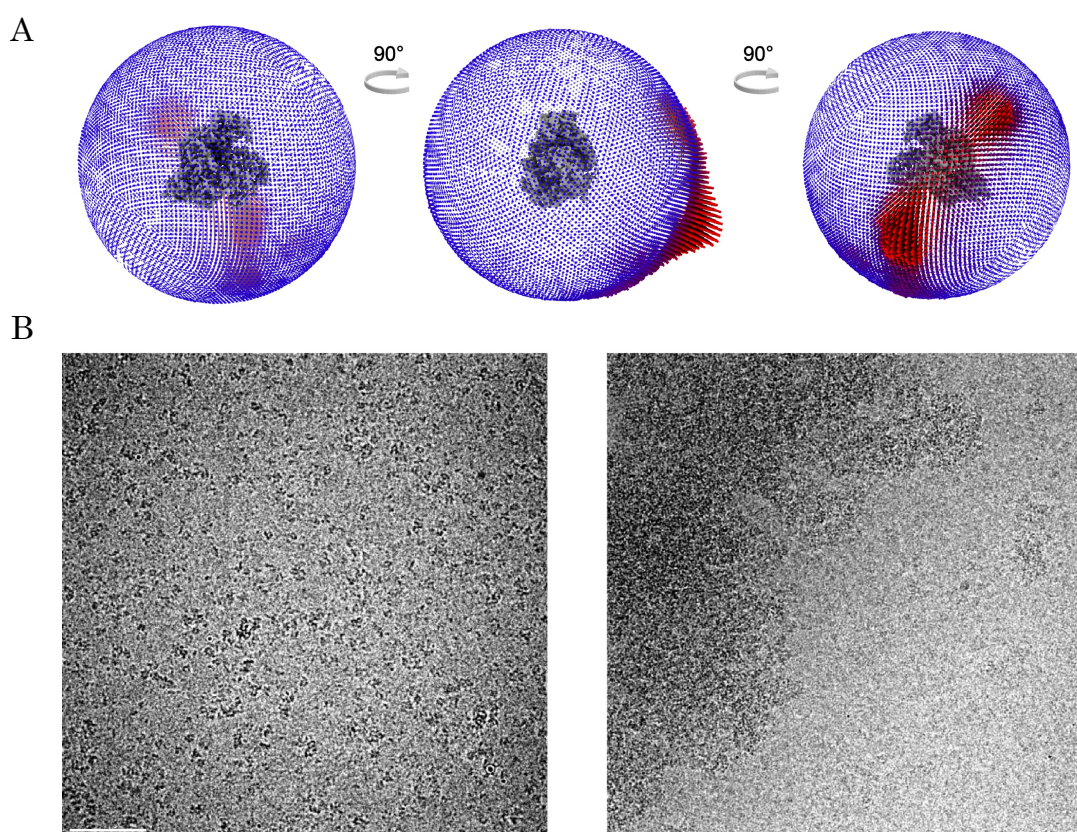


Figure 46. Les problèmes d'orientations préférentielles et de congélation de la HTNV-L. **A**, Les particules présentent une orientation préférentielle prononcée. **B**, Hétérogénéité importante (bonne répartition vs agrégats) en cryo-ME de la HTNV-L malgré l'optimisation des conditions de tampons et de congélation. Les deux images proviennent de la même grille. (Barre d'échelle = 100 nm).

Différents types de grilles ont ensuite été testées (Au/Au, Au/C, Cu/Rh, Cu/C) avec des tailles de mailles différentes (2/2, 1.2/1.3, 0.6/1). L'utilisation de grilles recouvertes d'un film de carbone de 2 nm ou l'application manuelle d'un film d'oxyde de graphène (Palovcak *et al.*, 2018) a également été testée pour pallier ce problème d'orientation préférentielle. Enfin, les conditions de congélation ont été

optimisées en modifiant le temps de blot, la force du blot, le temps de drainage et la température de congélation. Ainsi, de toutes les conditions testées, les seuls paramètres retenus consistent à congeler la HTNV-L fraîchement purifiée, sans étape de concentration préalable et d'effectuer la congélation à 4°C en utilisant des grilles Au/Au de maille 1.2/1.3 (Figure 46).

Il est important de mentionner que le criblage de toutes les conditions précédemment mentionnées par cryo-ME a été facilité par l'installation sur la plateforme de ME de l'IBS, en décembre 2018, du MET ThermoFisher Glacios, initialement équipé d'une Flacon II et du logiciel de collecte automatique EPU.

1.5. Caractérisation structurale de la HTNV-L liée à l'extrémité 5' de l'ARN_v en cryo-ME

Basé sur les précédents résultats obtenus chez la LACV-L (*Peribunyaviridae*) (Gerlach *et al.*, 2015), il a été montré qu'en raison de la forte complémentarité de séquence des extrémités 5' et 3' de l'ARN_v, leur hybridation *in vitro* les empêche de se fixer à leur site de liaison respectif. Ce problème est également présent pour les extrémités du génome viral du HTNV (*Hantaviridae*) et plus généralement chez les autres familles de l'ordre des *Bunyavirales*.

Pour pallier ce problème, l'une des solutions consistait à muter certains nucléotides du 5' *hook* qui interagissent avec la HTNV-L afin de diminuer le degré de complémentarité avec l'extrémité 3' de l'ARN_v. Bien évidemment, il est nécessaire de conserver sa structure secondaire nécessaire à la stabilisation des différents motifs catalytiques présents au sein de la RdRp. C'est pourquoi seuls les nucléotides interagissant avec la HTNV-L via leur groupements phosphates peuvent être mutés. Initialement, il était donc nécessaire de déterminer une première structure de la HTNV-L en complexe avec seulement l'extrémité 5' de l'ARN_v afin de déterminer son mode de fixation à la protéine L et pouvoir en déduire quels nucléotides muter.

Plusieurs reconstructions de la HTNV-L ont ainsi été obtenues à l'aide d'images collectées sur le Glacios, alors équipé d'une Falcon II. Malheureusement, cette caméra fonctionnant en mode linéaire ne permettait pas l'étude à haute

résolution ($< 4\text{\AA}$) d'objet asymétrique de masse d'environ 250 kDa. Par la suite, un jeu de données a donc été collecté sur le Titan Krios de l'ESRF équipé d'un filtre d'énergie et d'une K2 et m'a permis d'obtenir deux structures à des résolutions globales de 3.4\AA et 3.6\AA . Au sein de la première structure, la HTNV-L est liée à l'ARNv 5' 1-25 (Figure 47) tandis que dans la deuxième, elle ne l'est pas (forme *apo*). Il faut également noter que les structures obtenues sont légèrement anisotropiques en raison du problème d'orientation préférentielle précédemment mentionné.

Malgré la déformation de la densité, les résidus 227 à 1454, appartenant au cœur de la polymérase, ont pu être construits. Cela représente environ 60% de la totalité de la protéine, qui en contient 2151. La partie N-terminale (1-226) qui inclut le domaine ENDO (1-180) ainsi qu'une partie du domaine reliant cette dernière au cœur de la protéine (181-261) est manquante, tout comme la partie C-terminale qui contient notamment le CBD. Une partie des domaines du cœur de la

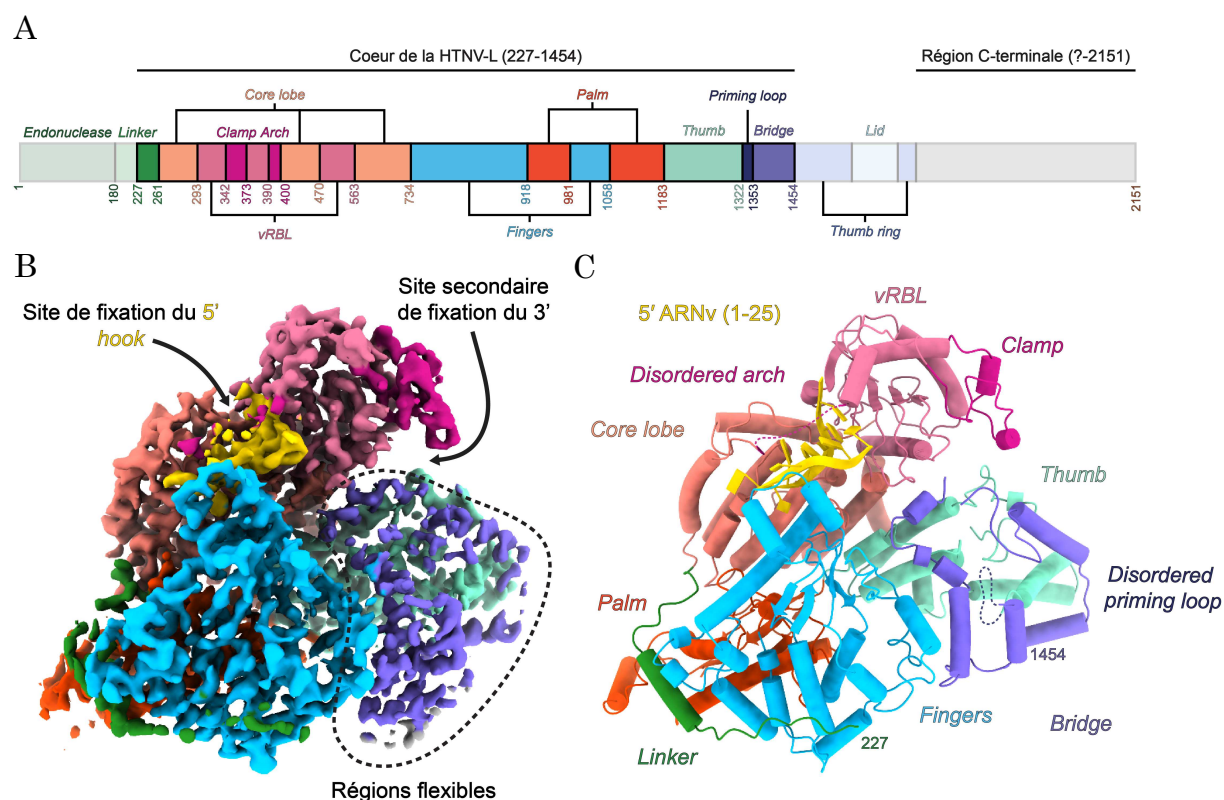


Figure 47. Structure de la HTNV-L résolue par cryo-ME. **A**, Représentation schématique des différents domaines de la HTNV-L. **B**, Structure de cryo-ME à 3.4\AA de résolution de la HTNV-L liée à l'extrémité 5' 1-25 de l'ARNv. Les différents domaines sont colorés comme en A et l'ARNv est colorée en jaune. **C**, Modèle atomique de la HTNV-L. Les limites N- et C-terminales du modèle construit sont indiquées et les domaines ainsi que l'ARN sont colorées comme en A et B.

HTNV-L, *i.e* le *lid*, le *thumb ring* et la *priming loop*, sont également trop flexibles pour présenter des densités me permettant de construire la structure de ces domaines dans leur totalité (Figure 47).

Les domaines *palm*, *fingers* et *thumb*, qui contiennent les différents motifs conservés et impliqués dans la réaction catalytique propre aux RdRps sont présents et forment une cavité interne. Dans la structure apo, le domaine des *finger tips* appartenant au motif F est flexible, *a contrario* de la structure qui lie l'extrémité 5' de l'ARNv. Dans cette dernière, les onze premiers nucléotides du 5' sont fixés à la HTNV-L et prennent la forme d'un crochet (*hook*) mais la résolution ne permet pas de construire un modèle de façon non ambiguë, deux modèles hypothétiques et différents pouvant être construits dans la carte (Figure 48).

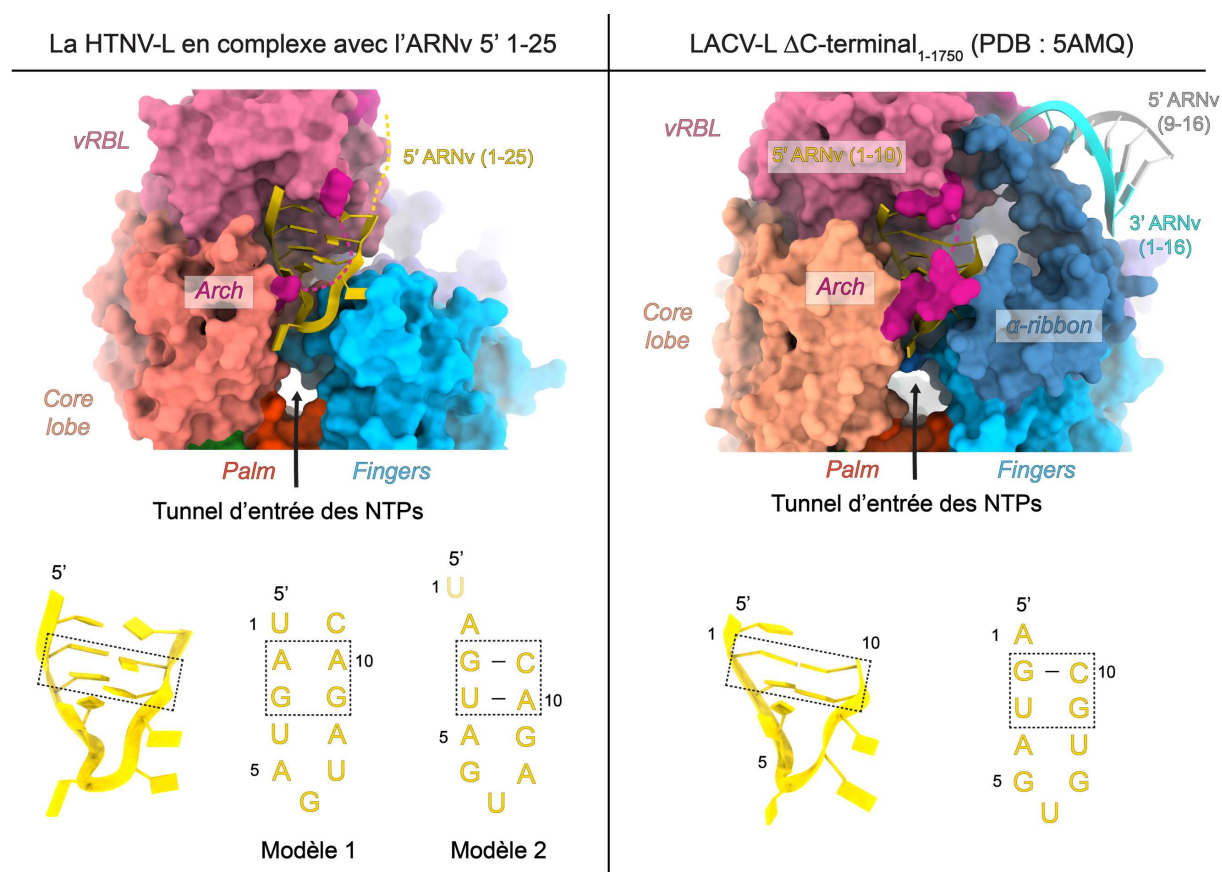


Figure 48. Comparaison de la fixation du 5' *hook* chez la HTNV-L en complexe avec l'ARNv 5' 1-25 et la LACV-L₁₋₁₇₅₀ Δ C-terminal (Gerlach *et al.*, 2015). Les différents domaines sont colorés et montrés comme isosurface. Les deux modèles potentiels concernant la fixation de l'ARNv 5' 1-25 de la HTNV-L sont indiqués. Pour chacun des 5', les bases susceptibles d'être mutées afin de limiter la complémentarité au 3' de l'ARNv sont encadrés.

La HTNV-L dispose également d'un site de liaison spécifique à l'extrémité 3' de l'ARNv, localisé à la surface de la protéine et correspondant au site B (ou site secondaire). Celui-ci est surplombé par les domaines *clamp* et *vRBL* (Figure 47).

Les deux reconstructions obtenues pour la HTNV-L présentent ainsi de nombreuses similitudes avec les structures déjà déterminées des protéines L chez les *Bunyavirales* (Gerlach *et al.*, 2015; Peng *et al.*, 2020; Wang *et al.*, 2020) ou les polymérasés viraux hétérotrimériques des *Orthomyxoviridae* (Pflug *et al.*, 2014; Reich *et al.*, 2014). Leurs organisations générales, que ce soit au niveau du cœur catalytique ou au niveau des différents sites de fixation spécifique de chacune des extrémités de l'ARNv localisés à la surface de la protéine L sont structurellement conservées (Figure 48 et 49).

La structure secondaire formée par l'extrémité 5' de l'ARNv et sa poche de fixation sont également similaires à ce qui a déjà été observé chez d'autres polymérasés des sNSVs. De plus, la structuration du motif F présent au niveau du site catalytique par la fixation du 5' a déjà été démontré chez la LACV-L₁₋₁₇₅₀ Δ C-terminal (Gerlach *et al.*, 2015). Malgré tout, la séquence des nucléotides du 5' au sein du *hook* reste à être déterminée de manière précise, l'une ou l'autre des solutions pouvant être considérées. Dans le but de limiter la complémentarité entre le 5' et le 3', deux options sont donc envisageables quant à la mutation des nucléotides interagissant avec la HTNV-L via leur groupement phosphates (Figure 48).

Enfin, on notera également l'absence du domaine *α -ribbon* qui, chez les *Peribunyaviridae* permet de séparer les deux extrémités de l'ARNv (Figure 48) (Gerlach *et al.*, 2015), et chez les *Orthomyxoviridae* (*β -ribbon*) de jouer un rôle stabilisateur de l'ARNv entre les étapes d'initiation et d'élongation de la transcription (Wandzik *et al.*, 2020). Chez les *Hantaviridae*, l'absence de ce domaine pourrait avoir un impact sur les différentes activités de répllication et de transcription du génome viral.

1.6. La HTNV-L et la LACV-L

Les deux structures obtenues de la HTNV-L sont malheureusement incomplètes en raison de l'hétérogénéité importante de l'échantillon au moment de la congélation et n'offrent aucune nouvelle information structurale nouvelle quant aux différents domaines C-terminaux de cette protéine. De ce fait, en parallèle des travaux réalisés sur la HTNV-L, je me suis également intéressé à l'étude structurale de la LACV-L. Ces deux protéines, bien qu'appartenant à deux virus de familles virales différentes (*Hantaviridae* et *Peribunyaviridae*) partagent respectivement 47% et 17% de similarité et d'identité de séquence (Figure 49).

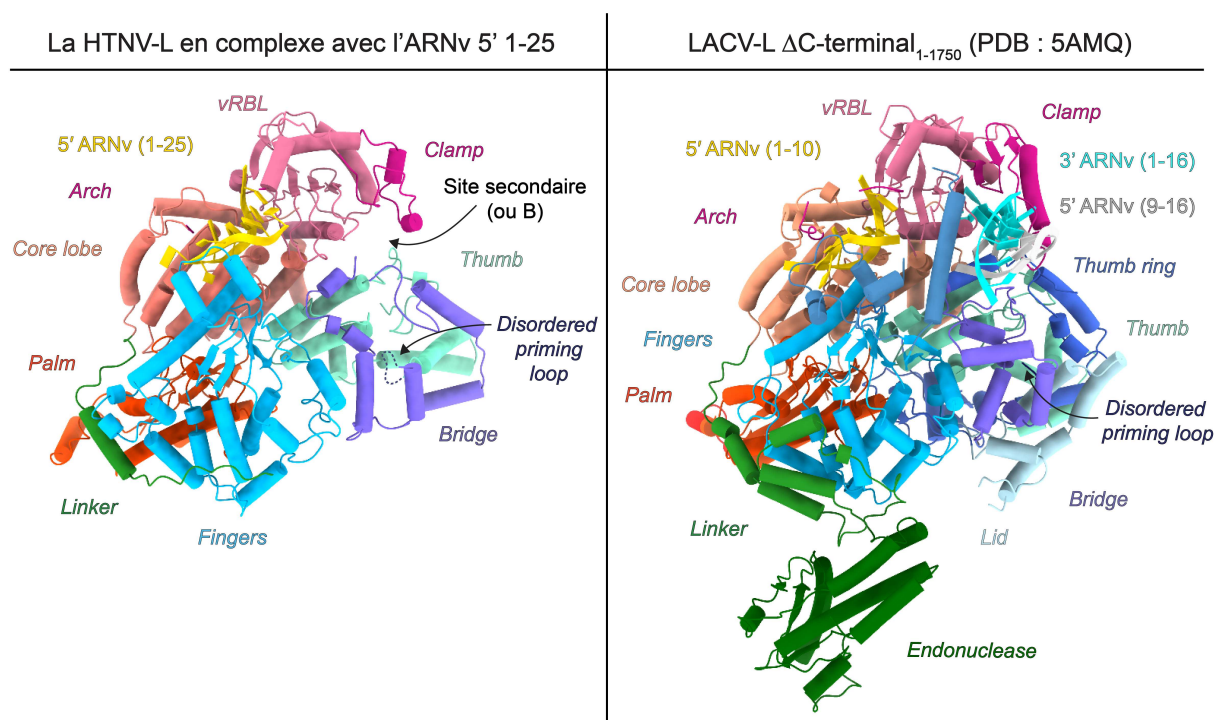


Figure 49. Comparaison de l'organisation générale de la HTNV-L en complexe avec l'ARNv 5' 1-25 et la LACV-L₁₋₁₇₅₀ ΔC-terminal (Gerlach *et al.*, 2015). Chacun des domaines sont colorés de manière identique entre les deux modèles.

Malgré ces hauts pourcentages de similarité et d'identité, la LACV-L bénéficie d'une stabilité importante et d'un bien meilleur comportement lors de sa congélation comme en attestent les différentes reconstructions 3D obtenues par cryo-ME et la résolution des structures de la LACV-L₁₋₁₇₅₀ ΔC-terminal par cristallographie aux rayons X, ces dernières étant tronquées des résidus 1751 à 2263 (ce qui représente 33% de la protéine) (Gerlach *et al.*, 2015). L'étude de ces deux polymérases virales reste donc complémentaire et les informations pouvant

être obtenues en termes de structures ou de fonctions sur l'une, peuvent servir à une meilleure caractérisation de l'autre. Cela est extrêmement bénéfique pour la compréhension générale des mécanismes de réplication et de transcription chez les *Bunyavirales*.

Actuellement, un nouveau doctorant de première année que j'ai en partie co-encadré, Quentin Durieux, a repris le flambeau sur la caractérisation structurale et fonctionnelle de la HTNV-L en disposant des multiples informations accumulées sur la LACV-L, notamment en ce qui concerne les avancées structurales et fonctionnelles obtenues et abordées dans les prochaines sections. Ensemble, ces informations devraient également déboucher sur de nouveaux résultats en ce qui concerne la compréhension des mécanismes de réplication et de transcription chez la famille des *Hantaviridae*.

2. La structure à haute résolution par cryo-ME de la LACV-L

2.1. Introduction

Au cours de mes 2^{ème} et 3^{ème} années de thèse, je me suis donc penché sur l'obtention de la structure entière de la LACV-L afin d'élucider l'organisation de son domaine C-terminal, absent de la structure cristallographique précédemment obtenue (Gerlach *et al.*, 2015).

A l'aide du MET Glacios, toujours équipé d'une Falcon II, j'ai pu obtenir plusieurs structures préliminaires à des résolutions comprises entre 5 et 6 Å. Ces dernières sont dans un état de pré-initiation où (i) l'extrémité 5' 1-10 et (ii) l'extrémité 3' 1-16 de l'ARNv sont respectivement fixées en tant que (i) *hook* et (ii) sur le site B de la LACV-L. Afin d'empêcher l'hybridation *in vitro* du 5' 1-10 et du 3' 1-16, l'extrémité 3' a été hybridée au préalable à un ARNv 5' 9-16. Ces reconstructions m'ont permis de distinguer différents domaines additionnels localisés à la périphérie du cœur de la protéine, mais également de repositionner le domaine ENDO dans une position complètement différente de celle présente dans la structure cristallographique (Gerlach *et al.*, 2015). Afin d'obtenir une reconstruction à plus haute résolution dans le but de construire un modèle des domaines présent en C-terminal de la LACV-L, un jeu de données de plus de 16000 films a été collecté sur le Titan Krios présent à l'ESRF. Il m'a permis, en combinant de multiples étapes de traitement d'images, de déterminer la structure de la LACV-L entière en état de pré-initiation. De plus, et de façon inattendue, l'incubation du 3' 1-16 et du 5' 1-10 qui sont quasiment complémentaires (seule une paire de bases ne l'est pas) a entraîné la formation d'un ARN double-brin qui s'est fixé à l'intérieur de la cavité centrale et de manière proximale au site actif de la RdRp, mimant ainsi un état d'élongation.

Ces structures ont révélé, pour la première fois, l'organisation complète d'une protéine L chez les *Bunyavirales* et ont permis de construire un modèle des différents domaines présents en C-terminal de la LACV-L. Enfin, en comparant les états de pré-initiation et d'élongation, ces structures ont permis de caractériser

les différents changements de conformation existant entre ces deux états et de les comparer avec ceux présents chez la polymérase du virus influenza (Kouba, Drncová and Cusack, 2019).

Ces résultats ont fait l'objet d'un deuxième article publié en tant que premier auteur dans le journal *Nature Communications* en Juillet 2020.

2.2. Article n°2 : Pre-initiation and elongation structures of full-length La Crosse virus polymerase reveal functionally important conformational changes

Pre-initiation and elongation structures of full-length La Crosse virus polymerase reveal functionally important conformational changes

Benoît Arragain ¹, Grégory Effantin¹, Piotr Gerlach ^{2,3}, Juan Reguera^{2,4}, Guy Schoehn ¹, Stephen Cusack ²✉ & Hélène Malet ¹✉

Bunyavirales is an order of segmented negative-strand RNA viruses comprising several life-threatening pathogens against which no effective treatment is currently available. Replication and transcription of the RNA genome constitute essential processes performed by the virally encoded multi-domain RNA-dependent RNA polymerase. Here, we describe the complete high-resolution cryo-EM structure of La Crosse virus polymerase. It reveals the presence of key protruding C-terminal domains, notably the cap-binding domain, which undergoes large movements related to its role in transcription initiation, and a zinc-binding domain that displays a fold not previously observed. We capture the polymerase structure at pre-initiation and elongation states, uncovering the coordinated movement of the priming loop, mid-thumb ring linker and lid domain required for the establishment of a ten-base-pair template-product RNA duplex before strand separation into respective exit tunnels. These structural details and the observed dynamics of key functional elements will be instrumental for structure-based development of polymerase inhibitors.

¹Université Grenoble Alpes, CNRS, CEA, Institute for Structural Biology (IBS), F-38000 Grenoble, France. ²European Molecular Biology Laboratory, Grenoble, France. ³Present address: Department of Structural Cell Biology, Max Planck Institute of Biochemistry, Munich, Germany. ⁴Present address: Aix-Marseille Université, CNRS, INSERM, AFMB UMR 7257, 13288 Marseille, France. ✉email: cusack@embl.fr; helene.malet@ibs.fr

Bunyavirales is a very large and diverse order of segmented negative-strand RNA viruses (sNSV) comprising more than 500 species classified into 12 families¹. It contains serious human pathogens such as La Crosse virus (LACV, *Peribunyaviridae* family), Hantaan virus (HTNV, *Hantaviridae* family), Crimean-Congo hemorrhagic fever virus (CCHFV, *Nairoviridae* family), Rift Valley Fever virus (RVFV, *Phenuiviridae* family), and Lassa fever virus (LASV, *Arenaviridae* family). Viruses from the *Bunyavirales* order are related to other sNSV and in particular to influenza virus, a major human pathogen belonging to the *Orthomyxoviridae* family.

Replication and transcription of sNSV viral genomic segments are performed by the virally encoded RNA-dependent RNA polymerase, also called L protein for *Bunyavirales*². These processes are performed in the cytoplasm of infected cells for Bunyaviruses, whereas they occur in the nucleus for influenza virus^{3,4}. Replication generates full-length genome or antigenome copies (vRNA and cRNA, respectively), whereas transcription produces capped viral mRNA that are recognized by the cellular translation machinery to produce viral proteins. Transcription is initiated by a “cap-snatching” mechanism, whereby host 5′ capped RNAs are bound by the L cap-binding domain (CBD), cleaved by the L endonuclease domain several nucleotides downstream, and then used to prime synthesis of mRNA^{2,5,6}.

Although the overall mechanism of transcription initiation is likely conserved between sNSVs, several elements suggest some divergence between viral families. First, the source of capped RNA differs. Whereas influenza polymerase interacts directly with the host RNA polymerase II to snatch the caps of nascent transcripts in the nucleus⁷, bunyavirus polymerases act in the cytoplasm. It is currently unclear which cytoplasmic capped RNAs are accessed by bunyavirus polymerases, and in what context, and if the polymerase contains specific domains that interact with host capped-RNA-bound proteins. Second, the length of the host-derived capped RNA primer generated after cleavage by the endonuclease differs between families⁵, 0–7 in *Arenaviridae*, 10–18 in *Peribunyaviridae*, 10–14 in *Orthomyxoviridae*, suggesting differences in the relative position of the endonuclease, CBD and polymerase active site. Third, CBD localization within L proteins remains unclear for several viral families primarily due to the absence of a definitive motif for the cap-binding site and because of the high divergence in sequence between polymerases, particularly in their C-terminal region. Identification of the CBD in the C-terminal region of L proteins has however recently been achieved for both Rift Valley Fever virus (RVFV, *Phenuiviridae*) and California Academy of Science virus (CASV, *Arenaviridae*), thanks to the determination of isolated CBD domain structures^{8,9}.

To understand the detailed mechanisms of replication and transcription, structures of the full-length polymerase are essential. Significant advances have recently been made on influenza polymerase with structures stalled at different steps of transcription now being available. They reveal that influenza CBD undergoes a 70° movement to bring the capped RNA from an orientation in which it can be cleaved by the endonuclease to one where it can enter the polymerase active site^{10–12}. A snapshot of transcription elongation has been captured revealing the presence of a nine-base-pair template-product RNA inside the active site cavity that is then separated into two single-stranded RNAs exiting through separated tunnels¹². In comparison, structural information on *Bunyavirales* polymerase remains limited with only the structure of a C-terminally truncated construct of LACV polymerase (residues 1–1750, LACV-L₁₋₁₇₅₀) being currently available¹³. LACV-L₁₋₁₇₅₀ is composed of an N-terminal protruding endonuclease domain (residues 1–185) and a polymerase

core containing the RNA synthesis active site (residues 186–1750). It was solved by X-ray crystallography in the pre-initiation state in complex with the 5′ and 3′ promoter ends. Both promoter ends bind sequence-specifically in separate pockets away from the active site, respectively, called the “5′ end stem-loop pocket” and the “3′ end pre-initiation pocket”. The LACV-L₁₋₁₇₅₀ structure also depicts the presence of an active site cavity with typical polymerase motifs as well as distinct template and product exit tunnels.

To reveal the structure of the C-terminal region of LACV-L and the overall architecture of the complete polymerase, we determined the structure of full length LACV-L (LACV-L FL) by X-ray crystallography and high-resolution cryo-EM. We uncover the structure of LACV-L CBD, which contains a specific insertion allowing interaction with the endonuclease. We find that the extreme C-terminal region of LACV-L FL is a zinc-binding domain (ZBD) that is absent in other sNSV polymerases of known structure and may correspond to a host–protein interaction platform. We also capture snapshots of LACV polymerase in both pre-initiation and elongation-mimicking states, thereby revealing, amongst other conformational changes, the movement of the priming loop that unblocks the active site cavity and permits accommodation of a 10-base-pair template-product duplex, characteristic of elongation.

Results

Structure determination of LACV-L FL protein. LACV-L FL was expressed in insect cells and purified to homogeneity based on the protocol described in Gerlach et al.¹³ (Supplementary Fig. 1a). Slight modifications were however necessary in order to stabilize LACV-L FL, in particular the addition of nucleotides 1–16 of the 3′ vRNA (3′OH-UCAUCACAUGAUGGUU) and complementary 8-mer corresponding to the nucleotides 9–16 of the 5′vRNA (5′OH-GCUACCAA) prior to the decrease to 150 mM of NaCl concentration in the buffer. LACV-L was then further stabilized by the addition of the first 10 nucleotides of the 5′ vRNA (5′pAGUAGUGUGC), that were added by crystal soaking or just before cryo-EM grid freezing. LACV-L FL was crystallized and its structure solved at 4.0 Å resolution by molecular replacement using LACV-L₁₋₁₇₅₀ as a template, revealing two molecules in the asymmetric unit. There was clear extra density showing repositioning of the endonuclease and for the previously missing C-terminal region, but the resolution was insufficient for building an accurate model (Supplementary Fig. 1b, Supplementary Table 1). LACV-L FL was subsequently characterized by cryo-EM, resulting in a 3.0 Å resolution structure. A 2.28 million particle dataset was collected on a Titan Krios equipped with a K2 direct electron detector (Supplementary Fig. 2a). 2D and 3D classifications revealed that the C-terminal region of the polymerase is extremely flexible and only 0.37 million particles displaying a defined density for the C-terminal region were kept for further structural analysis (Supplementary Fig. 2b). The resulting “stable dataset” was further 3D classified resulting in the separation of two defined states: (i) the expected pre-initiation state and (ii) an elongation-mimicking state in which the complementary 3′ and 5′ vRNA formed a double-stranded RNA that could be accommodated within the active site cavity (Fig. 1b and c, Supplementary Fig. 2c). Even though only the “stable dataset” was used for 3D classification, the C-terminal region (residues 1752–2263) remained poorly defined due to flexibility. Advanced image analysis (see “Methods” section) was necessary in order to determine the structure of all C-terminal domains between 3.0 and 3.5 Å resolution (Supplementary Figs. 2c, 3 and Supplementary Table 2). The complete model of LACV-L FL was manually built and refined (Fig. 1).

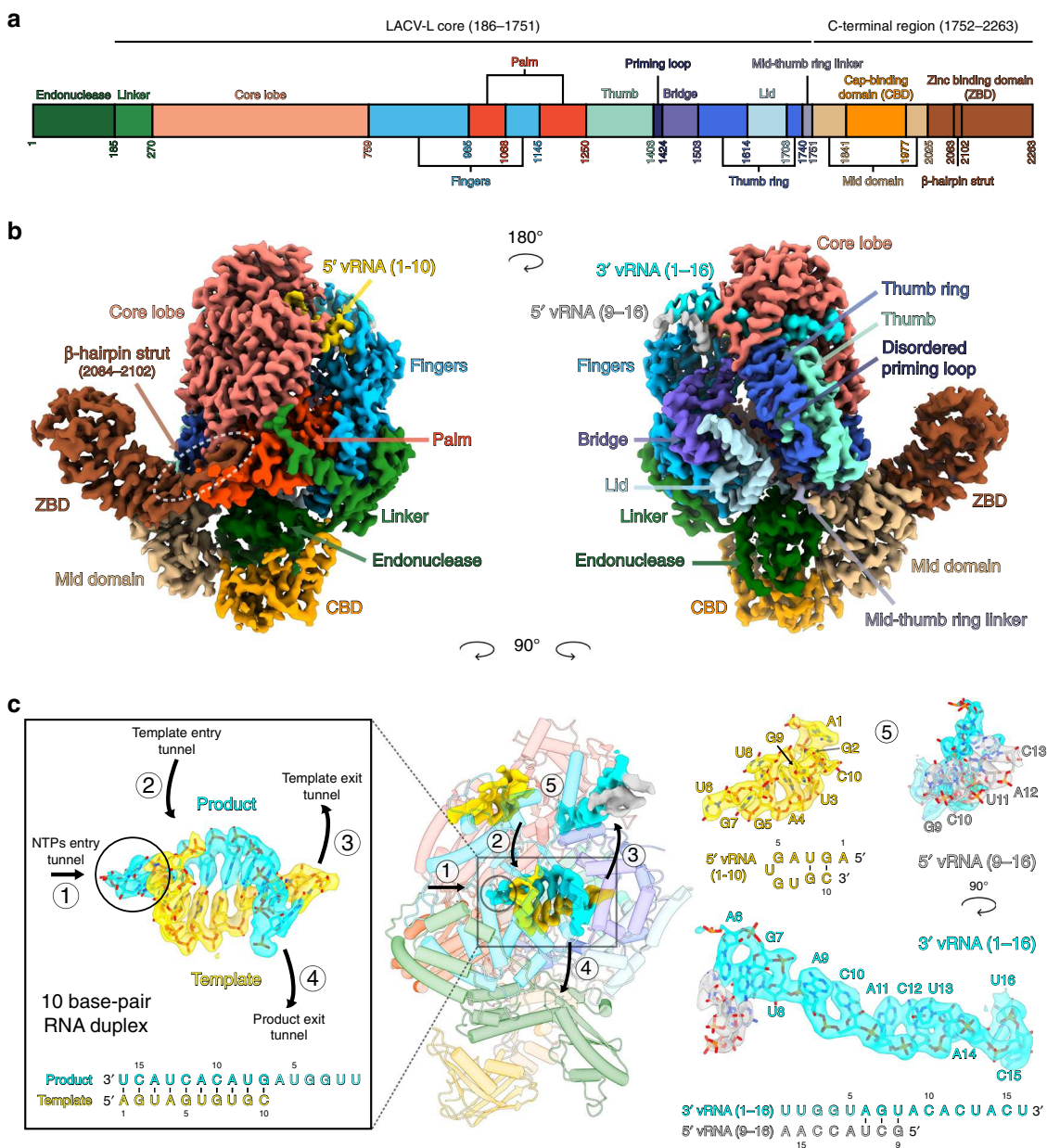


Fig. 1 Cryo-EM structures of LACV-L FL. **a** Schematic representation of LACV-L FL domain structure. **b** Two views of LACV-L FL cryo-EM map at pre-initiation stage. A composite cryo-EM map was assembled from individual maps of the (i) LACV-L core, endonuclease and mid domain, (ii) mid domain and CBD, and (iii) mid domain and ZBD. Domains and RNAs are indicated with arrows and colored as in **a**. **c** Cartoon representation of semi-transparent LACV-L FL cryo-EM structure at elongation-mimicking stage rotated of 90° compared to **b**. Close-up views of the Coulomb-potentials and models of all RNAs visible in the elongation-mimicking map. The position of the four tunnels are shown and numbered from 1 to 4. The 5' end vRNA (1-10) in its “5' end stem-loop pocket” is displayed in yellow, the 3' vRNA (1-16) in its “3' end pre-initiation pocket” is shown in cyan, the 5' vRNA (9-16) that hybridizes with the 3' vRNA (1-16) is colored in light gray. The RNA that mimics the template and product are shown in yellow and cyan, respectively. The sequence and secondary structures of nucleic acid moieties in each complex are displayed.

Overall structure of LACV-L FL. The X-ray and cryo-EM structures reveal the same overall arrangement of LACV-L FL. The structure of the polymerase core (residues 186–1751) is conserved compared to the LACV-L₁₋₁₇₅₀ construct (RMSD of 0.474 Å on 1187 Cα) but the endonuclease domain undergoes a large rotational movement of 180° (Supplementary Fig. 4). The previously unobserved C-terminal region (1752–2263) protrudes away from the core and forms an elongated arc-shaped structure that includes the mid domain (residues 1752–1841 and 1978–2025), the CBD (residues 1842–1977), and the ZBD (residues 2026–2263) (Fig. 1a and b). The C-terminal region is supported and stabilized by a β-hairpin strut (residues 2084–2102)

that emerges from the ZBD domain and bridges the entire C-terminal region to the core (Fig. 1b).

Endonuclease interactions with the other polymerase domains.

The endonuclease is held in place by hydrophobic interactions with a large number of residues from different domains (Fig. 2a). The N-terminus of the endonuclease and its un-cleaved TEV cleavage site are buried between the thumb ring (residues 1716, 1717, 1720) and the residue 714 of the core lobe β-hairpin (Fig. 2b and a). The α-helices 6 and 7 of the endonuclease face the palm domain (α-helix 38) and the finger domain (α-helix 30) (Fig. 2a

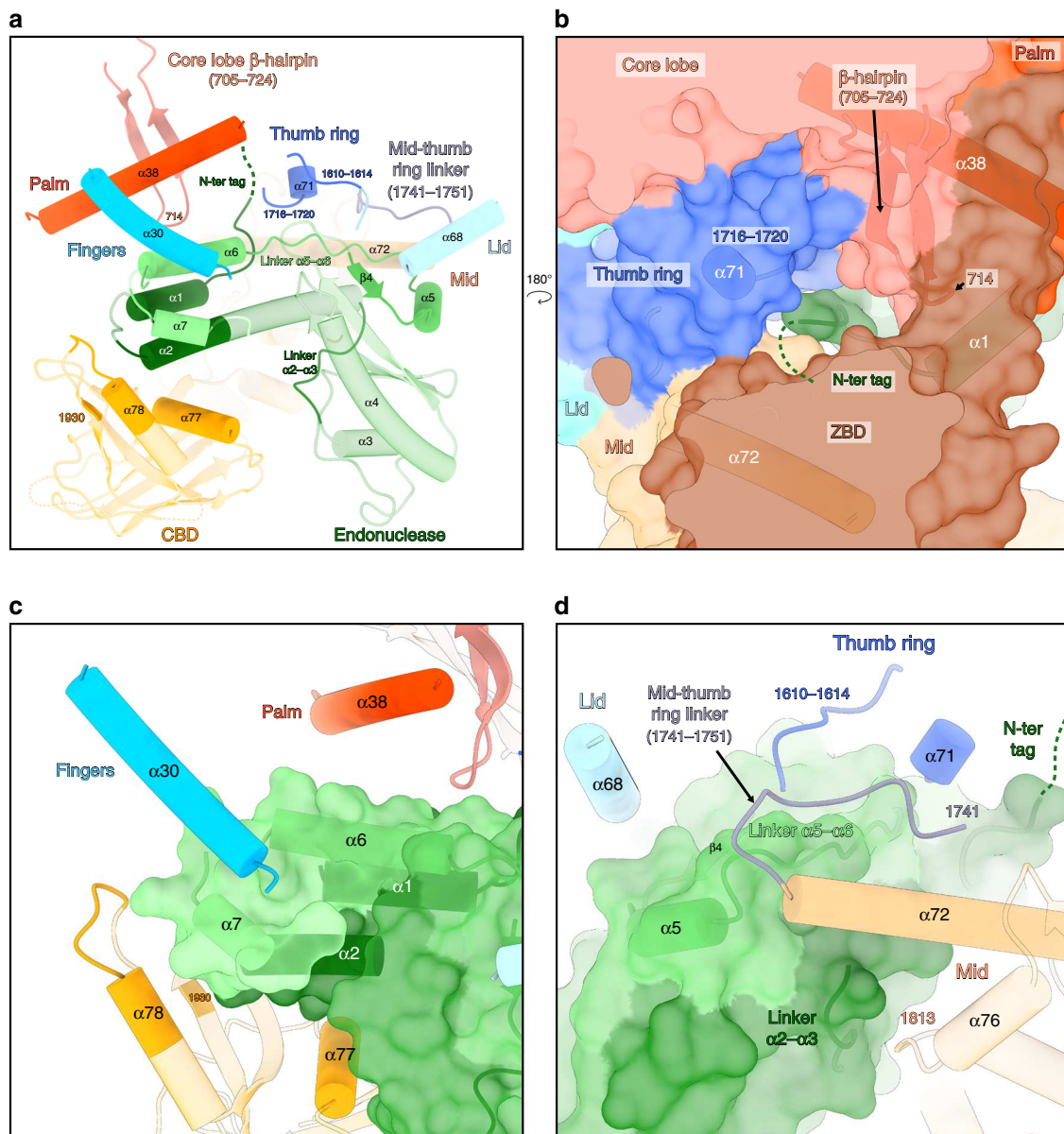


Fig. 2 Interactions of the endonuclease with the other polymerase domains. **a** Overview of the endonuclease interactions with the other polymerase domains. LACV-L orientation is the same as in Fig. 1c. Secondary structures and residues interacting with different parts of the endonuclease are shown in non-transparent and labeled. The endonuclease is colored from dark green on the N-terminus to light green on the C-terminus. The core lobe, fingers, thumb ring, lid, mid-thumb ring linker, mid, CBD are colored as in Fig. 1a. **b** Close-up view of the endonuclease N-terminus. Transparent surface of the polymerase is shown. The secondary structures of the core lobe and thumb ring domains that interact with the N-terminal tag are displayed, labeled and colored as in **a**. The N-terminal tag is represented as a dark green dotted line. **c** Close-up view of the interactions between the endonuclease and the following LACV-L domains: fingers, palm and CBD. The endonuclease is displayed as a transparent surface and colored as in **a**. **d** Close-up view of the interactions between the endonuclease and the following LACV-L domains: thumb ring, lid, mid-thumb ring linker and mid. The endonuclease is displayed as in **c**. The secondary structures are labeled and the N-terminal tag is represented as a dark green dotted line.

and c). Hydrophobic interactions are depicted between residues 46–53 (corresponding to the linker between α -helices 2 and 3) and the mid domain (residue 1813 and residues of the α -helix 76) (Fig. 2d). Residues 137–159 of the endonuclease (corresponding to the β -strand 4, the α -helix 5 and the linker between the α -helices 5 and 6) interact with the residues 1610–1614 of the thumb ring domain, the α -helix 68 of the lid domain, the α -helix 71 of the thumb ring domain, the mid-thumb ring linker (residues 1741–1751), and the α -helix 72 of the mid domain (Fig. 2d). Finally, the endonuclease interacts with the CBD (Fig. 2a), through interactions that change depending on the CBD position (described in the paragraph below).

Structure and mobility of the cap-binding domain of LACV-L. The CBD is composed of a five-stranded anti-parallel β -sheet (β 34, β 35, β 36, β 37, β 41) packed against the α -helix 77 that is flanked by a three-stranded antiparallel β -sheet (β 38, β 39, β 40), the α -helix 78 and long loops (Fig. 3a). There is a disordered loop between the first two strands of the CBD five-stranded β -sheet (Fig. 3 a, b) that contains a number of residues highly conserved in all *Peribunyaviridae* L proteins, although there is no density for them. The m^7 GTP-binding sites of RVFV and influenza virus CBD^{8,14} which share the same overall fold are located in an equivalent loop (Supplementary Fig. 5). The m^7 GTP cap-binding site of LACV-L can therefore be predicted to be composed of

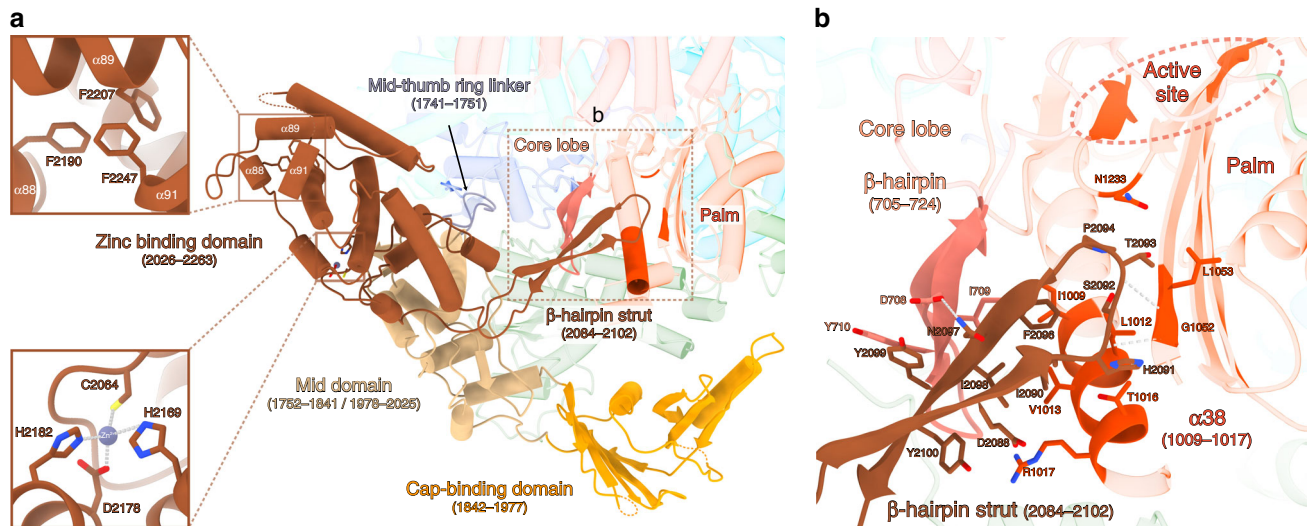


Fig. 4 C-terminal region of LACV-L FL. **a** The structure of LACV-L C-terminal region is shown as non-transparent, the rest of the polymerase is shown as transparent, except the residues interacting with the ZBD β -hairpin strut. The domains are colored as in Fig. 1. Position of the last α -helix 91 that interacts with α -helices 88/89 is indicated and a close-up view of hydrophobic amino acids interaction is shown. Close-up view on the coordination of the presumed zinc ion from the ZBD. Conserved residues implicated in ion coordination are shown and labeled. The ZBD β -hairpin strut that protrudes towards the core is surrounded by a dotted rectangle corresponding to panel **b** view. **b** Close-up view of the ZBD β -hairpin strut. Residues from the core lobe β -hairpin (705–724) and the palm domain that interact with the ZBD β -hairpin are indicated. The active site is surrounded by a dotted line.

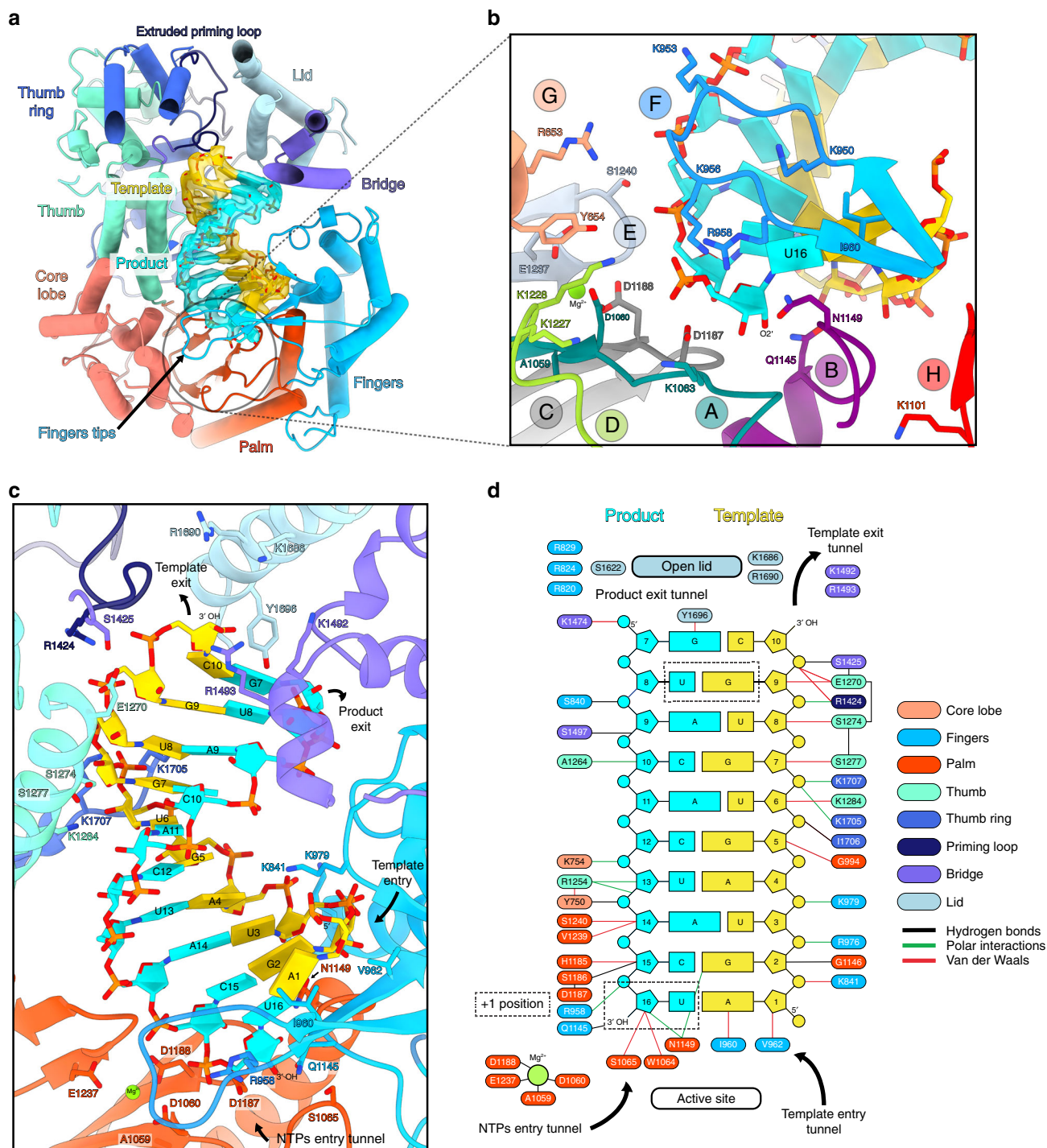
residues 1009–1017, L1053, and N1233 of the palm domain, proximal to the polymerase active site (Fig. 4b). The extreme C-terminal α -helix 91 of the ZBD (Fig. 4a and Supplementary Fig. 1c) is connected to the rest of the domain via a long flexible loop permitting large movements. In the crystal structure, α -helix 91 protrudes away to bind to a hydrophobic pocket present in the ZBD of the second polymerase of the asymmetric unit, forming a domain-swap dimer (Supplementary Fig. 1c). In the cryo-EM map, the polymerase is monomeric and some density present at low threshold suggests that the α -helix 91 may fold back into the same hydrophobic pocket of the ZBD, although the binding might be rather labile (Fig. 4a).

Elongation-mimicking state. Based on the RNA promoter sequences with which the polymerase was incubated, we expected to obtain only the pre-initiation state. However, extensive 3D classification resulted in identification of an alternative RNA-bound subset of particles in which a 10-base-pair duplex is unexpectedly visible in the active site cavity (Supplementary Fig. 2c). This structure, mimicking an elongation state with a bound product–template duplex, is determined at 3.0 Å resolution, enabling to distinguish unambiguously purine and pyrimidine bases (Fig. 5a). It can thus be deduced that the RNA duplex corresponds to the hybridization between the nearly complementary 5' and 3' promoter ends (5'p-(1)AGUAGU-GUGC(10) and 3'OH-(16)UCAUCACAUG(7)), corresponding to nucleotides 1–10 for the 5' and 7–16 for the 3' (Fig. 5a, c, d). Visualization of this state shows that a small fraction of the “stable dataset subset” (59,152 particles out of 370,497, Supplementary Fig. 2c) was able, in the in vitro conditions used and with 3' and 5' promoter ends in excess, to internalize the promoter duplex in the active site cavity. This is in addition to the 3' and 5' RNA promoter ends also being bound in their respective “3' end pre-initiation pocket” and “5' stem-loop pocket”, in positions identical to the ones observed at pre-initiation, showing the RNA binding compatibility between all these separate RNA binding sites (Fig. 1c). Although not a true elongation state, the structure obtained fortuitously mimics this state and gives insight into the mechanisms of (i) RNA binding in the active site

cavity and (ii) template–product separation after formation of a 10-base-pair double-stranded RNA in the active site cavity.

The backbone and some bases of the template–product duplex contact many residues lining the polymerase active site chamber via both van der Waals and polar interactions (Fig. 5c, d). The active-site proximal part of the template-mimicking RNA (nucleotides 1, 2 and 3) interacts with the finger domain, the central part (nucleotides 4, 5) binds to the palm, while the distal part (nucleotides 6–10) interacts with the thumb and the thumb ring domains (Fig. 5a, c, d). The proximal part of the product-mimicking RNA (nucleotides 14–16) is surrounded by the palm domain, the central part of the product (nucleotides 10–13) interacts with the core and the core lobe, while the distal part of the product-mimicking RNA (nucleotides 7–9) mainly binds to the bridge and the finger domains (Fig. 5a, c, d). The LACV-L catalytic core shares with other viral RNA-dependant RNA-polymerases the six conserved structural motifs (A–F)¹⁹ (Fig. 5b). In addition, motifs G and H that are specific to sNSV polymerases are also visible¹³ (Fig. 5b).

The polymerase conformation mimics a post-incorporation, pre-translocation elongation step in which an incoming nucleotide would just have been incorporated into the product. During the nucleotide addition cycle, viral polymerase active sites undergo small structural changes that enable NTP-binding, NTP-incorporation, and subsequent RNA translocation²⁰. Whereas a particular configuration of (i) the nucleotide to be incorporated, (ii) the product RNA to be elongated, (iii) two magnesium ions, and (iv) the three aspartic acids of motifs A and C is necessary for catalysis of the phosphoryl transfer reaction, the organization of the active site changes subsequently²⁰. Such post-incorporation structural changes are visible in LACV-L elongation mimicking stage. Perhaps related to the fact that there is no pyrophosphate (since there was no reaction), the two magnesium ions that control nucleotide addition are not present in the catalytic configuration. Instead, a presumed Mg^{2+} ion, coordinated by residues D1188 (motif C), E1237 (motif E), and the carbonyl oxygen of A1059 (motif A) is present, typical of the inactive open state of the polymerase active site (Fig. 5b, Supplementary Fig. 6a). Some other motifs are also in post-incorporation conformation,



such as the fingertips residues R958 and I960 that, respectively, stack the bases of the product and template nucleotides in the +1 position, thereby stabilizing them (Fig. 5b). The motif B loop, which is implicated in the selection of the correct nucleotide to be incorporated, also adopts a conformation compatible with a post-incorporation state. Its residue Q1145 contacts the +1 position nucleotide base of the last incorporated product nucleotide, while residue N1149 interacts with the 2' hydroxyl group of the template nucleotide (Fig. 5b). In summary, the elongation-mimicking structure represents a post-incorporation, pre-translocation elongation step containing a 10-base-pair template-product RNA in the active site chamber.

Conformational changes between pre-initiation and elongation. Comparison between the pre-initiation and elongation-mimicking states reveals key movements of the L protein in action. The priming loop (residues 1404–1424) is an essential element that usually stabilizes the first ‘priming’ nucleotide of the product during replication initiation. In the pre-initiation structure, this loop protrudes towards the active site but is disordered probably due to the absence of RNA and nucleotides (Fig. 6a). As part of the initiation to elongation transition, it extrudes from the active site via the template exit tunnel, thereby freeing space for the 10-base-pair RNA to fit in the active site chamber (Fig. 6b). The fully ordered and extruded priming loop is located on the

Fig. 5 Cryo-EM structure of the LACV-L FL at elongation-mimicking stage. **a** Cut-away view of the LACV-L FL at elongation-mimicking stage. Its orientation corresponds to a top view visualization of Fig. 1b left. The main domains are depicted and colored as in Fig. 1. The Coulomb potential map of the 10-base-pair product-template RNA is shown in cyan for the product (3'OH-UCAUCAUG, nucleotides 7–16) and gold for the template (5'p-AGUAGUGUGC, nucleotides 1–10). The extruded priming loop is shown in dark blue. The active site position is indicated as a dotted circle. **b** View of the LACV-L FL active site showing the conserved RNA-dependent RNA polymerase functional motifs A–H (G and H are only conserved in sNSV polymerases). They are respectively colored turquoise, purple, gray, light green, light blue, blue, beige, and red for A–H. Template-mimicking and product-mimicking RNA are colored as in **a**. Presumed magnesium ion, is shown as a green sphere. **c** Interactions of the 10-base-pair product-template RNA in LACV-L active site cavity. Principal residues from the active site in palm domain (A1059, D1060, S1065, D1187, D1188, E1237), fingertips (R958), fingers (K841, I960, V962, K979, Q1145), priming loop (R1424), bridge (S1425, K1492, R1493), thumb (E1270, S1274, S1277, K1284), and lid (K1686, R1690, Y1696) are displayed. NTPs entry, template entry, template entry/exit tunnel directions are shown. Ion position is shown as in **b**. Nucleotides are labeled according to RNA promoter sequence. **d** Schematic representation of RNA-protein contacts in the active site cavity. Residues are colored according the domain to which they belong. Template and product RNA are numbered according to their position in the 5' end promoter (5'p-AGUAGUGUGC, nucleotides 1–10) and the 3' end promoter (3'OH-UCAUCAUG, nucleotides 7–16). The U-G mismatch that is due to the non-perfect complementarity between the 5' and 3' promoters is surrounded by a dotted rectangle. Interaction type are color coded as indicated. Ion is shown as a green circle. Active site and lid domain positions are indicated. Nucleotide U16 corresponds to the nucleotide in position +1 of the product and is identified as such.

surface of the thumb ring and lid domains (Fig. 6b) with which it interacts mainly through hydrophobic contacts involving residues V1572, Y1576, A1751, M1753, N1658, L1661 (Fig. 6d). Interestingly, the priming loop movement is coupled with the reorganization of mid domain residues 1752–1761 from an α -helix to an extended loop. This results in an extension of the mid-thumb ring linker from residues 1741–1751 at pre-initiation to 1741–1761 at elongation. As a result of the helix unwinding, the mid-thumb ring linker loop extremity (residues 1750–1753, at pre-initiation), is displaced by 8 Å in the transition to elongation (Fig. 6d) and interacts with the priming loop residues 1414–1418 mainly through hydrophobic contacts (Fig. 6d).

The transition from pre-initiation to elongation is also coupled to coordinated domain movements. The lid domain rotates by 12° compared to the thumb ring, resulting in the opening of the template and the product exit tunnels (Fig. 6c). Separation of the template-product RNA duplex is made possible by the α -helix 70 of the lid domain that faces the distal part of the double-stranded RNA. Its residue Y1696 interacts with the product 5' end nucleotide, thereby forcing strand-separation of the RNA duplex (Fig. 5c). Domain movements occurring between pre-initiation and elongation are nicely captured by a 3D variability analysis of the dataset (Supplementary Movie 2). It reveals a coordinated rotation of the endonuclease and the C-terminal region compared to the core, using the mid domain as a hinge, and resulting in 4.5 and 8 Å displacement of the endonuclease and the ZBD respectively (Supplementary Movie 2).

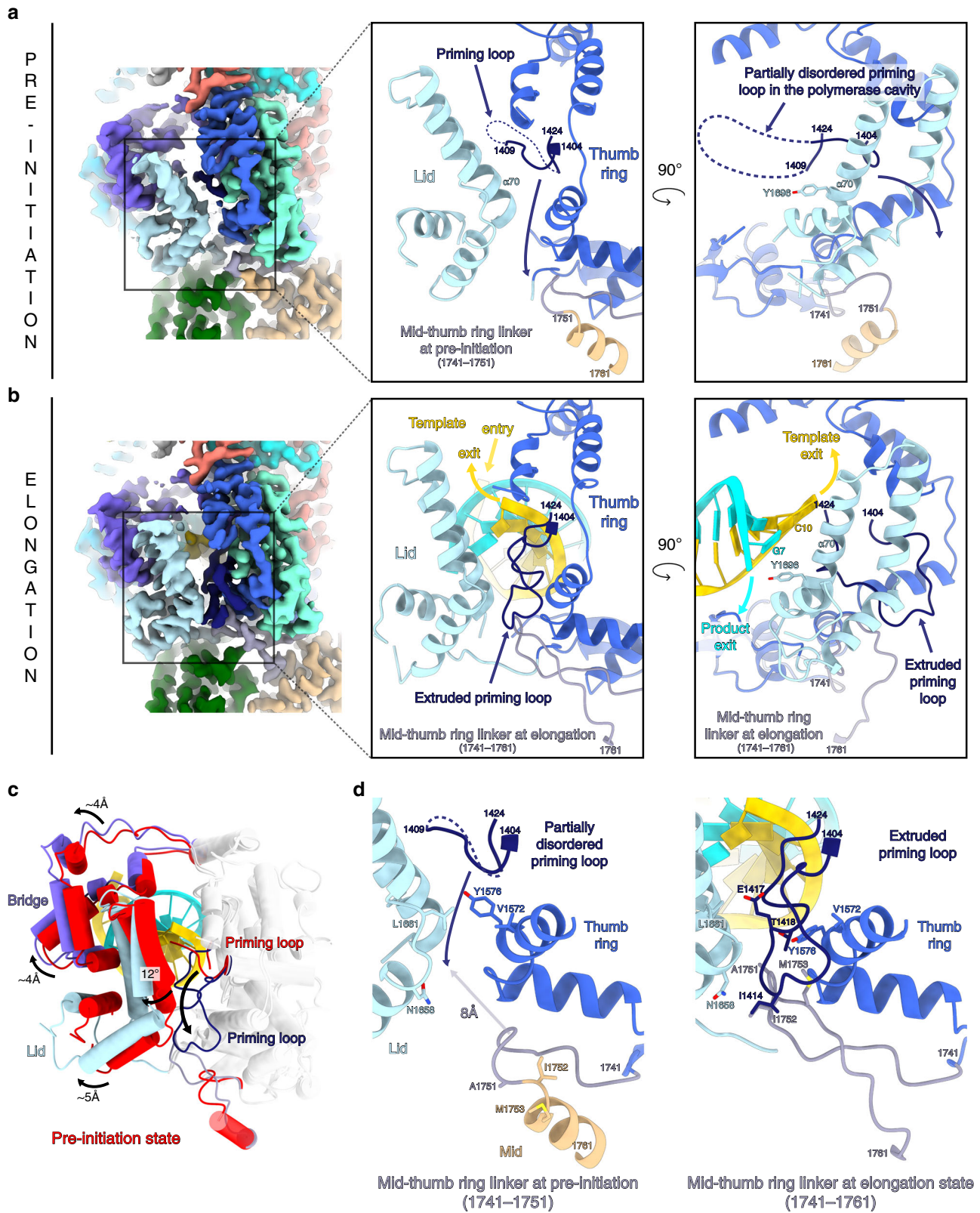
Discussion

The structure presented here reveals the organization of the entire LACV-L protein. The newly described C-terminal domains can be compared with equivalent parts of *Phenuiviridae*, *Arenaviridae*, and *Orthomyxoviridae* polymerases. LACV-L CBD shares a conserved fold with equivalent structures from RVFV (*Phenuiviridae*)⁸, CASV (*Arenaviridae*)⁹, and influenza virus (*Orthomyxoviridae*)¹⁴, consisting of an antiparallel β -sheet stacked against an α -helix (Supplementary Fig. 5). In addition, LACV-L CBD contains a family-specific insertion consisting of a three-stranded β -sheet (β 38, β 39, β 40), α -helix 78 and charged loops (1932–1936 and 1956–1963). This insertion is likely related to the role of the CBD in transcription initiation as it is involved in interactions with the endonuclease (Fig. 3d). LACV C-terminal region also contains a mid domain and a ZBD. The mid domain fold is conserved between *Orthomyxoviridae* and *Arenaviridae* polymerases^{9,11} (Supplementary Fig. 7b), whereas the distal C-terminal region differs between the three families. *Orthomyxoviridae* and *Arenaviridae*, respectively, have a PB2 627-domain and D1-III domain that are structurally related^{9,11}, whereas the

ZBD present in LACV has a different fold. This suggests that *Orthomyxoviridae* and *Arenaviridae* polymerases are more closely related (Supplementary Fig. 7c). The presence of the ZBD or PB2 627-domain in the C-terminal region in a protruding position compared to the polymerase core suggests that, as for the PB2 627-domain²¹, LACV-L ZBD may play a role as a platform to recruit host cytoplasmic proteins. For instance it could interact with host proteins that bind capped RNA, host factors that mediate the transcription-translation coupling observed in *Bunyavirales*^{5,22} or be involved in replication-related activities. It should however be noticed that, contrary to influenza PB2 627-domain, it is not possible to identify a non-conservative mutation of a residue in LACV-L ZBD that would differentiate one LACV strain to another and might reflect host specificity.

The endonuclease and the CBD have essential roles in transcription initiation. For influenza polymerase, structures depicting their coordinated movements explain how cap-dependent transcription is initiated^{12,23}. The present LACV-L FL structure reveals two relative positions of the endonuclease and the CBD with respect to the core, that we call “LACV-L conformation 1” and “LACV-L conformation 2” (Supplementary Fig. 8a, Fig. 3c). In both conformations, the endonuclease is stabilized in position by making hydrophobic interactions with the thumb ring, core lobe, mid, palm, finger and cap-binding domains (Fig. 2). Its orientation however differs by around 180° compared to LACV-L_{L1-1750} (Supplementary Fig. 4) where the endonuclease protrudes away from the polymerase core, making only few interactions with it. Concerning the CBD, its position in the LACV-L FL dataset is variable, with extreme conformations being 35° apart, respectively, corresponding to LACV-L conformation 1 and LACV-L conformation 2 (Fig. 3c, Supplementary Fig. 8a, Supplementary Movie 1).

These conformations are likely to correspond to functional states. LACV-L conformation 1 may correspond to a pre-initiation transcription state compatible with the binding of cellular capped RNA, as its CBD is exposed towards the exterior (Supplementary Fig. 8a). In LACV-L conformation 2, the cap-binding site is closer to the endonuclease and we speculate that in this conformation the RNA can be cleaved by the endonuclease (Supplementary Fig. 8a). However, the observed location of the endonuclease prevents rotation of the CBD into a position where it can direct the capped primer into the active site for transcription initiation (Supplementary Fig. 8a), so a relative repositioning of the two domains is expected. In addition, the endonuclease active site is located close to the product exit (Supplementary Fig. 8a). In order to prevent degradation of the product when it exits the polymerase, one can speculate that a large movement of the endonuclease may be necessary at a late elongation stage.



The position of both the endonuclease and the CBD in LACV-L conformation 1 and 2 significantly differ from the observed disposition of these domains in all the functional states described for influenza polymerase (Supplementary Fig. 8b). They also diverge from the recently described structures of full-length L

protein from Machupo virus (*Arenaviridae* family)¹⁷ and Severe Fever with Thrombocytopenia Syndrome virus L (*Phenuiviridae* family)^{24,25} published while this manuscript was under review (Supplementary Fig. 8c, d). However, without being sure that these latter structures represent functionally active states, it is too

Fig. 6 Conformational changes between pre-initiation and elongation-mimicking states. **a** LACV-L FL in pre-initiation state. Left panel: cryo-EM map colored as in Fig. 1. Middle panel: close-up view. Right panel: close-up view rotated 90° compared to the middle panel. For more clarity, only the lid, thumb ring, and mid domains are shown in the middle and right panels and are colored in light blue, blue, and beige. The mid-thumb ring linker (1741–1751) is shown in purple gray. The priming loop is shown in dark blue. All these elements are labeled. The lid domain is closed and the residue Y1696 from the α -helix 70 points towards the cavity. The priming loop is disordered between residues 1410 and 1423 (dotted line). Priming loop movement occurring during elongation is depicted with a dark blue arrow. **b** LACV-L FL in elongation-mimicking state. Equivalent panels as in **a** are shown with domains and loops colored as in **a**. The mid-thumb ring linker contains residues 1741–1761 at elongation. The template tunnel entry and exit, the product exit are shown. The lid domain is open and Y1696 from the α -helix 70 interacts with the product-mimicking RNA. The priming loop is ordered and interacts with the lid, the thumb ring, and the mid-thumb ring linker. **c** Superimposition of the thumb and thumb ring (shown in transparent gray) in pre-initiation and elongation-mimicking state of LACV-L FL. The bridge, the lid, and the priming loop in pre-initiation are shown in red. The bridge, the lid and the priming loop in elongation-mimicking state are shown in purple, light blue and dark blue. Their rotation between the two states is labeled. Movement of the priming loop between the two states is shown with an arrow. **d** Priming loop and mid-thumb ring linker at pre-initiation (left) and elongation (right). Numbering of both elements are indicated. Residues from the lid and thumb ring that interact with the priming loop at pre-initiation and/or elongation are shown. The 8 Å displacement between the mid-thumb ring linker extremity at pre-initiation and elongation is indicated with an arrow.

early to draw any conclusions on differences in mechanism between these various viral polymerases. The insight provided here on LACV-L constitutes a basis to address in more detail the exact mechanisms underlying *Peribunyaviridae* transcription initiation in the future, notably by determining structures in complex with capped RNA.

The active sites of LACV and influenza polymerases in presence of RNA are remarkably similar (Supplementary Fig. 6). Previous studies of influenza polymerase have depicted the post-incorporation pre-translocation (Supplementary Fig. 6b)²³ and post-incorporation post-translocation (Supplementary Fig. 6c)¹² states. LACV-L FL bound to duplex RNA appears to mimic an intermediate between these two states as it displays a product nucleotide at the +1 position (as if it had been newly incorporated) with a magnesium ion and the catalytic aspartic acids in an inactive open state conformation (Supplementary Fig. 6a). Such an intermediate resembles the equivalent state in poliovirus polymerase (PDB: 3OL9)²⁶ although the observed movements of the active site between pre/post translocations are restricted to side chains movements in LACV/influenza polymerase structures, whereas they involve larger movements of motifs in poliovirus polymerase. It should however be noticed that the observed state in LACV-L was obtained by internalization of a double-stranded RNA and might not exactly reflect a true elongation intermediate notably because of the absence of pyrophosphate.

More generally, the cavities that bind the template-product RNA duplex are very similar in LACV-L and influenza polymerases. The mechanism of template-product separation by the lid domain is remarkably conserved with the equivalent helix being implicated¹². The tyrosine that prevents double-strand continuation strand differs however slightly in position: Y1696 of LACV-L interacts with the RNA product whereas Y207 of influenza PB2 stacks with the last nucleotide of the template (Supplementary Fig. 9). Movements occurring between pre-initiation and elongation also differ between the two viral polymerases. For instance, the priming loop takes a completely different position and organization when it extrudes from the active site in the two cases (Supplementary Fig. 9). Influenza polymerase priming loop, which is 35-amino acid long, has 17 residues extruded into the solvent in a disordered loop, and projects towards the PB1–PB2 interface helical bundle¹² (Supplementary Fig. 9b). LACV-L priming loop, which is 20-amino acid long, orders itself at the surface of thumb ring and lid domains (Supplementary Fig. 9a). Its extrusion is coupled with a *Peribunyaviridae*-specific unwinding of the α -helix 72 of the mid-domain into a loop, enabling an interaction between the priming loop and the extended mid-thumb ring linker (Fig. 6d). Conformational changes of the whole C-terminal region upon the transition from pre-initiation to elongation visualized in the 3D variability

analysis (Supplementary Movie 2) have also not been observed in influenza polymerase.

Altogether, the complete structure of LACV-L FL captured in pre-initiation and elongation-mimicking states and the rotational mobility of the CBD provide mechanistic insight into bunyaviral transcription. This, reinforced by the atomic details of the polymerase active site, establishes a firm basis for future structure-based drug design that could target essential activities or critical conformational changes of *Peribunyaviridae*L proteins.

Methods

Cloning, expression, and purification. Sequence-optimized synthetic DNA encoding a N-terminal his-tag, a TEV protease recognition site, and the LACV-L FL (strain LACV/mosquito/1978, GenBank: EF485038.1, UniProt: A5HC98) was synthesized (Geneart) and cloned into a pFastBac1 vector between NdeI and NotI restriction sites (Supplementary Table 3). The LACV-LFL expressing baculovirus was generated via the standard Bac-to-Bac method (Invitrogen). For large-scale expression, *Trichoplusia ni* High 5 cells at 0.5×10^6 cells/mL concentration were infected by adding 0.1% of virus. Expression was stopped 72 h after the day of proliferation arrest. The cells were disrupted by sonication for 3 min (10 s ON, 20 s OFF, 50% amplitude) on ice in lysis buffer (50 mM Tris–HCl pH 8, 500 mM NaCl, 20 mM Imidazole, 0.5 mM TCEP, 10% glycerol) with EDTA-free protease inhibitor complex. After lysate centrifugation at $48,384 \times g$ during 45 min at 4 °C, protein from the soluble fraction was precipitated using $(\text{NH}_4)_2\text{SO}_4$ at 0.5 mg/ml and centrifuged at $104,630 \times g$ for 45 min, 4 °C. Supernatant was discarded, proteins were resuspended back in the same volume of lysis buffer and centrifuged at $104,630 \times g$ during 45 min at 4 °C. LACV-L FL was purified from the supernatant by nickel ion affinity chromatography after a wash step using 50 mM Tris–HCl pH 8, 1 M NaCl, 20 mM Imidazole, 0.5 mM TCEP, 10% glycerol and eluted using initial lysis buffer supplemented by 300 mM Imidazole. LACV-L FL fractions were pooled and dialyzed 1 h at 4 °C in heparin-loading buffer (50 mM Tris–HCl pH 8, 250 mM NaCl, 0.5 mM TCEP, 10% glycerol). Proteins were loaded on heparin column and eluted using 50 mM Tris–HCl pH 8, 1 M NaCl, 0.5 mM TCEP, 5% glycerol. LACV-L FL was then mixed in a 1:3 molar ratio with both 3' (1–16, 3'-UCAUCAUGAUGGUU-5') and 5' (9–16, 5'-GCUACCAA-3') vRNA oligonucleotide ends which had been pre-annealed by heating at 95 °C for 2–5 min followed by cooling down on bench at room temperature. During overnight dialysis at 4 °C in a gel filtration buffer (20 mM Tris–HCl pH 8, 150 mM NaCl, 2 mM TCEP) LACV-L FL formed a complex with vRNA, which was ultimately resolved on the S200 size exclusion chromatography column.

Crystallization and X-ray crystallography. For crystallization, LACV-L FL in complex with pre-annealed 3' (1–16) and 5' (9–16) vRNA was concentrated to 5 mg/ml. The 5' (1–10) vRNA end (5'-AGUAGUGGC-3') was later soaked into crystals in 1:2 molar ratio. Initial hits were dense and round precipitates that appeared in 100 mM Tris pH 8.0, 100 mM NaCl, and 8% PEG 4000. Upon manual reproduction in hanging drops, they grew as thin hexagonal plates, but were soft and fragile and diffracted only to ~ 8 Å. To improve the resolution, crystals were soaked in a stepwise manner with increasing concentration of the glycerol cryo-protectant, reaching 30%. Diffraction data were collected at the European Synchrotron Radiation Facility (ESRF), using a helical collection strategy and maximum transmission of the ID29 beamline. Crystals are of space-group C2, diffracting at best to a maximum resolution of 4.0 Å. Data were integrated with STARANISO²⁷ to account for the anisotropy (Supplementary Table 1). The structure was solved with PHASER²⁸ using LACV-L₁₋₁₇₅₀ (PDB code: 5AMQ)¹³ as a model after removal of the endonuclease. There are two L protein complexes per asymmetric unit (Supplementary Fig. 1b). The initial map after molecular replacement revealed that the core of the L protein and bound

RNA were little changed but there was clear density for the endonuclease in a new position. In addition, there was extra density for the previously missing C-terminal domain. This density was improved by multi-crystal and non-crystallographic two-fold averaging using PHENIX²⁹. Based on secondary structures predicted from an extensive multiple sequence alignment of *Peribunyaviridae* L proteins³⁰, it was possible to build an approximate model of much of the C-terminal domain (which is largely helical), except the CBD for which there is no density, including identification of the zinc-binding site co-ordinated by highly conserved cysteine and histidine residues. The two C-terminal domains from the two complexes in the asymmetric unit interact around a non-crystallographic two-fold axis in such a way that the extreme terminal helix of one packs against the C-terminal domain of the other, forming a domain swapped dimer. When the accurate structure of the C-terminal domain was obtained by cryo-EM, the X-ray model could be improved (Supplementary Fig. 1c, Supplementary Table 1). The two polymerases of the crystallographic asymmetric unit are very similar with an RMSD of 0.79 Å over 2009 Ca, although there is a slight difference in C-terminal domain orientation. They are also very similar to the cryo-EM structure at pre-initiation with RMSDs of, respectively, 3.022 and 2.951 Å over 2009 Ca. A significant difference concerns the C-terminal linker and last α -helix that form a domain swapped dimer interaction in the X-ray structure but folds back on the ZBD in the monomeric cryo-EM structure.

Electron microscopy. For cryo-EM experiments, LACV-L FL in complex with pre-annealed 3' (1–16) and 5' (9–16) vRNA at 0.2 mg/ml was mixed with 5' (1–10) vRNA hook in a 1:2 molar ratio. UltraAuFoil grids 300 mesh, R 1.2/1.3 were negatively glow-discharged at 30 mA for 1 min. 3.5 μ l of the sample was applied on the grids and excess solution was blotted away with a Vitrobot Mark IV (FEI) (blot time: 2 s, blot force: 1, 100% humidity, 20 °C), before plunge-freezing in liquid ethane. Grid screening and cryo-EM initial datasets were collected on a 200 kV Thermofischer Glacios microscope equipped with a Falcon II direct electron detector.

A high-quality cryo-EM grid pre-screened on a 200 kV Thermofischer Glacios microscope was used to collect data on a Thermofischer Titan Krios G3 operated at 300 kV equipped with a Gatan Bioquantum LS/967 energy filter (slit width of 20 eV) coupled to a Gatan K2 direct electron detector camera³¹. Automated data collection was performed with SerialEM using a beam-tilt data collection scheme³², acquiring one image per hole from nine holes before moving the stage. Micrographs were recorded in super-resolution mode at a 165,000 \times magnification giving a pixel size of 0.4135 Å with defocus ranging from -0.8 to -3.5 μ m. In total, 16,498 movies with 40 frames per movie were collected with a total exposure of 50 $e^-/\text{Å}^2$ (Supplementary Table 2).

Image processing. Movie drift correction was performed in Motioncor2 using all frames, applying gain reference and cameras defect correction³³. Images were binned twice, resulting in 0.826 Å/pixel size. Further initial image processing steps were performed in cryoSPARC v2.14.2³⁴. CTF parameters were determined using “patch CTF estimation” on non-dose weighted micrographs. Realigned micrographs were then manually inspected using “Curate exposure” and low-quality micrographs were manually discarded for further image processing resulting in a curated 16,015 micrographs dataset (Supplementary Fig. 2c). LACV-L FL particles were then picked with “blob picker” using a circular particle diameter ranging from 90 to 150 Å, manually inspected and selected using “inspect particle picks”, extracted from dose-weighted micrographs using a box size of 300 \times 300 pixels², resulting in a 4,065,475 particles dataset. Successive 2D classifications were used to eliminate bad quality particles displaying poor structural features resulting in 2,279,573 particles suitable for further image processing (Supplementary Fig. 2b). Per particle CTF was calculated. The LACV-L FL initial 3D reconstruction was generated with “ab-initio reconstruction” in cryoSPARC using a small subset of particles. Subsequent steps were all performed in Relion 3.1^{35,36}. The entire dataset was divided in four (~570,000 particles per subset) and subjected to 3D classification with coarse image-alignment sampling (7.5°) using a circular mask of 170 Å and 10 classes (labeled “1st 3D classification”, Supplementary Fig. 2c). LACV-L FL classes displaying a stable C-terminal conformation were kept and merged for further classification resulting in 566,025 selected particles (dotted squared, green maps in Supplementary Fig. 2c). Selected particles were subjected to a 2nd run of 3D classification using finer angular sampling (3.7°) with a circular mask of 170 Å and was restricted to 10 classes (labeled “2nd 3D classification”, Supplementary Fig. 2c). Particles from three 3D classes that display stable C-terminal regions without neighboring particles were selected to perform further high-resolution analysis, resulting in a 370,497 particle dataset. A previously obtained 3D class from the 2nd run of 3D classification was low-pass filtered to 15 Å, extended by 10 pixels with a soft edge of 5 pixels and used as a mask for a final 3rd run of 3D classification using finer angular sampling (1.8°) (labeled “3rd 3D classification”, Supplementary Fig. 2c). Two classes (in dark blue, 59,152 particles) displayed a 10-base pair template-product RNA duplex in the LACV-L FL active site cavity and mimic an elongation state and one class (in cyan, 57,660 particles) displays a typical pre-initiation state with both 5' 1–10 and 3' 1–16/5' 9–16 promoters bound. Both LACV-L FL pre-initiation and elongation-mimicking state subsets were submitted to 3D auto-refine using previous initial mask giving reconstructions at, respectively, 3.13 and 3.17 Å resolution using the FSC 0.143 cutoff criteria before post-processing. Masks to perform sharpening were generated using 3D refined

maps low-pass filtered at 10 Å, extended by 4 pixels with 8 pixels of soft-edge. Post-processing was done using an applied B -factor of -40 Å² and resulted in a map at 3.02 Å resolution for the elongation state and 3.06 Å resolution for the pre-initiation state using the FSC 0.143 cutoff criteria (Supplementary Figs. 2c, and 3a).

In order to deal with the high mobility and the small size of the CBD within LACV-L FL particles, the following advanced strategy was applied. 131,058 particles displaying a stable CBD (corresponding to its extreme position 1), originating from 4 out of the 10 classes from the third 3D classification (in orange in “third 3D classification-CBD view”, Supplementary Fig. 2c) were submitted to 3D auto-refine in order to get the best global accuracy alignment. A mask excluding LACV-L₄₆₋₁₇₅₁ residues, low-pass filtered to 10 Å, resampled and extended of 4 pixels with a soft edge of 6 pixels was then used for signal subtraction followed by particle re-centering on the mask center-of-mass. The resulting subtracted particles containing Endo₁₋₄₅-Mid-CBD densities were subjected to 3D masked auto-refine with local angular searches. The obtained map was sharpened with an applied B -factor of -90 Å² and resulted in a 3.54 Å resolution map according to the FSC 0.143 cutoff criteria.

A similar strategy was applied to deal with the ZBD flexibility. 287,363 particles displaying an ordered ZBD, originating from 7 out of the 10 classes from the third 3D classification (in red in “third 3D classification-ZBD view”, Supplementary Fig. 2c) were submitted to 3D auto-refine in order to get the best global accuracy alignment. A mask excluding LACV-L_{1-1751/1841-1983} residues, low-pass filtered to 10 Å, resampled and extended of 4 pixels with a soft edge of 6 pixels was used for signal subtraction followed by particle re-centering on the center-of-mass of this mask. Resulting subtracted particles containing mid-ZBD densities were classified without alignment in order to detect potential heterogeneity. The most stable subset containing 51,842 particles was subjected to 3D masked auto refine with local angular searches. The resulting map was post processed with an applied B -factor of -83 Å², which resulted in a 3.49 Å resolution map according to the FSC 0.143 cutoff criteria.

For each final map, local resolution variations were estimated in Relion 3.1 (Supplementary Fig. 3). The 3D variability analysis was performed in cryoSPARC filtering resolution to 4 Å and using three modes.

Model building in the cryo-EM maps. All the cryo-EM maps, namely pre-initiation map, elongation-mimicking map, CBD-mid domain map, and ZBD-mid domain map were superimposed using Chimera³⁷ previous to model building. The partial model determined in the 4.0 Å X-ray structure was used as a starting point to manually build into the cryo-EM maps using COOT³⁸. The map chosen for model building was the one corresponding to the best resolution in the region built. The pre-initiation and elongation-mimicking maps were used to build the LACV-L core, the endonuclease domain, the mid domain and the β -hairpin strut of the ZBD. The CBD-mid domain map was used to build the CBD. The ZBD-mid domain map was used to build the ZBD devoid of its β -hairpin strut. The sequence of the RNA duplex visible in the active site cavity at elongation state was deduced based on purines and pyrimidines, clearly visible in the density. As the 3' and 5' vRNA incubated respectively contain 16 and 10 nucleotides, 6 nucleotides of the vRNA should be present in the product exit tunnel and none in the template exit tunnel. Blurred density is visualized in the product exit tunnel due to the large flexibility of the nucleotides. Blurred RNA density is also visible in the template exit tunnel suggesting that some particles have encapsidated the RNA in different positions than the one shown in Figs. 1c and 5d. After initial manual building in COOT, the models were iteratively improved by Phenix-real space refinement³⁹ and manual building in COOT. Validation was performed using the Phenix validation tool, and model resolution was estimated at the 0.5 Fourier shell correlation (FSC) cutoff. Figures were generated using ChimeraX⁴⁰.

Multiple alignment. Six sequences of major Peribunyaviruses (L_BUNYW for Bunyamwera virus, L_SVVBG for Schmallenberg virus, L_MCAV for Macau virus, L_WBV for Wolkberg virus, L_Oya for Oya virus) were aligned using MUSCLE⁴¹ and presented with ESPript⁴².

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Coordinates and structure factor have been deposited in the Protein Data Bank and the Electron Microscopy Data Bank with the accession codes: LACV-L pre-initiation complex (X-ray crystallography) PDB 6Z6B, LACV-L pre-initiation complex PDB 6Z6G, EMDB EMD-11093, LACV-L elongation complex PDB 6Z8K, EMDB EMD-11118, LACV-L CBD and mid domain map EMDB EMD-11095, LACV-L ZBD and mid domain map EMDB EMD-11107. Other data are available from the corresponding authors upon reasonable request.

Received: 24 March 2020; Accepted: 25 June 2020;
Published online: 17 July 2020

References

- Maes, P. et al. Taxonomy of the order Bunyvirales: second update 2018. *Arch. Virol.* **164**, 927–941 (2019).
- Reguera, J., Gerlach, P. & Cusack, S. Towards a structural understanding of RNA synthesis by negative strand RNA viral polymerases. *Curr. Opin. Struct. Biol.* **36**, 75–84 (2016).
- Sun, Y., Li, J., Gao, G. F., Tien, P. & Liu, W. Bunyavirales ribonucleoproteins: the viral replication and transcription machinery. *Crit. Rev. Microbiol.* **44**, 522–540 (2018).
- Eisfeld, A. J., Neumann, G. & Kawaoka, Y. At the centre: influenza A virus ribonucleoproteins. *Nat. Rev. Microbiol.* **13**, 28–41 (2015).
- Olschewski, S., Cusack, S. & Rosenthal, M. The cap-snatching mechanism of Bunyaviruses. *Trends Microbiol.* <https://doi.org/10.1016/j.tim.2019.12.006> (2020).
- Patterson, J. L. & Kolakofsky, D. Characterization of La Crosse virus small-genome transcripts. *J. Virol.* **49**, 680–685 (1984).
- Lukarska, M. et al. Structural basis of an essential interaction between influenza polymerase and Pol II CTD. *Nature* **541**, 117–121 (2017).
- Gogrefe, N., Reindl, S., Günther, S. & Rosenthal, M. Structure of a functional cap-binding domain in Rift Valley fever virus L protein. *PLoS Pathog.* **15**, e1007829 (2019).
- Rosenthal, M. et al. Structural insights into reptarenavirus cap-snatching machinery. *PLoS Pathog.* **13**, e1006400 (2017).
- Reich, S. et al. Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* <https://doi.org/10.1038/nature14009> (2014).
- Pflug, A., Guilligay, D., Reich, S. & Cusack, S. Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* <https://doi.org/10.1038/nature14008> (2014).
- Kouba, T., Drncová, P. & Cusack, S. Structural snapshots of actively transcribing influenza polymerase. *Nat. Struct. Mol. Biol.* **26**, 460–470 (2019).
- Gerlach, P., Malet, H., Cusack, S. & Reguera, J. Structural insights into Bunyavirus replication and its regulation by the vRNA promoter. *Cell* **161**, 1267–1279 (2015).
- Guilligay, D. et al. The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat. Struct. Mol. Biol.* **15**, 500–506 (2008).
- Yap, T. L. et al. Crystal structure of the dengue virus RNA-dependent RNA polymerase catalytic domain at 1.85-angstrom resolution. *J. Virol.* **81**, 4753–4765 (2007).
- Liang, B. et al. Structure of the L protein of vesicular stomatitis virus from electron cryomicroscopy. *Cell* **162**, 314–327 (2015).
- Peng, R. et al. Structural insight into arenavirus replication machinery. *Nature* **579**, 615–619 (2020).
- Holm, L. Benchmarking fold detection by DaliLite v.5. *Bioinformatics* **35**, 5326–5327 (2019).
- Ferrero, D., Ferrer-Orta, C. & Verdaguer, N. Viral RNA-dependent RNA polymerases: a structural overview. *Subcell. Biochem.* **88**, 39–71 (2018).
- Shu, B. & Gong, P. Structural basis of viral RNA-dependent RNA polymerase catalysis and translocation. *PNAS* **113**, E4005–E4014 (2016).
- Nilsson, B. E., Te Velthuis, A. J. W. & Fodor, E. Role of the PB2 627 domain in Influenza A virus polymerase function. *J. Virol.* **91**, e00523–17 (2017).
- Barr, J. N. Bunyavirus mRNA synthesis is coupled to translation to prevent premature transcription termination. *RNA* **13**, 731–736 (2007).
- Wandzik, J. M. et al. A structure-based model for the complete transcription cycle of Influenza polymerase. *Cell* **181**, 877–893.e21 (2020).
- Vogel, D. et al. Structural and functional characterization of the severe fever with thrombocytopenia syndrome virus L protein. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkaa253> (2020).
- Wang, P. et al. Structure of severe fever with thrombocytopenia syndrome virus L protein elucidates the mechanisms of viral transcription initiation. *Nat. Microbiol.* <https://doi.org/10.1038/s41564-020-0712-2> (2020).
- Gong, P. & Peersen, O. B. Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase. *Proc. Natl Acad. Sci. USA* **107**, 22505–22510 (2010).
- Tickle, I. J. et al. STARANISO Anisotropy & Bayesian Estimation Server. <http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi> (Cambridge, United Kingdom: Global Phasing Ltd., 2018).
- McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D* **75**, 861–877 (2019).
- Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389–W394 (2015).
- Kandiah, E. et al. CM01: a facility for cryo-electron microscopy at the European Synchrotron. *Acta Crystallogr. D* **75**, 528–535 (2019).
- Schorb, M., Haberbosch, I., Hagen, W. J. H., Schwab, Y. & Mastrorade, D. N. Software tools for automated transmission electron microscopy. *Nat. Methods* **16**, 471–477 (2019).
- Li, X. et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 (2013).
- Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
- Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, e42166 (2018).
- Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
- Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
- Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D* **74**, 531–544 (2018).
- Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
- Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
- Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).

Acknowledgements

We thank Karine Huard, Angélique Fraudeau, and Alice Aubert for technical advices on expression and purification; Ambroise Desfosses, Leandro Estrozi, and Alister Burt for discussion on image processing; Aymeric Peuch for help with the usage of the EM computing cluster; Dominique Housset for general discussion on the project and Irina Gutsche for support. This work used the platforms of the Grenoble Instruct-ERIC center (ISBG; UMS 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology (PSB), supported by FRISBI (ANR-10-INBS-05-02), and GRAL, financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). The electron microscope facility is supported by the Auvergne-Rhône-Alpes Region, the Fondation Recherche Médicale (FRM), the fonds FEDER, and the GIS-Infrastructures en Biologie Santé et Agronomie (IBISA). We acknowledge the European Synchrotron Radiation Facility (ESRF) for provision of beam time on CM01. We thank all platform staff that enabled us to perform these analyses. This work was supported by the IDEX IRS G7H-IRS17H50 and the ANR-19-CE11-0024-02 to H.M.

Author contributions

B.A., P.G., and J.R. expressed and purified LACV-L FL. P.G. crystallized LACV-L FL. S.C. P.G. and J.R. solved the X-ray structure. S.C. built an initial model based on X-ray crystallography data. B.A. prepared cryo-EM grids. B.A. and H.M. collected cryo-EM data on a Thermofischer Glacios EM thanks to advices and training from G.S. who set up and maintains the IBS EM platform. G.E. set up SerialEM data collection scheme and collected the high-resolution cryo-EM dataset on the Thermofischer Krios EM at ESRF (CM01). B.A. performed advanced cryo-EM image processing and 3D reconstructions. H.M., B.A., and S.C. built the models based on the cryo-EM maps and performed structural analysis. H.M. and G.S. co-supervise B.A.; J.R. and S.C. co-supervised P.G. The project was conceived by H.M. and S.C. This project used funding obtained by H.M., G.S., and S.C. The manuscript was written by H.M. and B.A. with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-17349-4>.

Correspondence and requests for materials should be addressed to S.C. or H.M.

Peer review information *Nature Communications* thanks Yuan He, Aartjan te Velthuis and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

SUPPLEMENTARY INFORMATION

**Pre-initiation and elongation structures of full-length La Crosse virus polymerase
reveal functionally important conformational changes**

Benoît Arragain et al.

Supplementary Table 1 Data collection and refinement statistics

	LACV L FL Pre-initiation (6Z6B)
Data collection	
Beamline (ESRF)	ID29
Wavelength (Å)	0.9724
Space group	C2
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	371.19, 145.32, 234.19
α , β , γ (°)	90, 116.33, 90
Resolution (Å)	172.95-3.96 (4.35-3.96) *
<i>R</i> _{meas}	0.163 (1.285)
<i>I</i> / σI	6.5 (1.6)
CC(1/2)	0.995 (0.470)
Completeness (%)	
Spherical	60.8 (12.4)
Ellipsoidal	91.3 (55.8)
Redundancy	11.6 (13.3)
Refinement	
No. reflections work/free	57697/1514
<i>R</i> _{work}	0.264 (0.323)
<i>R</i> _{free}	0.299 (0.370)
No. protein atoms	34640
Average B-factors (Å ²)	189.8
R.m.s. deviations	
Bond lengths (Å)	0.002
Bond angles (°)	1.156
Validation	
Ramachandran favoured	90.0
Ramachandran outliers	1.04
Molprobity score	2.11
Clash score	3.07

*Values in parentheses are for highest-resolution shell.

Supplementary Table 2: Cryo-EM data collection, refinement and validation statistics

	LACV L Pre-initiation	LACV L CBD	LACV L ZBD without β-hairpin	LACV L Elongation
	(6Z6G) (EMD-11093)	(6Z6G) (EMD-11095)	(6Z6G) (EMD-11107)	(6Z8K) (EMD-11118)
Data collection and processing				
Microscope	ThermofischerTitan Krios			
Camera	Gatan K2 Summit + GIF			
Magnification	165000			
Voltage (kV)	300			
Number of frames	40			
Electron exposure (e ⁻ /Å ²)	50			
Defocus range (μm)	0.8-3.5			
Pixel size (Å)	0.4135 in super-resolution, 0.827 used for processing			
Symmetry imposed	C1			
Initial/Final micrographs (no.)	16498/16015			
Final particles (no.)	57660	131058	51842	59152
Map resolution (Å) 0.143 FSC threshold	3.06	3.54	3.49	3.02
Map resolution range (Å)	2.9-5	3.5-4.3	3.4-4.6	2.9-5
Refinement				
Map content during refinement	Entire pre-initiation map	CBD and mid domains	ZBD and mid domains	Entire elongation-mimicking map
Protein/ligand content during model refinement For RNA, if present but not built due to flexibility: indicated in italics	-Protein: endonuclease, core, mid thumb ring linker, mid, ZBD β-hairpin (residues 1-1841, 1978-2025, 2084-2102) -RNA: *5' vRNA (1-10): 5'-AGUAGUGU GC-3' *5' vRNA (9-16): 5' GCUACCAA-3' *3' vRNA (1-16): 3'-UCAUCACAU GAUGGUU-5'	-Protein: residues 1842-1977	-Protein: residues 2026-2083, 2103-2263 -Zinc ion	-Protein: same residues as in Pre-initiation -RNA: *5' vRNA (1-10): 5'-AGUAGUGU GC-3' *5' vRNA (9-16): 5' GCUACCAA-3' *3' vRNA (1-16): 3'-UCAUCACAU GAUGGUU-5' *Template mimicking RNA: 5'-AGUAGUGU GC-3' *Product mimicking RNA: 3'-UCAUCACAU GAUGGUU-5'
Initial model used	5AMQ	-	-	5AMQ
Model resolution (Å) 0.5 FSC	3.06	4.03	3.80	3.12

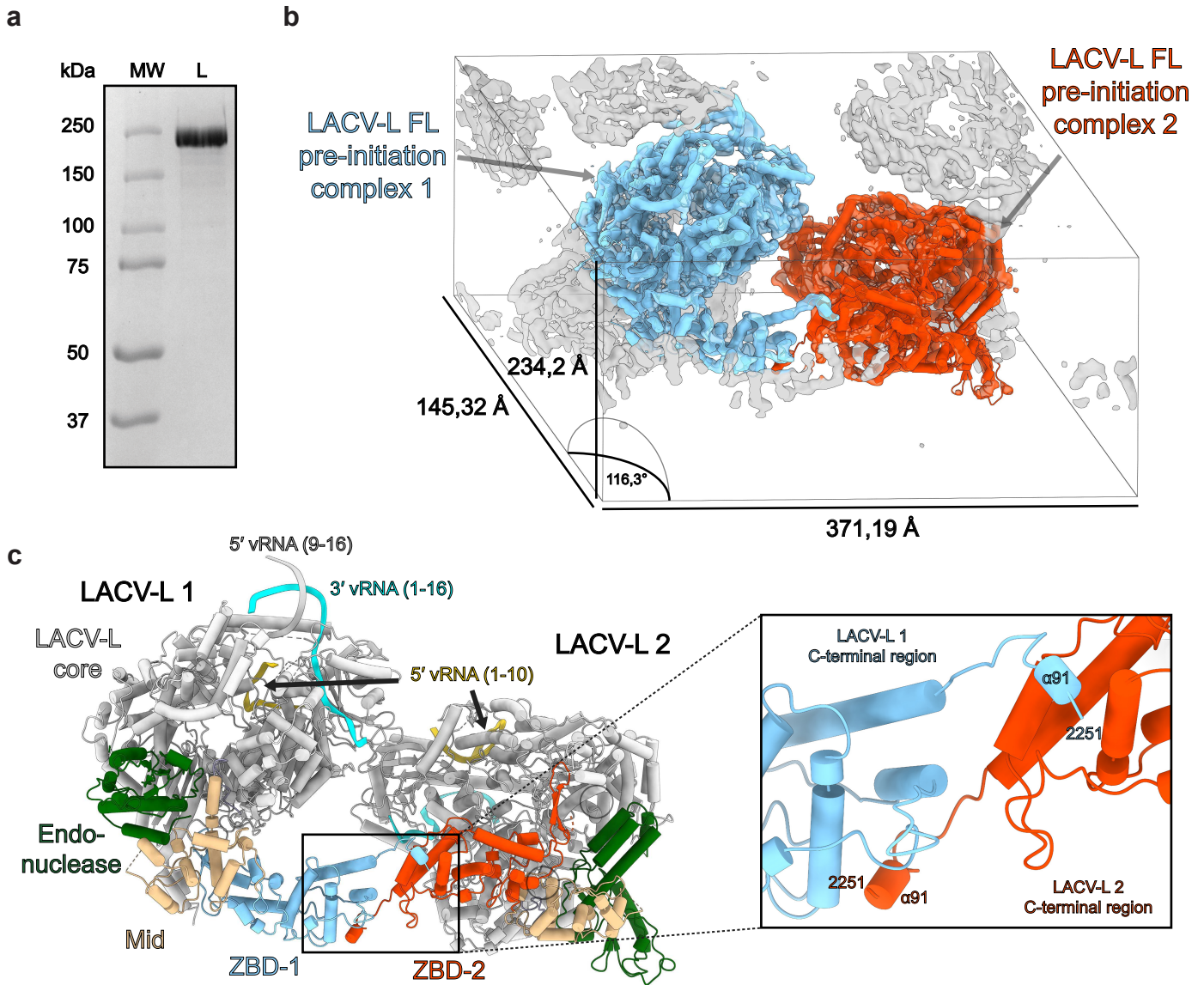
threshold				
Map sharpening B factor (Å ²)	-40	-90	-83	-40
Model composition				
Protein residues	1815	129	201	1831
Ligands	29	0	1	46
B-factor (Å ²)				
Protein	12.93-145.07	69.72-127.37	55.55-	40.59-158.48
min-max (mean)	(55.74)	(93.35)	135.31(84.74)	(74.67)
R.m.s deviations				
Bond lengths (Å)	0.008	0.005	0.005	0.009
Bond angles (°)	0.889	0.924	0.844	0.931
Validation				
MolProbity score	1.46	1.88	1.80	1.54
Clashscore	3.57	4.97	3.99	3.79
Poor rotamers (%)	0.24	0.00	0.53	0.53
Ramachandran plot				
Favored (%)	95.53	87.07	93.85	94.57
Allowed (%)	4.47	12.97	6.15	5.43
Disallowed (%)	0	0	0	0

Supplementary Table 3 codon-optimized gene sequence

atggccaccatcatcaccaccatgattatgatattccaactaccgagaatttggattttcaggggatggactaccaggaatatcaacagtcttggcagctatc
aacaccgcacgtgacgcttgcgtggccaaggatattgatgacgactgctgagccgtcacgactatfctggacgcgaactctgtaaaagctgaaacatc
gaataccgtaacgacgtgaccttcatcgacatcattctggacatcgtcctgaggtgacccgctgaccatcgacgcaccgcacatcaccctgataactatct
gtacattaataacgtgctgtacatcattgattacaagggttagcgtctccaatgaaagcagcgtcattacgtacgacaagtactcgaactgaccgcgatact
ccgaccgtctgacatcccgatcgaattgttattaccgtattgatccagtggccgcatctgcatatcaactccgatcgtttcaaggaaattgatccgactat
cgtggctgataattaactttaaccagttcttgaccicaagcaactgctgtacgaaaagttcggcgtgacgaggaattctgctgaaggctgcacatggtgatttt
acctgactgccccgtggtgtaagcggctgtccgaaatttggagcaccgattfacaaggagttcaagatgagcatgccggtgccgaacgtcgtctg
ttcgaggagcgtcaaaftcaacgctacgagtcgagcggctggaacaccaactggttaaaatccgtgaatacactaaaaggactatagcgagcaca
tagcaagtctgcgaaaacatttctgctgctcgggttctacaagcaactcaaaaaacgagatttgaaggctggaccctgatggtgagcgtgtccag
gaccaacgcgagatctaaaagcctgcacgatcaaaagccctctatccatttcatctgggtgcccacaaccgggaaacgcaacaatgcacgttcaa
gttgatcgtgtagcaaaagcttgcagagcattaaaggatctccactacaccgaggcttttaaatccctgggcaaaatgatggatattggtgacaaggca
atcgatgacaggagttctgatgacgctgaaaagcaaggcgcgcagcagctggaagcagattatgaataagaaagctggaactaaagcaatcaaac
gactggtcctgtgggaacaacagttatgatcaacaatgacctgatgataaaagcgaagaaagctgaaagcttcaaaaatttctgctgacgttgaatacaca
agcagttcaagaacaagatgctggaggacgtggaggtcagcaaacgaagattctggaatttgacgacccaatgtacctggccagcctgacctgatg
gaacagagcaaaaagatcctgccaagagcaatggtctgaaacctgacaactttatcttgaacgaattcggcagccgtatfaaagatgcaaacaggaaac
gtacgataatgcaaaaactttgaaacgggttattggcaatgtatcagcgatttcagcagcgtcatgaaagaacatcttgagcgtacgccaatataatcgc
acaatactttcgaattgcatgctgcgcaacaataacgcttctgcgacgtcttccgagcggcgaatfaaaacaaaaggctacggtggtttacagcat
cattgttctgcacaaggaggaagaaacatcttcaatcccggctgctgacggtaccttcaagtgcagatgaaaggctacatcagcctctcgcgccatccgc
ctggataaggagcgttgcaaacgcatcgtgagctcccaggacttcttaccacctgcttcttcaagatgataccccacgctggtgatgagcgtat
attatgaactttgacattacaccagcctgtccattaccaagctgtcttgaagcttgaccgagccagcccgttatatgatcatgaactctctggtatcagcagca
acgtcaaggactatctgcgaaaaattctctccgtacagaaagacctgttctcgtctacatgaccgcttgattaaagatgctgcttcgacgatacgtat
cagcgtcagcgcgttcaactgctgcgacatttactgtctgactatgatacaccaaaaaggtatcaaaagacaatcgcgaattgacctcctctgttcccgg
cagcgttactgtgaaggaaatctgaccagatctacctgcaatttactcaatgccaaggcctgcatgagaagcaccgctacgtatggtgacctggcgaag
actatcctgagatcgagtgtagcagcgtgaaaatatacaggagatctggagcactaattgcagcaaacagaccgttaacctcaagatcctgattcactcc
ctgtgcaaaaactgctggcggacacgtctctgcatataaccttgcgcgaaccgcaattgaaaaccgtaataacttccgctgttctatcagccatcagcacttt
acgtccagcaagcttctgctgaagattggtgacttctgtaaaagaaagagcgtgcaatccgtfaagcagaaagagatcctggaggtgcaatccgtaagatg
cgtttggcctaaccgatgtttgtgacggagcaggtttgctggaggttggccattgcaactacgaaatgctccgtaacgcaatgcccgaattacaccgatt
acatttccaccaaggtctttgatcgctgtacgaactgttgacaagaaagcttgcagggataaacctgtgattgagcagatcatggatgatgatcaccac
aagaagtctatttcaacttcttaaaagggccagaaaacgtcaaaagaccgcaaatttctgctggatgaaatgaaagctaaaatgctgatgacgctgga
acgtatcgtcaaaagcgttgcaaacactcaaccagacgaaatgattagcagccgggagacggtaaagctgaaagctggaacaaaagctcagcaaga
aatccgtttctcgttgagaccactcgcagaaaaaccgtgaaatgacgaagccatcgaggcttggctacggaaggatacgaagcaacctgggtaaa
atcgagaaatgctctggtaaggcgaaggcctgaagatggagatcaacgccgatafgagcaaatggaagcgtcaggatggttctataaatacttctgg
ttgatcgcgctggaccgattttgatccgagagaaggacgcaattctacttctgcaattacatgacaaggagctgatcctgccagacgagctgc
tgttaatctgtggacaaaagggtggcgtaccagaatgacattatcgcgacctgaccaatcaactgaattcaaacaccgtcgtgattaaactgtaactgctg
cagggtaatttcaactacacttctgctacgttcacagctgcgcgatgagcgtctacaaaagatcttgaagagcgttactctctggagcgtctatcttg
gtaacagcctcgttacttctgatgataaccagacctctaccattgtccaagataaaatggagaacgacaagatcactgacttccatgaaagagttcga
acgtgcttgcctgacgttggctgtcaggcgaatataagaaaactacgtcaccatgtcatcaaggagttgttagctgtttaaactgtatggtgagccgtt
cagcatttaccggccgttctgctgaccagcgtcggcactcgcgtacattgttctttaggagctggcagccgatcagctccgccagaccgcaat
caagcagcgttggccgagcctggttgggtgagcagcaccattgaccactggatgacgagcctgacttacaatgctgccgggtcagagcaatgacc
cgatcgattacttccggcggaacccgtaaggacatccctatcgaactgaacggtgttctggagcgtccttctgtatgatttccaccgtgggtctgaaagc
ggaaatctgatttctgatcaagctcctgagcaaatatacggcgtgatgcaaaaagcgtgagtcctgttaaccagattgcagaggctcaagaattggaag
ttgaggactgactgacaacgaaatcttgcctcaaaaactcctccgctacttggttctggtatgaggagatggaccgagcgaatcatggcgaacactccg
acatgocggctgtagcattctgacgccagtaaatcagaccgcgggtgacctccgtaagttgtaacgctttagcaagtaccaggatcgtttgtctcccc
gggtggtatggtgagctgttctacgtatctgctggagaacctgagctgctggttactaaggcgaaggacatgaagactacatggaagcgtgattttccg
ctacaacgcaaacgtttaaagaatcttggacattcgaacccccgcaactgttcttgaacaaatctgttcagccacaagccgtcattgactttccgg
cattcgcgataatacattatgtatgactcccgccttggaaaaggacccggacatcctcggtaaaagtacgttccaggaagcgtaccgtctgctgatg
cgtgatttgaactctgagctgactaacgatgacatccaggtcacttctctatattctgaacgaccgatgatgaccatcgcgaaacccacat
cttgcatttaccggttccccgcaacgccgatggcagctctcgcagcagcagcgggaatttctgaaatctgaaagctgattcaccacagccagccctggtc
ctgctgcttatttcaagaacaatccggatatacaggctgcccagcagcgaatggccgtgacctgggtgcatctgaaagaaattcgtcaaaactaac
ctggaagaaagatgaaagtcgcacgctatgaaagctgaaggtcgaaggtcgaactgtgacattgttttcaagcgtgaaagagatgacgcgttttaccagttt
gctacgagacgtgaaatcaccggagcacaagatcaaaagcttctcctcccgcaagagctataccaccagcagcttctgctccctgatgagggcaatt
tgattaaagcaaaagatggtacaccgtccacttctgaaacagattctgtccggtggccataaggctatcatgcaacacaacgcgaccagcgagcagaac

atcgctttgagtgctcaagctgatcactcactttgaggatagcttcattgacagcttgctcgtagcgcgttctgcagctgatcattgacgagtttagctataa
agacgttaaagtgagcaagttgtacgacattatfaagaacggctacaaccgfaccgatttcaccccctgctgtccgfactggtagcctgcgccaggcagac
ctggataaatacagcgaatgaagtcccacgagcgcgtcacctggaacgattggcaaacgtcccgtcatttggacatgggtccattaacctgaccattacc
ggctataaccgtagcatcactatcatcggtaggacaataagttgacctaccgagctgtgtctgacgcgcaagactcctgaaaacattaccatctccggtc
gtaaattgctgggcagccgtcatggttgaagttcgaaaacatgagcaagatcagacctatccgggtaactactataattacatctgtaaaaaggatcgtcat
cagttcgtctaccaaatccactcccacgaaagcattacgcgtcgtaacgaagacacatggcgattcgtaccgtatctacaatgagattacgccggctctgtg
ttgtaatgtcggcagggtggagagatcagcgtattctgattcgtagcctggactatttgaataatgacattttagcctctcccattaaagtggctggac
gaattcgtaccatcaagaaagcgattttagcaagatggtgtcttcgagggacctccattaagactggtctgctggatctgaccgagctgatgaagagcc
aggatctgtgaatctgaattacgacaacattcgcaatagcaacctgatttccttctctaaactgatctgctgcgagggtagcgataacatcaacgacggtctg
gagttcctgagcgacgatccgatgaatttcaccgagggcgaggcaatccacagcaccctgatttcaatattactattctaaactgggtgaacgtcacatga
cctaccgtaatcgatataactgctgatcagcgtgaaaccaagatctttgaggaggecttcacgttttagcgaaaatggcttcattagccctgaaaacctggg
atgcctggaagccgtcgtgagcctcatcaagctgctgaagacgaatgagtggtctaccgtgatcgacaaatgtattcacatctgtctcatcaagaacggaatg
gatcacatgtaccacagcttcgagtgccgaaatgtttatgggaaacccgattactcgtgacattaactgggtcatgttccgtgagttatcaacagcctgccg
ggtaccgacatcccggcggagcgtcatgactgagaactcaagaagaatgtatcgtctgattaacagcaaatggagacgcaacgcgacttctctgag
tttacgaaactgatgaaaaaggaggcggtcgtccaacattgattcactaatga

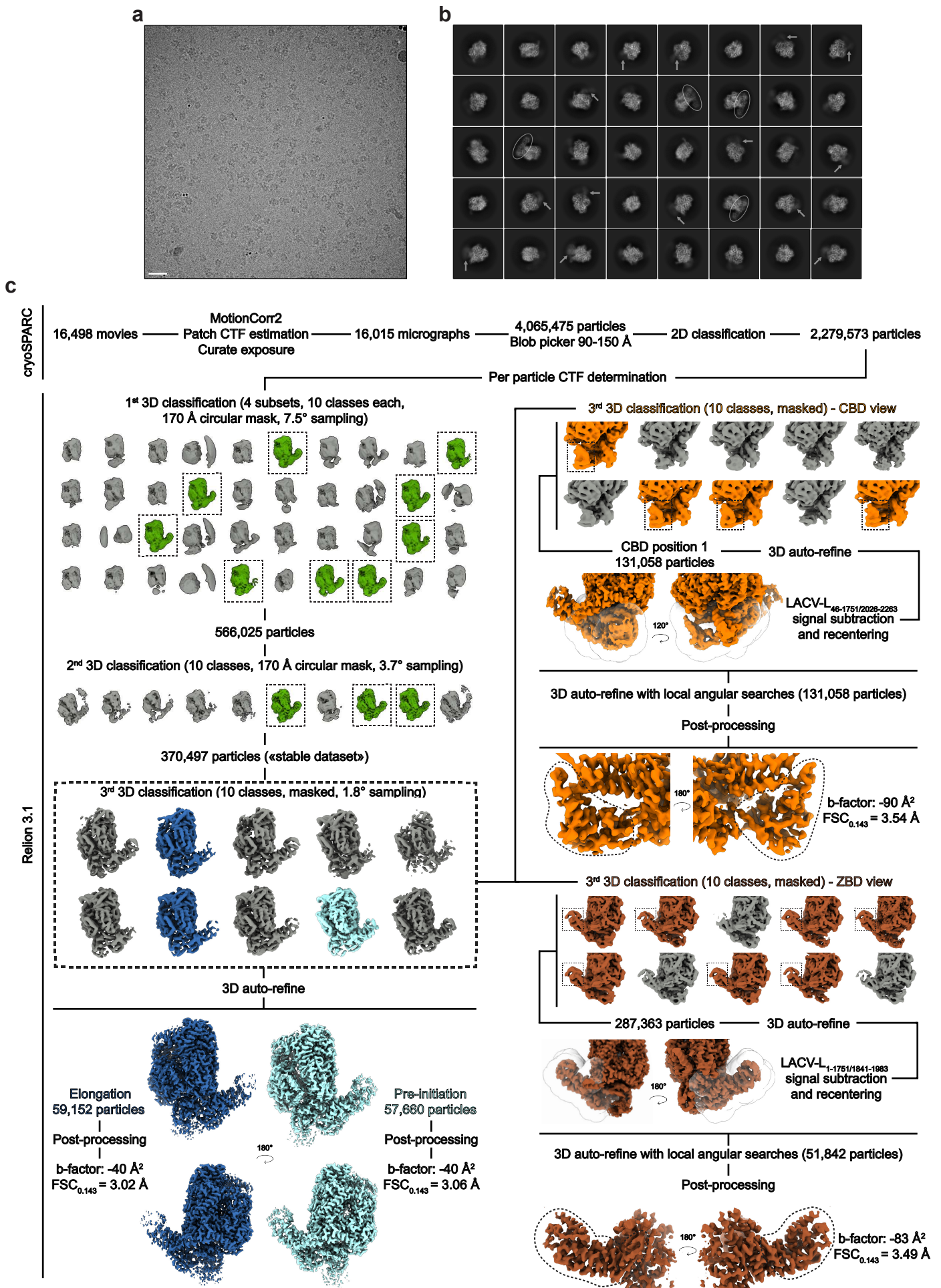
Supplementary Figure 1



Supplementary figure 1: Purification and X-ray crystal structure

a, 8% SDS-PAGE gel of LACV-L FL purification after gel filtration step (MW: Molecular weight; L: LACV-L FL). **b**, Electron density map of one asymmetric unit. The two polymerase complexes at pre-initiation are respectively colored in blue, red and labelled. Asymmetric unit dimension and angles are displayed. **c**, X-ray crystallography structure shown in the same orientation as in **b** displaying the 2 polymerase complexes of the asymmetric unit. The core regions are colored in white, the endonucleases in green, the mid domains in beige and the ZBD according to LACV-L 1/LACV-L 2 colors in blue and orange. The swap of the last α -helix 91 is shown in a close up view and the C-terminal residue 2251 is labelled. The RNA promoter positions are shown.

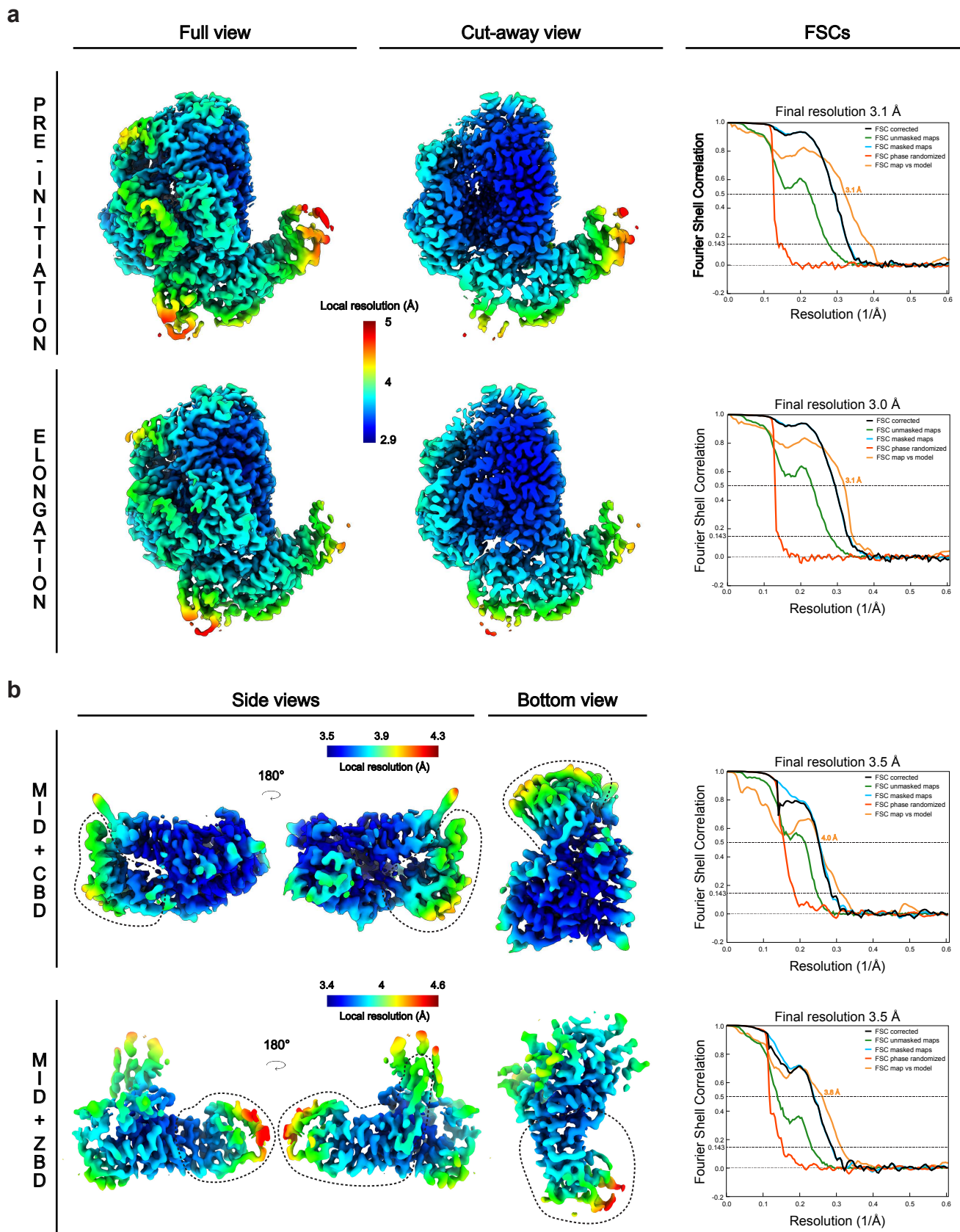
Supplementary Figure 2



Supplementary figure 2: Cryo-EM data collection and image processing

a, Cryo-electron micrograph of LACV-L FL collected on the ESRF CM01 Titan Krios equipped with a K2 direct electrons detector camera at -2 μm defocus. Scale bar = 200 Å. **b**, Representative 2D classes of LACV-L FL. Flexibility of the LACV-L C-terminal region seen in some class averages is highlighted with white arrows. Stabilized LACV-L C-terminal regions visible in other class averages are surrounded by a white circle. **c**, Cryo-EM image processing pipeline is described in detail in the “Image processing” method section.

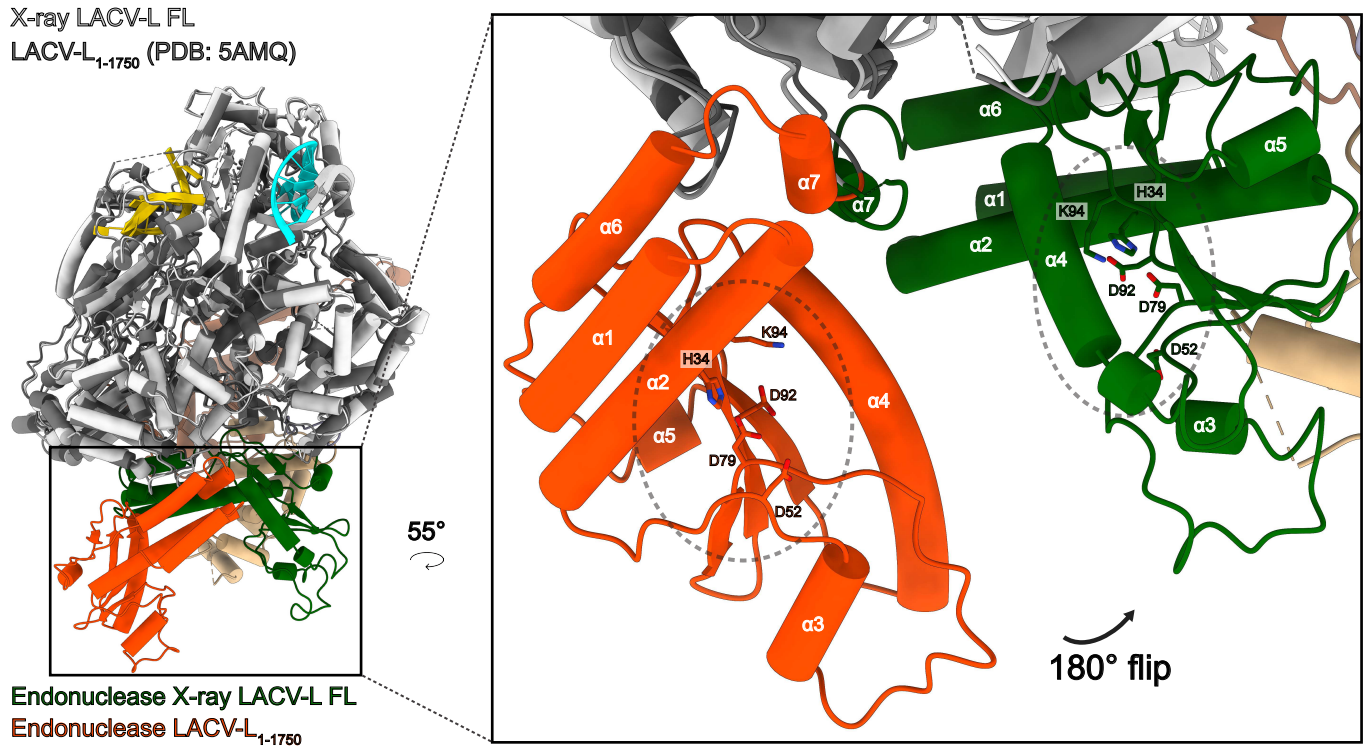
Supplementary Figure 3



Supplementary figure 3: Cryo-EM maps local resolution and FSCs

a, Full view and cut-away view of the pre-initiation and elongation-mimicking cryo-EM maps. Maps are filtered and colored according to their local resolution. Resolution range coloring is indicated. FSCs corrected, FSCs of unmasked maps, FSCs of masked map and phase randomized FSCs are respectively displayed in black, green, light blue and red. The FSC of the map and the model are displayed in orange. The model refined in the pre-initiation and the elongation-mimicking cryo-EM maps corresponds to the core, the mid domain, the endonuclease domain and the ZBD β -hairpin. The gold-standard Fourier shell correlation (FSC) of masked maps indicates a respective resolution of 3.1 Å and 3.0 Å for the pre-initiation and the elongation-mimicking cryo-EM maps with the FSC = 0.143 criteria. **b**, Side view and bottom view of the subtracted maps containing (i) the CBD and the mid domain, (ii) the ZBD and the mid domain. The maps are filtered and colored according to their local resolution. Resolution range coloring is indicated. The CBD position is surrounded by a dotted line in the map containing the CBD and the mid domain. The ZBD position is surrounded by a dotted line in the map containing the ZBD and the mid domain. FSCs are displayed and colored as in **a**. Gold-standard Fourier shell correlation of masked maps indicate a resolution of 3.5 Å for both the CBD-mid domain map and the ZBD-mid domain map (FSC = 0.143 criteria).

Supplementary Figure 4



Supplementary figure 4: Endonuclease movement between LACV-L FL and LACV-L₁₋₁₇₅₀

Superimposition of LACV-L FL and LACV-L₁₋₁₇₅₀ (PDB: 5AMQ). LACV-L FL core is shown in light grey, LACV-L₁₋₁₇₅₀ core in dark grey. LACV-L FL endonuclease is shown in green, LACV-L₁₋₁₇₅₀ endonuclease in red. The 180° rotation of the endonuclease domain is indicated. On the close-up view, the endonuclease active site is surrounded by a dotted ellipse, residues are displayed and α -helices are numbered.

Supplementary Figure 5

a

Bunyavirales

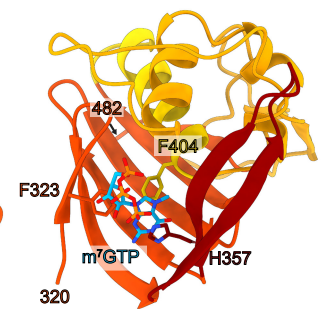
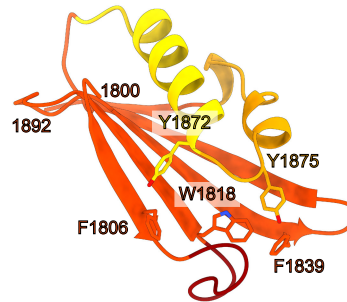
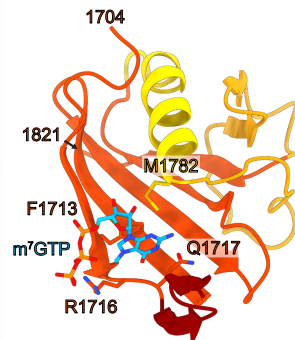
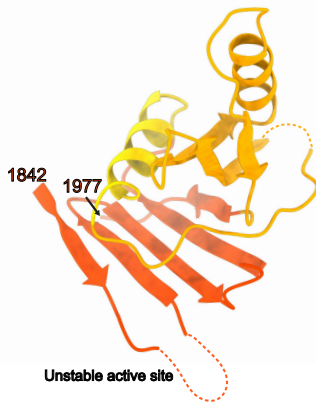
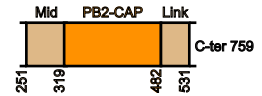
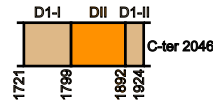
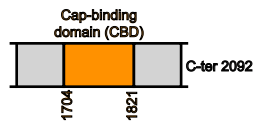
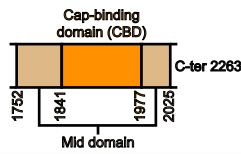
Articulavirales

LACV
Peribunyaviridae

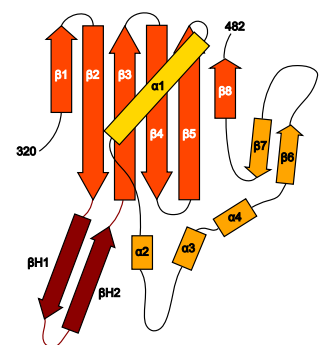
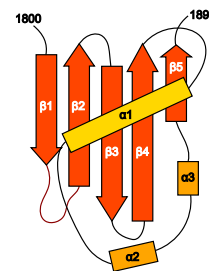
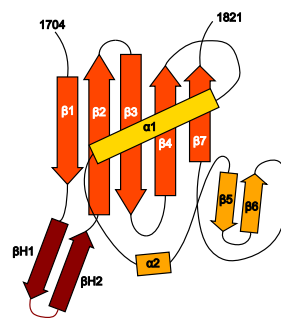
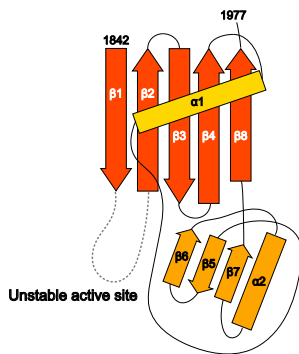
RVFV
Phenuiviridae

CASV
Arenaviridae

Influenza A
Orthomyxoviridae



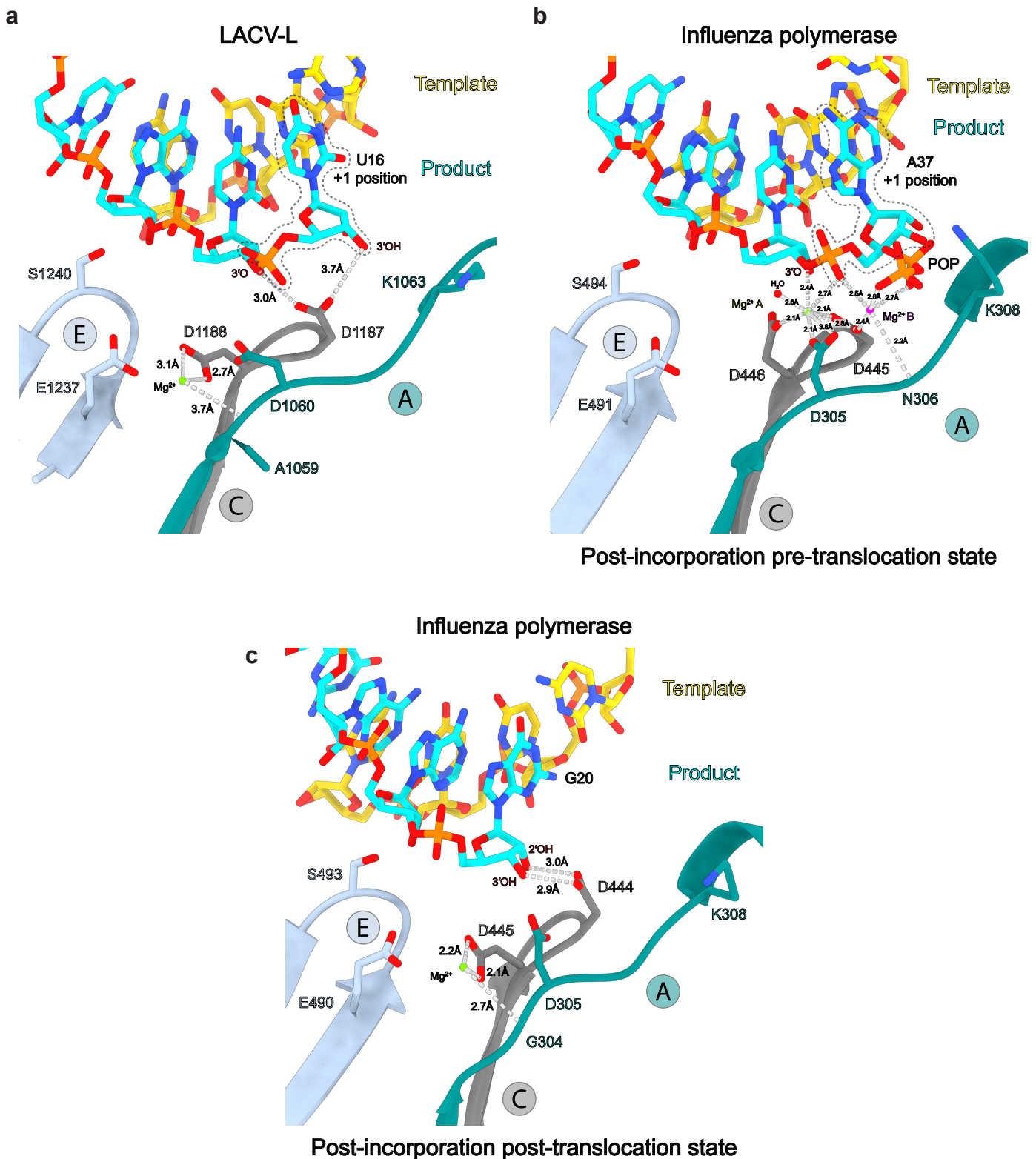
b



Supplementary figure 5: Structural comparison of sNSV CBD

a, Schematic representation of domains surrounding the CBD of LACV, RVFV, CASV and influenza A virus. Structures of LACV-L CBD, RVFV-L CBD (PDB: 6QHG), CASV-L putative CBD (PDB: 5MUZ) and influenza A PB2-CAP (PDB: 2VQZ) shown as ribbon. Similar structure elements are depicted in the same color: structurally conserved β -sheet in dark orange, α -helix in gold and specific insertions in orange. Residues involved in cap binding are shown and labelled. m^7GTP molecules are colored in cyan. **b**, Schematic representation of CBDs shown as in **a**. LACV-L CBD loop suggested to contain the m^7GTP binding site is shown as a dotted line. For clarity, secondary structures are numbered in the same way starting from $\alpha 1$ and $\beta 1$.

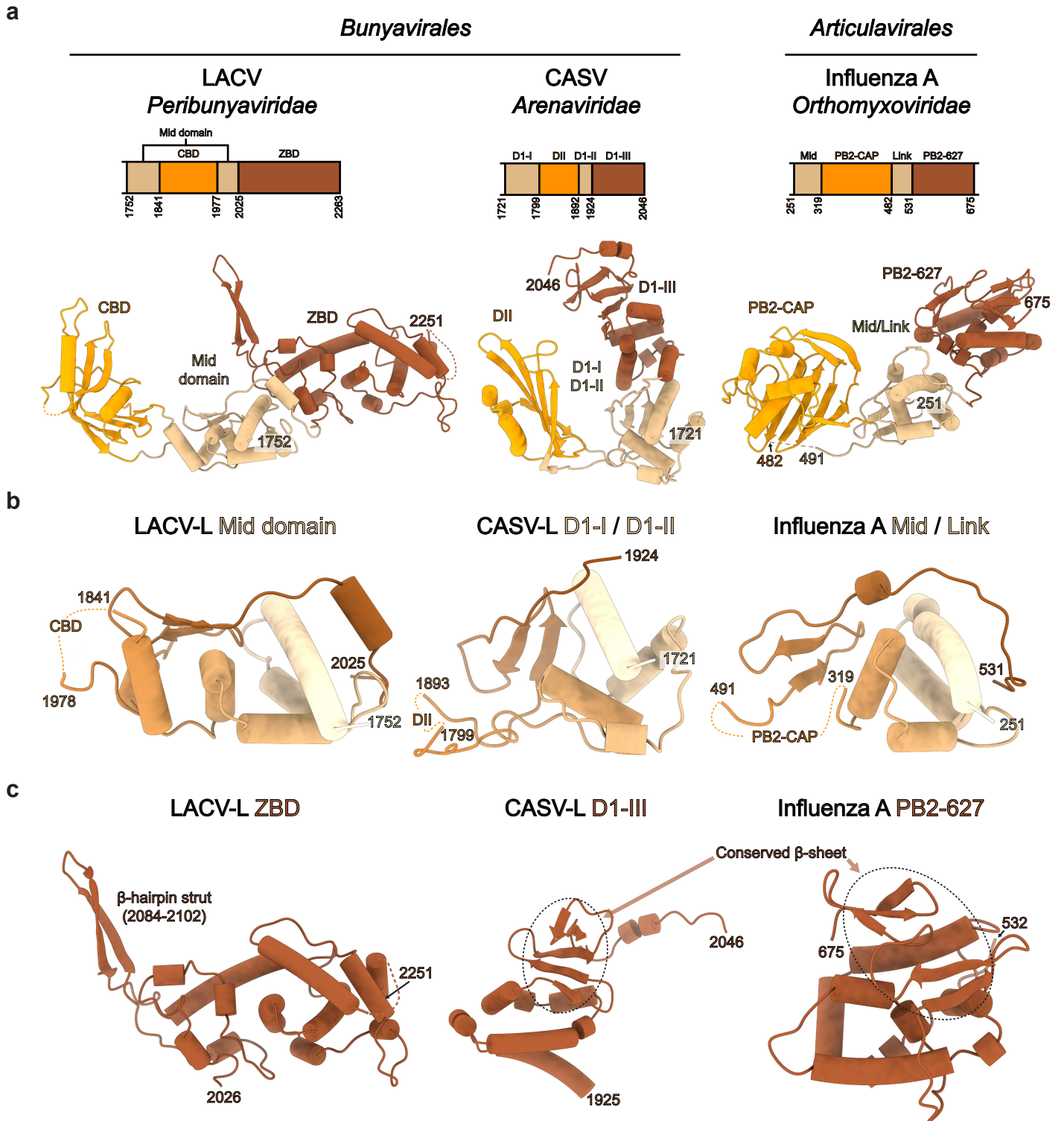
Supplementary Figure 6



Supplementary figure 6: Comparison between LACV-L and influenza polymerases active sites states.

a, b, c, The template, product, motif A, C and E are respectively displayed in yellow, cyan, sea green, grey and dark grey. The last incorporated nucleotides are circled by a dotted line. Distances between active site residues and coordinated ion(s) are labelled and shown as grey dotted lines. **a,** LACV-L active site at elongation-mimicking stage. Residues A1059, D1060, D1188 and E1237 coordinate a magnesium ion. D1187 interacts with the 3'O of the previous nucleotide. **b,** Influenza polymerase active site in a post-incorporation pre-translocation state (PDB: 6SZU). Magnesium ion in position A is colored in green, coordinated by D305, D445, D446, 3'O of the previous nucleotide and a water molecule (H₂O) colored in red. Magnesium ion in position B is colored in purple, coordinated by N306, D445, the +1 nucleotide phosphate and a pyrophosphate (POP). **c,** Influenza polymerase active site in a post-incorporation post-translocation state (PDB: 6QCT). Residues G304, D305, D445 and E490 coordinate a magnesium ion. D444 interacts with the 2'OH and 3'OH of the last nucleotide.

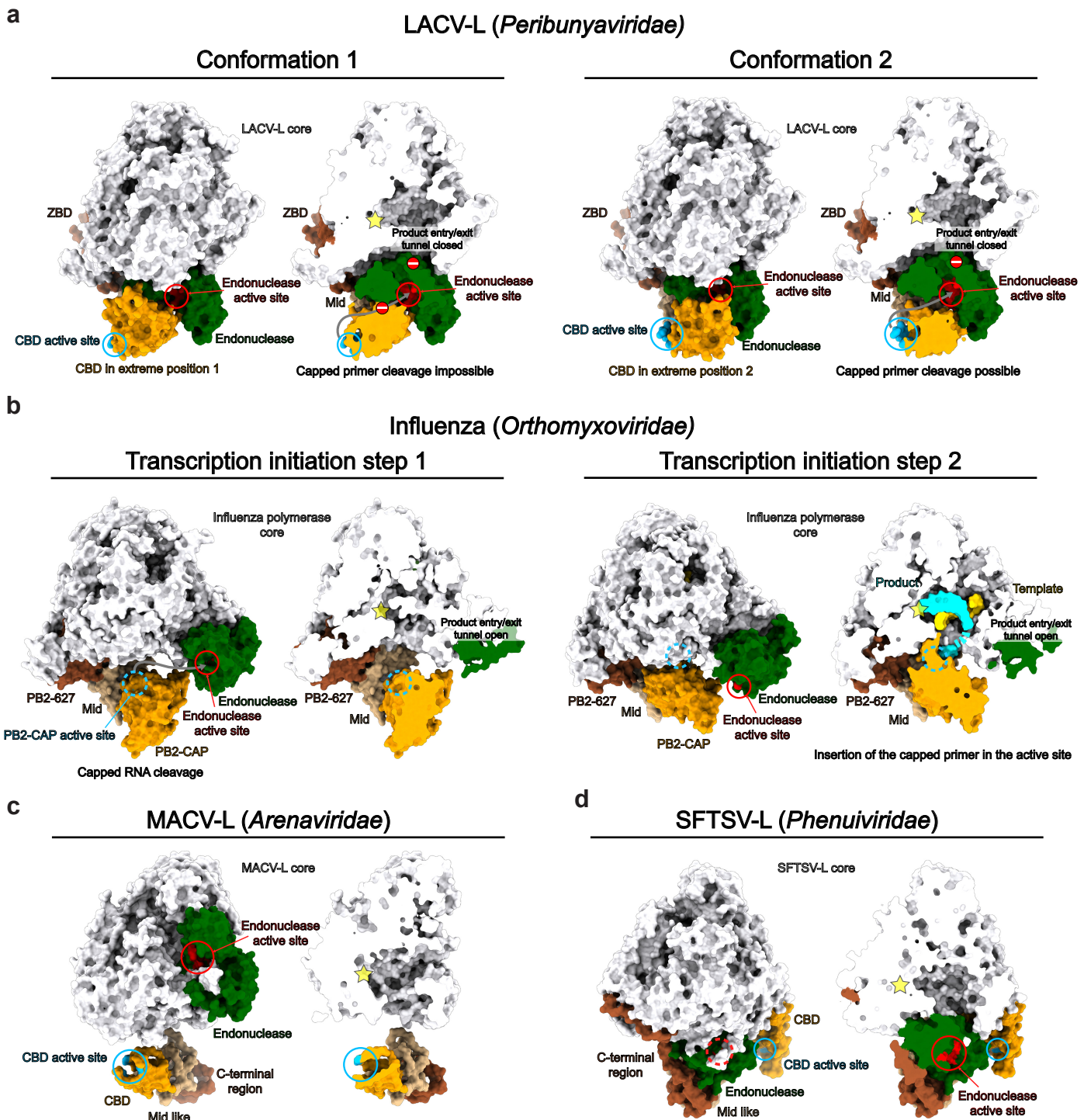
Supplementary Figure 7



Supplementary figure 7: Comparison of C-terminal domain organization between sNSV polymerases

a, C-terminal domains of LACV-L, CASV-L (PDB: 5MUS) and influenza A virus polymerase (PDB: 4WSB): schematic representation (top) and structure (bottom). Similar domains are colored in the same way. **b**, Superimposed structures of LACV-L mid domain, CASV D1-I/D1-II and influenza A mid/link. Rainbow colors from beige to brown is used to color from the N-terminal to the C-terminal. **c**, Structure of LACV-L ZBD, CASV D1-III and influenza A PB2-627. CASV D1-III and influenza A PB2-627 domains are superimposed. Their conserved β -sheet is surrounded by a dotted line.

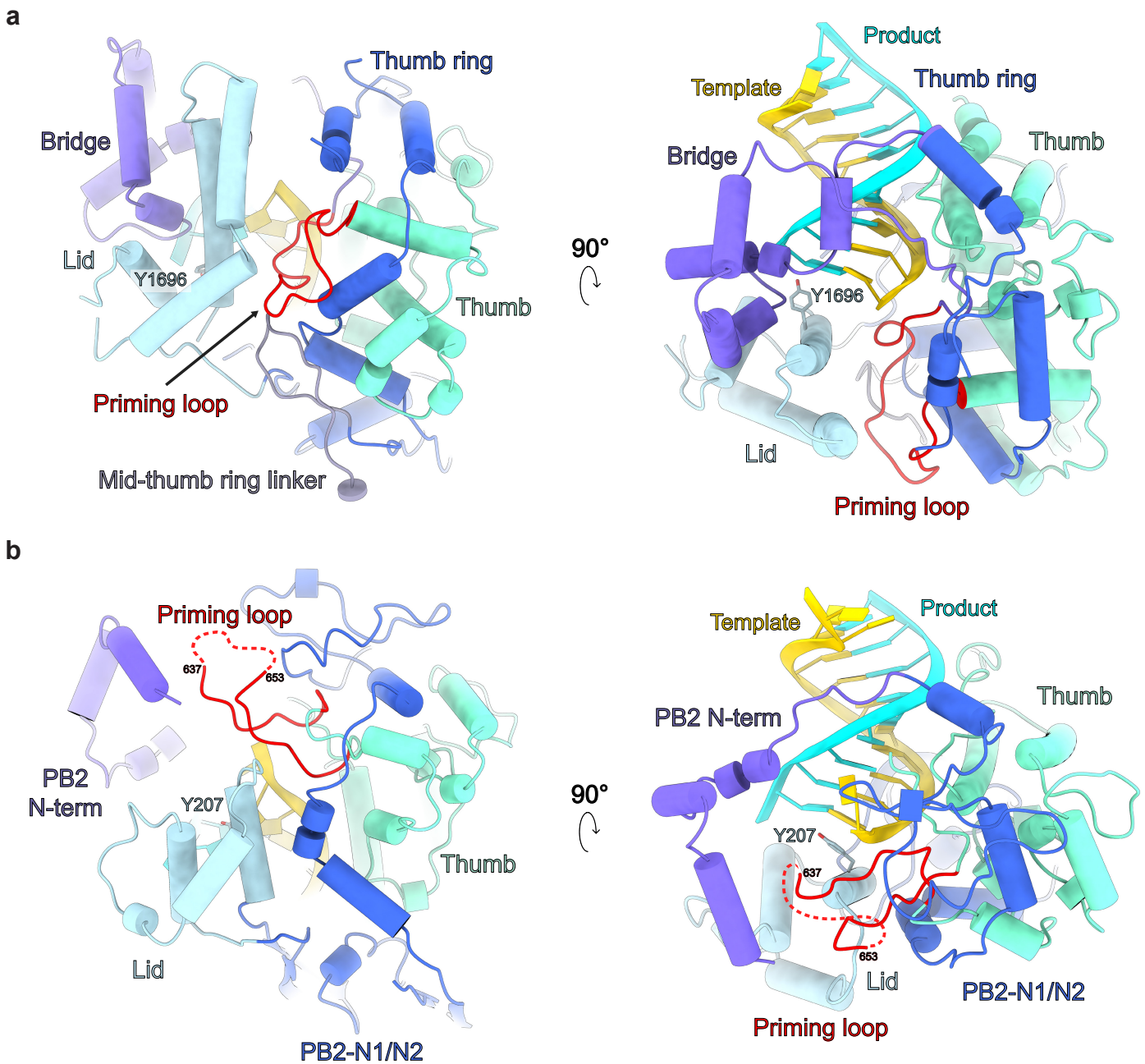
Supplementary Figure 8



Supplementary figure 8: Position of the endonuclease and C-terminal regions of LACV-L, influenza virus polymerase, Machupo polymerase (MACV-L) and Severe fever with thrombocytopenia syndrome virus polymerase (SFTSV-L)

All polymerase structures are aligned according to the polymerase core and are displayed as molecular surface. For each polymerase and/or conformation, both full and cut views are displayed. **a**, LACV-L in conformation 1 and 2. LACV-L core is shown in white, the endonuclease in green with its active site in red. LACV-L CBD is colored in orange with its active site in blue. Both mid domain and ZBD are colored respectively in beige and brown. Polymerase core active site location is shown as a yellow star. Grey arrow mimics the path of the bound capped primer through the endonuclease active site. Its access is closed in CBD extreme position 1 (stop sign) and opened in CBD extreme position 2. The product entry/exit tunnel is closed in both conformation (stop sign). **b**, Influenza polymerase in transcription initiation step 1 (PDB: 4WSB) and 2 (PDB: 6QCT). Same colors as in **a** are used for similar and/or related domains. PB2-CAP active site is not visible but represented as a dotted blue circle. Grey arrow mimics the capped RNA cleavage in step 1. In step 2, product and template are respectively colored in cyan and yellow. Product entry/exit tunnel stay open in both conformations. **c**, MACV-L monomer bound to vRNA was extracted from the polymerase dimer (PDB: 6KLH). **d**, SFTSV-L (PDB: 6L42). The endonuclease active site (not visible on the full view) is represented as a dotted red circle.

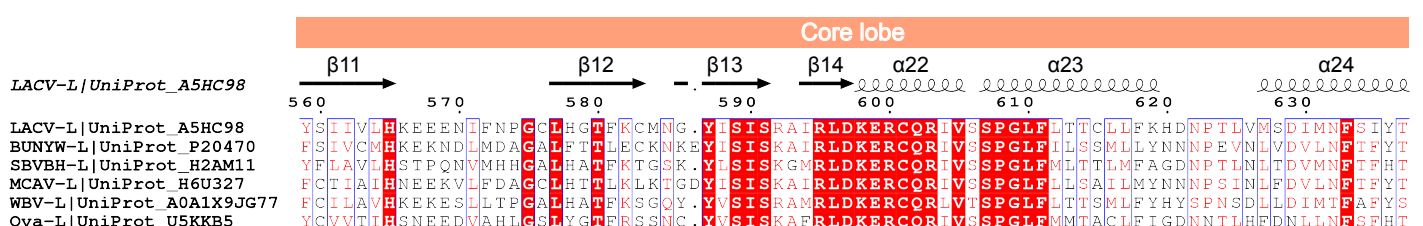
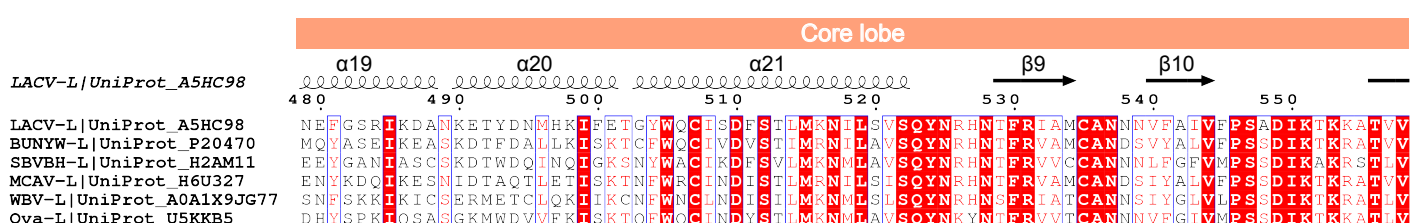
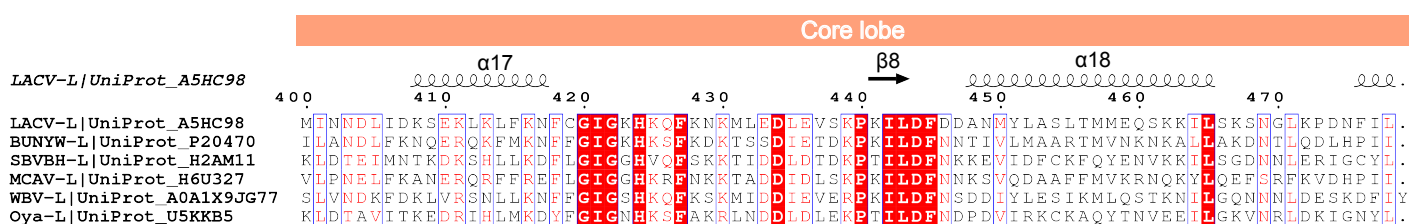
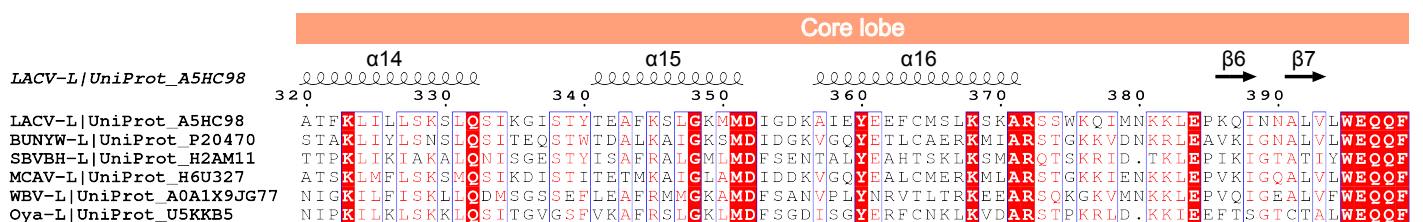
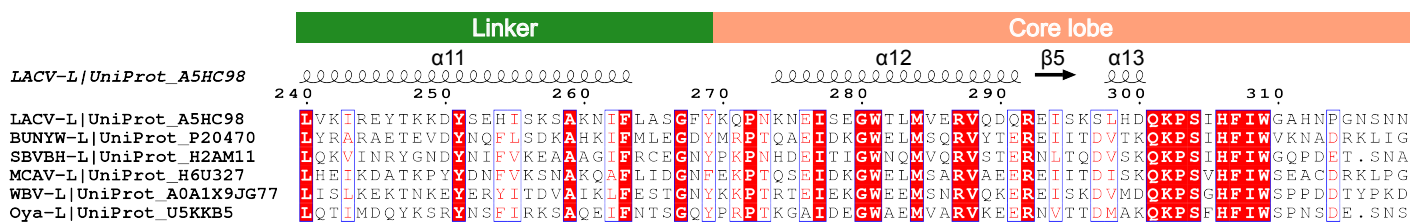
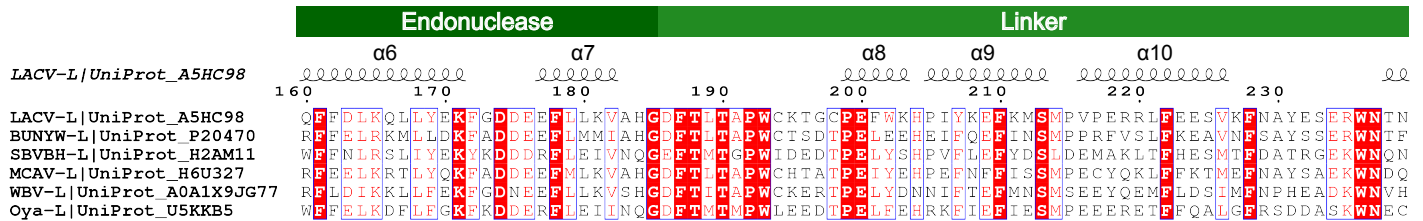
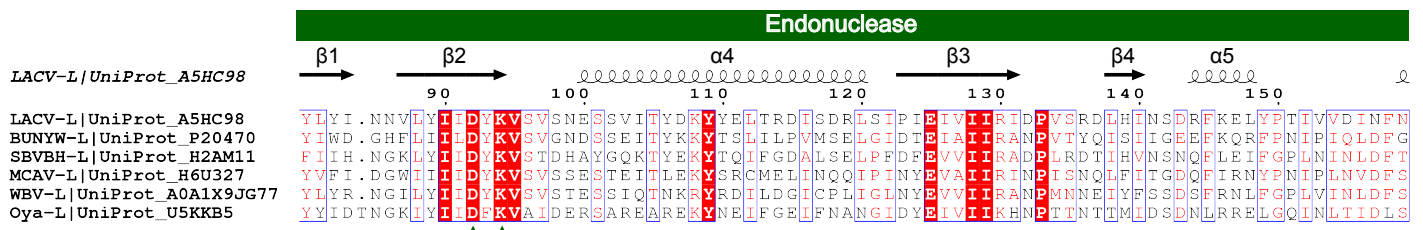
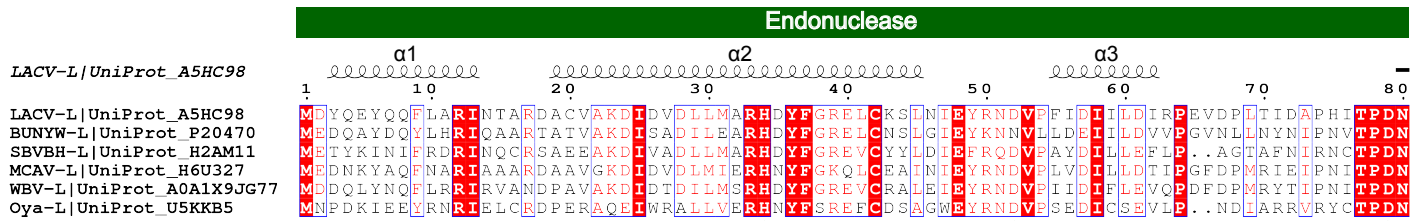
Supplementary Figure 9



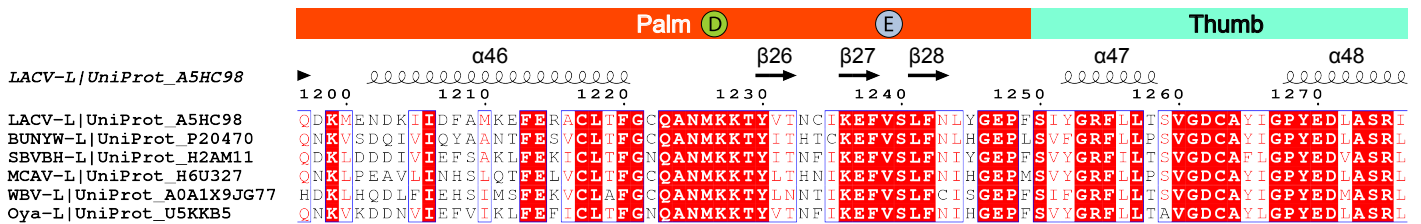
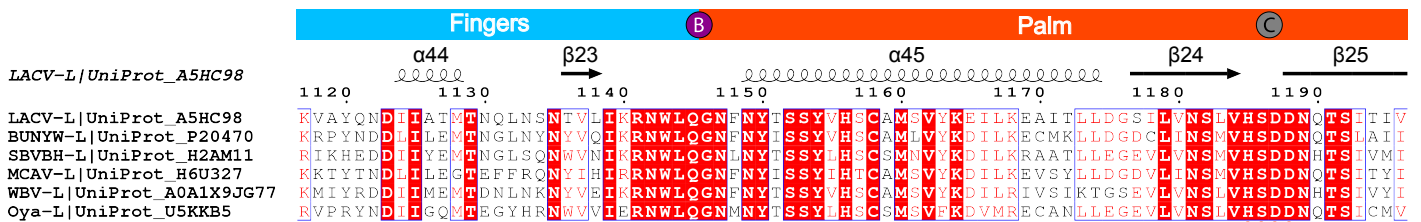
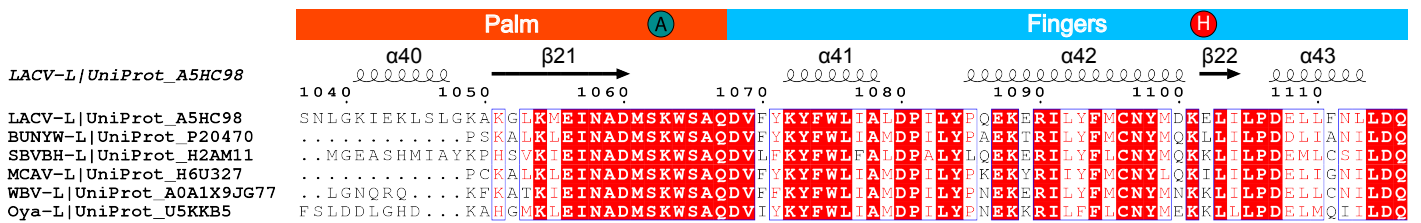
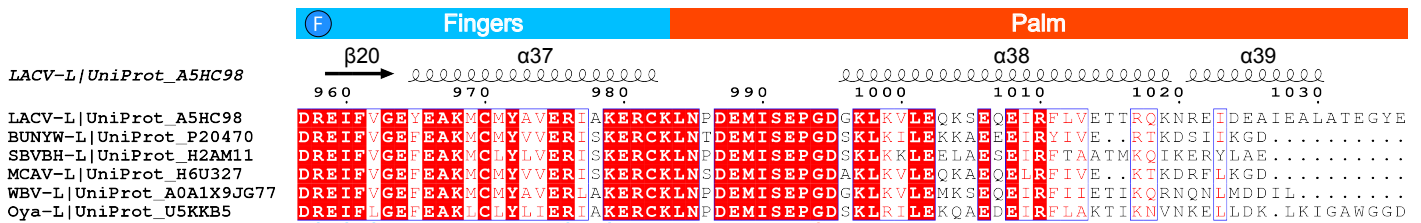
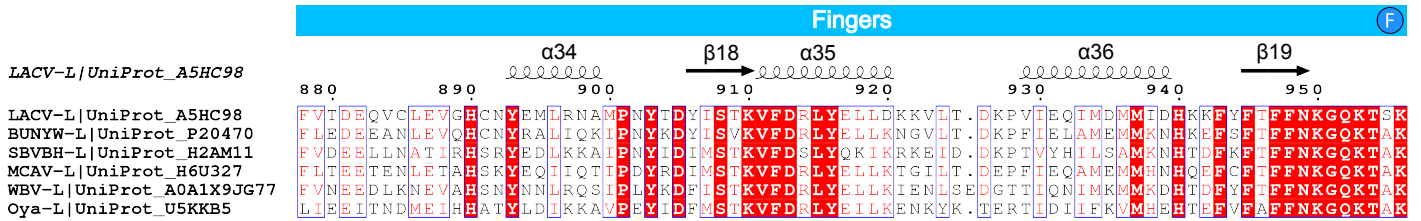
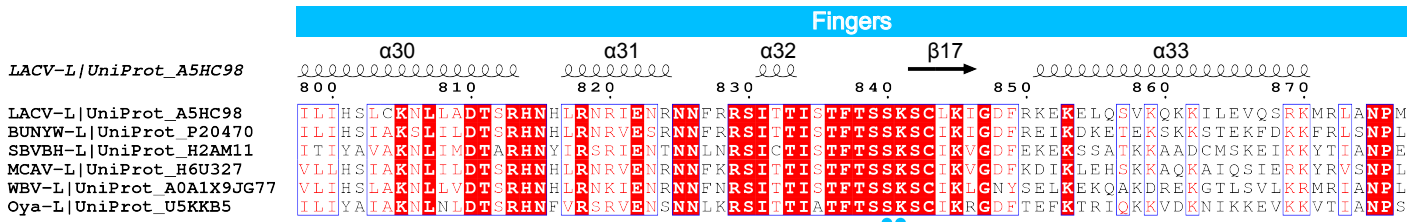
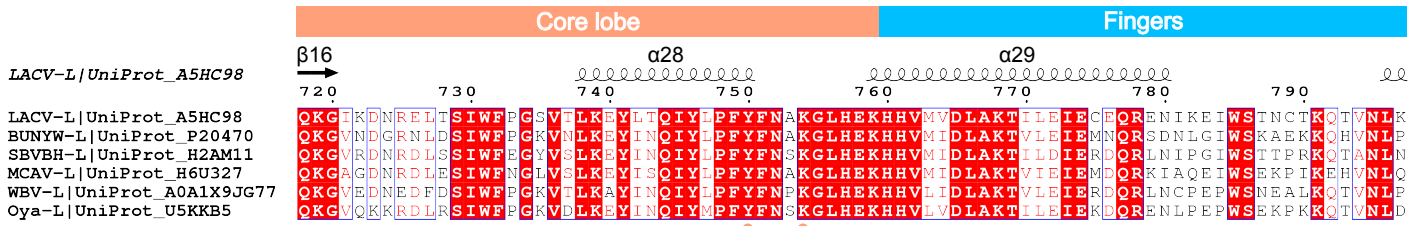
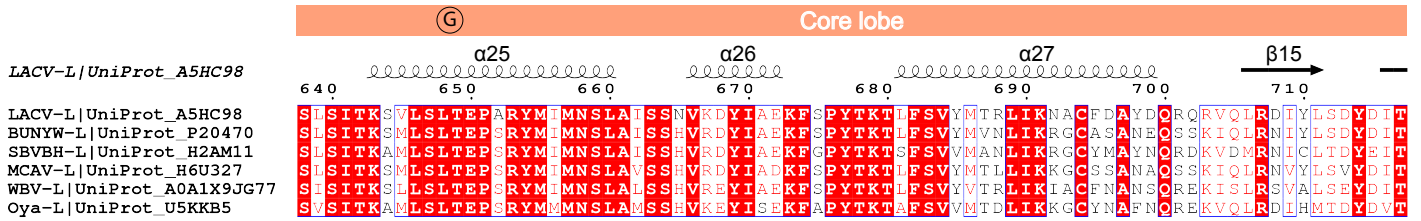
Supplementary figure 9: Comparison between LACV-L and influenza polymerase priming loops at elongation

a, LACV-L thumb, thumb ring, bridge and lid domains are displayed in turquoise, blue, purple and light blue respectively. The priming loop is shown in red. The template RNA is shown in yellow and the product in cyan. dsRNA top view (left) and side view (right) are shown. **b**, Equivalent elements are shown using the same color code in influenza virus polymerase (PDB: 6QCT). The priming loop extremity is disordered and is shown as a dotted line.

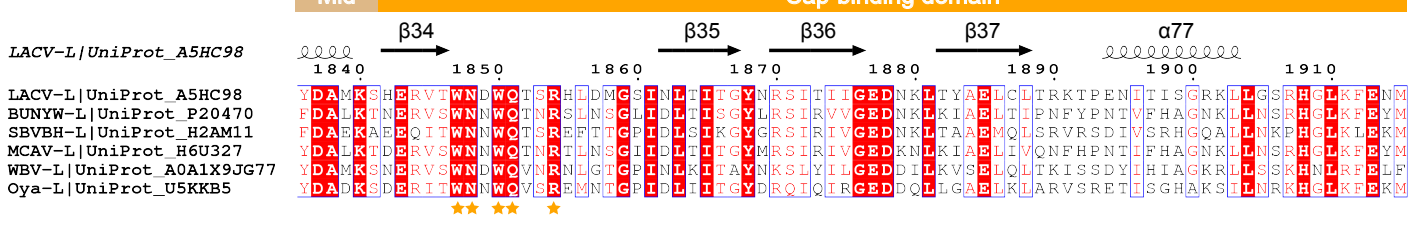
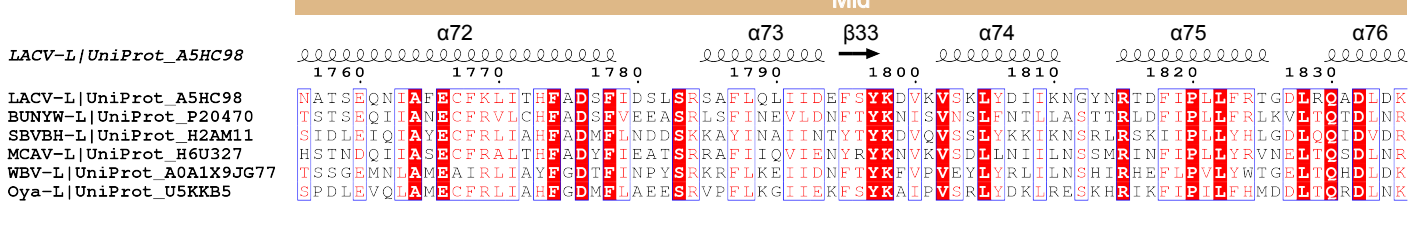
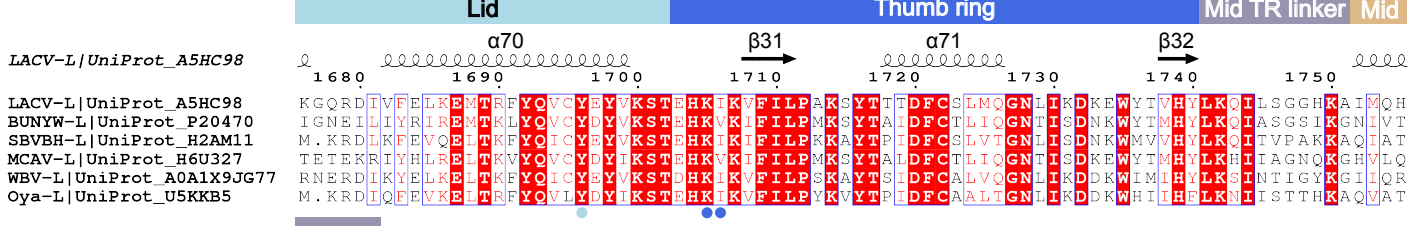
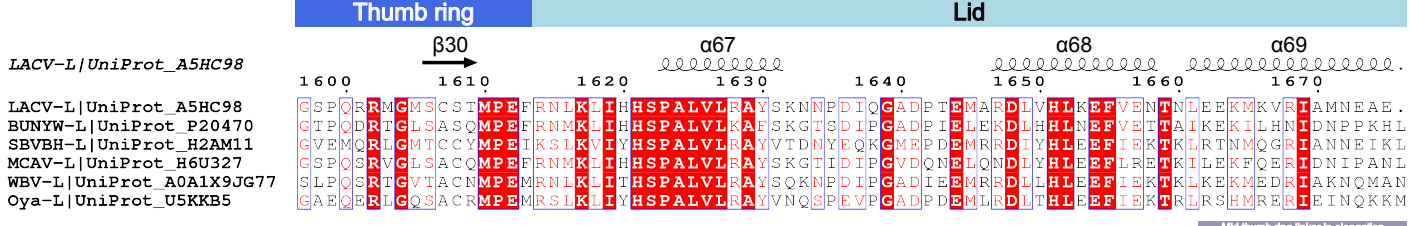
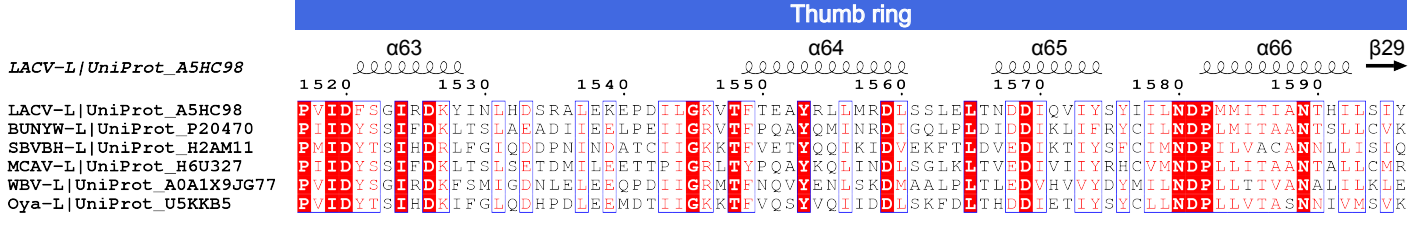
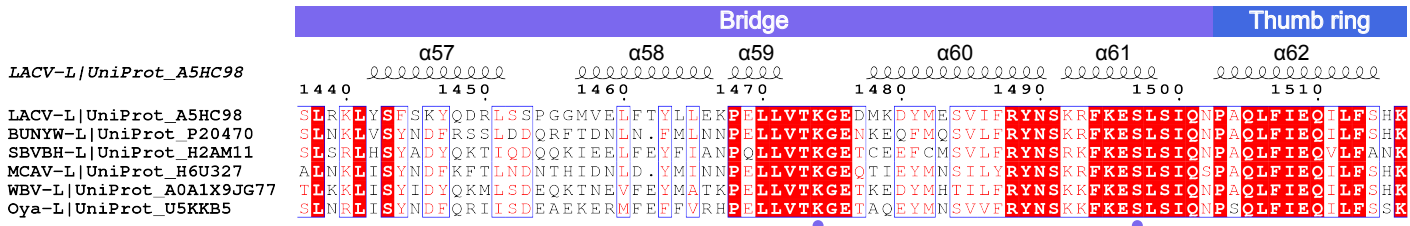
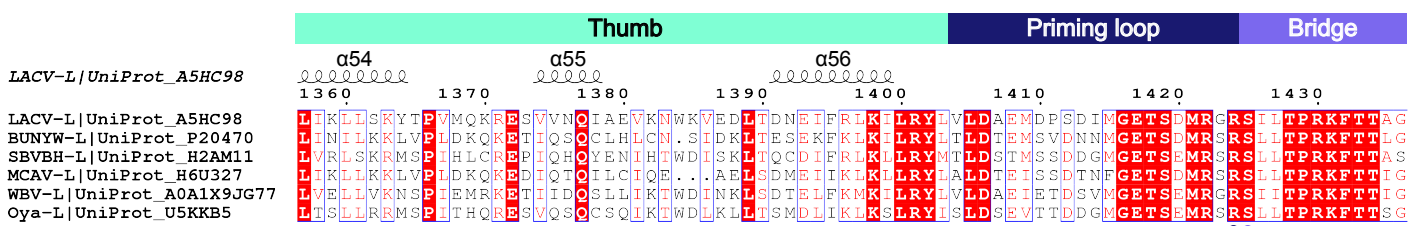
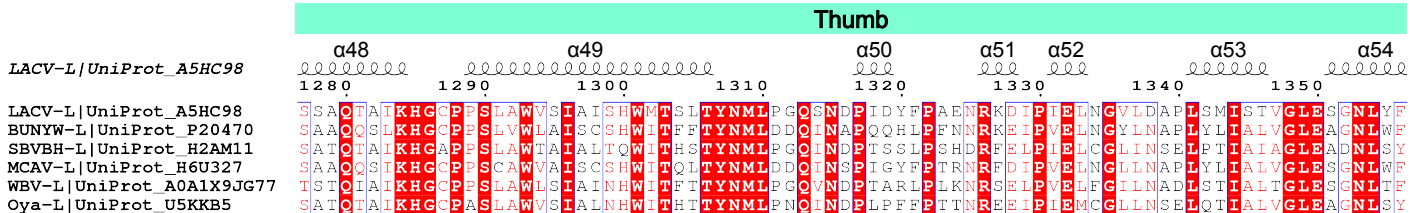
Supplementary Data (1/4)



Supplementary Data (2/4)



Supplementary Data (3/4)



Supplementary Data (4/4)



Supplementary data: Multiple alignment of *Peribunyaviridae* L proteins

Multiple alignment of six *Peribunyaviridae* L proteins: LACV, Bunyamwera virus (BUNYW), Schmallenberg (SBVBH), Macaue virus (MCAV), Wolkberg virus (WBV) and Oya virus. LACV-L secondary structures are shown and numbered. Domain positions and motifs are indicated. Endonuclease and CBD active site residues are labelled with green and gold stars respectively. Residues that coordinate the zinc are shown with a brown star. Residues that interact with the template/product RNA are labelled with a circle colored based on their domain localization.

2.3. Discussion

L'obtention de deux structures de la LACV-L, l'une dans un état de pré-initiation et l'autre dans un état mimant l'élongation, a permis de mettre en lumière les différents domaines localisés au sein de la partie C-terminale de la protéine L et de caractériser les différents changements de conformation induits lorsqu'un ARN double brin est présent au niveau de la cavité interne de la RdRp. Bien que ces résultats soient importants dans la compréhension générale de l'organisation structurale des protéines L chez les *Bunyavirales*, elles ne reflètent pas différents états actifs de la LACV-L et ce pour plusieurs raisons.

Premièrement, aucuns ions métalliques divalents ou NTPs n'ont été mis en présence de la LACV-L, qui ne peut donc synthétiser un nouveau brin d'ARN à partir de l'extrémité 3' du génome viral. De ce fait, l'internalisation des ARNv hybridés 5' 1-10 et 3' 1-16, bien que mimant un état d'élongation, n'est biologiquement pas pertinente.

Deuxièmement, la quasi-complémentarité des extrémités 5' et 3' de l'ARNv nécessite l'addition préalable de l'ARNv 5' 9-16 afin qu'il s'hybride au 3' 1-16 pour permettre aux extrémités du génome viral de se lier à leur site de fixation respectif. De ce fait, l'ajout de cet ARN empêche la translocation de l'extrémité 3' de l'ARNv de son site de fixation B vers son site de fixation A, localisé au niveau du site actif de la RdRp.

Troisièmement, des études *in vivo* réalisées en minireplicon par notre collaboratrice, le Dr. Florence Baudin (EMBL, Heidelberg), ont montré que l'insertion de différents tags aux extrémités N- et C-terminales de la LACV-L diminuaient drastiquement l'activité du virus, or les structures précédemment obtenues disposent d'un tag-poly-histidine localisé en N-terminal.

Plusieurs hypothèses ont donc dû être testées afin d'obtenir une activité suffisante pour permettre de caractériser simultanément (i) l'activité *in vitro* de la LACV-L et d'effectuer (ii) une caractérisation structurale par cryo-ME de différents états clés présents au cours des mécanismes de répllication et de transcription.

3. Caractérisation structurale et fonctionnelle de la LACV-L

3.1. Introduction

Dans le but d'effectuer une caractérisation fonctionnelle et structurale de la LACV-L et pallier ces problématiques soulevées dans les paragraphes précédents, deux solutions ont été établies pour obtenir une enzyme active *in vitro*.

Premièrement, afin de limiter l'hybridation entre les deux extrémités de l'ARNv, la séquence de l'extrémité 5' a été mutée en prenant soin de ne pas perturber sa fixation à la LACV-L. En effet, les dix premiers nucléotides du 5' forment une structure secondaire conservée chez les sNSVs, le *hook*, tandis que les nucléotides suivants forment un duplex distal à la surface de la protéine L avec l'extrémité 3' du génome (Pflug *et al.*, 2014; Reich *et al.*, 2014). Chez la LACV-L, les nucléotides du 5' A1, A4, G5, U6, G7 et U8 interagissent avec plusieurs résidus de la LACV-L de manière bases-spécifiques ce qui empêche leur mutation. En revanche, les nucléotides G2, U3, C10 et G9 interagissent à travers leur groupements phosphates, tandis que leurs bases respectives s'opposent les unes aux autres (Gerlach *et al.*, 2015; Arragain *et al.*, 2020). C'est pourquoi ces derniers ont été respectivement mutés en C2, G3, G10 et C9 limitant ainsi la complémentarité du 5' et du 3' mais en conservant l'appariement des bases du 5' non mutées. La structure secondaire en forme de *hook* est donc conservée.

Deuxièmement, le tag-poly-histidine localisé en N-terminal a été supprimé et relocalisé à l'intérieur de la séquence de la protéine. Basé sur les structures de la LACV-L précédemment obtenues (Gerlach *et al.*, 2015; Arragain *et al.*, 2020), un domaine appelé *California insertion* semblait remplir toutes les conditions nécessaires pour y placer ce tag. En effet, cette insertion, localisée au niveau du domaine *palm*, est constituée de deux hélices α reliées par une boucle flexible dirigée vers le solvant. Après insertion d'un strep-tag sur cette boucle, des études *in vivo* en minireplicon ont montré une récupération d'environ 85% de l'activité du virus.

Ces modifications ont permis de visualiser une activité de réplication et de transcription *in vitro*, activité que j'ai ensuite cherché à optimiser en jouant sur différents paramètres : les températures, les concentrations et les longueurs des ARN_v 5'/3', les NTPs ajoutés et les concentrations en ions divalents métalliques (magnésium ou manganèse).

Puis, grâce aux améliorations apportées au Glacios avec l'implémentation du logiciel d'automatisation des collectes SerialEM et l'intégration de la caméra K2 Summit précédemment installée sur le Polara, j'ai pu établir des conditions permettant de bloquer la LACV-L dans différents états actifs en cours de réplication et de transcription de l'ARN_v afin d'analyser par cryo-ME les différents changements de conformations que subit la protéine L.

Il en résulte l'obtention de sept structures à des résolutions comprises entre 2.7 et 4 Å. Quatre ont été obtenues sur le Glacios et trois sur le Titan Krios localisé à Heidelberg (projet Instruct), ce dernier étant équipé d'une caméra K3. Ces travaux, réalisés au cours de ma 3^{ème} et 4^{ème} année de thèse grâce à l'obtention d'une bourse d'extension de 9 mois financée par la Fondation pour la Recherche Médicale, font l'objet d'un article soumis dans lequel je suis premier auteur.

3.2. Article n°3 : Structural snapshots of La Crosse virus polymerase reveal the mechanisms underlying *Peribunyaviridae* replication and transcription

1 **Structural snapshots of La Crosse virus polymerase reveal the mechanisms**
2 **underlying *Peribunyaviridae* replication and transcription**

3
4 **Benoît Arragain¹, Quentin Durieux Trouilleton¹, Florence Baudin², Jan Provaznik³, Nayara**
5 **Azevedo³, Stephen Cusack^{4,*}, Guy Schoehn¹, H  l  ne Malet^{1,5*}**

6 ¹ Univ. Grenoble Alpes, CNRS, CEA, IBS, F-38000 Grenoble.

7 ²European Molecular Biology Laboratory (EMBL), Structural and Computational Biology Unit,
8 Heidelberg, Germany.

9 ³European Molecular Biology Laboratory (EMBL), GeneCore, Heidelberg, Germany.

10 ⁴European Molecular Biology Laboratory (EMBL), Grenoble, France.

11 ⁵Institut Universitaire de France (IUF).

12 *Correspondence: helene.malet@ibs.fr ; cusack@embl.fr

13
14 **ABSTRACT**

15 Segmented negative-strand RNA bunyaviruses encode a multi-functional polymerase that
16 performs genome replication and transcription. Here, we establish conditions for *in vitro*
17 activity of La Crosse virus polymerase and visualize by cryo-electron microscopy its
18 conformational dynamics, unveiling the precise molecular mechanics underlying its essential
19 activities. Replication initiation is coupled to distal duplex promoter formation,
20 endonuclease movement, prime-and-realign loop extension and closure of the polymerase
21 core that direct the template towards the active site. Transcription initiation depends on C-
22 terminal region closure and endonuclease movements that firstly prompt primer cleavage
23 and secondly promote primer entry in the active site. Product realignment after priming,
24 observed in replication and transcription, is triggered by the prime-and-realign loop. Switch
25 to elongation results in polymerase reorganization and core region opening to facilitate
26 template-product duplex formation in the active site cavity. The detailed uncovered
27 mechanics will be crucial for future design of antivirals counteracting bunyaviral life-
28 threatening pathogens.

29
30 **INTRODUCTION**

31 *Bunyavirales* is a large order of segmented negative sense single stranded RNA viruses
32 (sNSV). It comprises 12 families and more than 500 viruses, amongst which human
33 dangerous pathogens such as La Crosse (LACV, *Peribunyaviridae* family), Hantaan (HTNV,
34 *Hantaviridae* family), Rift Valley Fever, Severe Fever with Thrombocytopenia Syndrome
35 (RVFV and SFTSV, *Phenuiviridae* family), Lassa (LASV, *Arenaviridae* family) and Crimean
36 Congo Haemorrhagic fever viruses (CCHFV, *Nairoviridae* family)^{1,2}. There is currently no
37 treatment or vaccine to counteract them. In this context, we focus our interest on LACV and
38 more precisely on essential steps of its viral cycle: genome replication and transcription.
39 These processes are catalysed by the virally-encoded RNA-dependent RNA-polymerase
40 (RdRp), also called L protein (LACV-L)³. This large 260kDa multi-functional and monomeric
41 enzyme performs replication of the viral genome (vRNA) into a complementary RNA (cRNA),
42 which is then used as a template to generate nascent vRNA that have a size and composition
43 identical to the parental vRNA. Replication initiation is performed *de novo*, in the absence of
44 primer. LACV-L and other *Peribunyaviridae* are suspected to initiate their replication
45 internally at position 4 of the RNA template to produce a primer that then realigns to the
46 template end. This process, called “prime-and-realign”, is made possible by a triplet
47 nucleotide repetition at the 3'-vRNA template end (3'-UCAUCA...-5' for LACV) and has been
48 reported for several families in the *Bunyavirales* order, although with family-dependent
49 specificities⁴⁻⁶. Concerning transcription, LACV-L catalyses its initiation through a cap-
50 snatching mechanism, whereby LACV-L steals cellular capped mRNA, binds them through its
51 cap-binding domain (CBD), cleaves them 9 to 17 nucleotides downstream the cap with its
52 endonuclease domain (ENDO), before using them as a primer for transcription initiation^{7,8}.
53 The prime-and-realign mechanism has also been observed at this stage resulting in the
54 insertion of a triplet (5'-AGU-3') or multiple triplets (5'-AGUAGU...-3') complementary to the
55 3'-template inserted between the cap-snatched mRNA primer and the viral RNA
56 transcript^{7,9,10}.

57 Structural determination of these essential and complex sNSV polymerases has long been a
58 challenge in the field. Following determination of the X-ray structure of the C-terminally
59 truncated construct of LACV-L¹¹, structures of full-length LACV-L, LASV-L, Machupo-L (MACV-
60 L) and SFTSV-L have been determined by cryo-EM¹²⁻¹⁵. They contain a conserved central core
61 comprising the polymerase active site, an N-terminal ENDO and a flexible C-terminal region
62 (CTER) that includes the CBD. They were solved in their apo (SFTSV-L) or pre-initiation form

63 (LACV-L, Machupo-L and LASV-L) with the 3'-vRNA promoters bound in a secondary site at
64 the surface of the polymerases, away from the RNA synthesis active site. The 5'-vRNA
65 promoter, present only in LACV-L structures, is bound as a hook in a distinct and specific site
66 on the polymerase surface.

67 These structures were important advances in the field, but they do not shed light on the
68 detailed mechanisms by which this key multi-functional molecular machine performs
69 genome replication and transcription. In the present article, we uncover the conditions
70 necessary for LACV-L *in vitro* activity and determine LACV-L structures stalled at specific
71 stages of both replication and transcription revealing with unprecedented detail the critical
72 elements triggering activation of these essential processes.

73 **RESULTS**

74 **LACV-L replication activity**

75 Visualization of LACV-L activity necessitated careful optimization of (i) the construct, (ii) the
76 tag position and (iii) the composition of the vRNA promoters. Using a mini-replicon system,
77 LACV-L constructs with an N- or C-terminal tag were found to display low activity compared
78 to a construct without a tag (**Fig. 1a**). Addition of a Strep-tag in an internal exposed loop
79 called the California insertion (LACV-L_{Citag}) displayed 85% activity compared to the LACV-L
80 without a tag and was thus chosen for subsequent purification and *in vitro* activity assays
81 (**Fig. 1a**). To abolish unspecific RNA degradation *in vitro*, we introduced the H34K mutation
82 known to prevent ENDO activity¹⁶. Despite the optimized construct, the homogeneously
83 purified LACV-L_{Citag_H34K} (**Supplementary Fig. 1a**) incubated with wild-type (WT) 5'/3'-vRNA
84 promoters, NTPs, and MgCl₂ at 30°C for 4h did not generate replication products (**Fig. 1b,**
85 **lane 1**). This behaviour can be explained by the high complementarity of the 5'/3'-vRNA
86 promoters that tend to form stable double-stranded RNA, preventing their binding as single-
87 stranded RNA at their specific binding site on LACV-L. This situation is specific to *in vitro*
88 reconstitution as *in vivo* each promoter is kept in separate binding sites on the polymerase.
89 To avoid 5'/3'-vRNA duplex formation, we analyzed the structure of the 5'-vRNA promoter
90 and its interaction with the polymerase. We generated a 17-base-pair modified 5' RNA (5'-1-
91 17BPm), by mutating the nucleotides G2, U3, A9 and C10 of the 5'-end into C2, G3, C9 and
92 G10, thereby preserving the hook structure and its interaction with LACV-L while
93 significantly decreasing the 5'/3'-vRNA complementarity (**Fig. 1c, pre-initiation vs initiation**).
94 Equivalent activity assays with the 5'-1-17BPm in place of the WT 5'-vRNA led to the
95 formation of a 25-nucleotide replication product (**Fig. 1b, lane 2, Supplementary Fig. 1b**).
96 Next Generation RNA sequencing (NGS) confirms the presence of expected 25-mer
97 replication products (**Supplementary Fig. 2a**). It also identifies 25-mer products that contain
98 a misincorporated terminal 5' nucleotide and therefore start with GGU instead of AGU.
99 Slower-migrating products are also visible on gel and NGS identifies them as being products
100 extended in 5' with GU and GGU sequences. These correspond to products in which
101 realignment occurred twice. In addition, NGS identifies that some products are extended in
102 3' by 3 to 6 nucleotides with the sequence CUU to CUUGGU. Visualisation of replication
103 activity with LACV-L_{Citag_H34K}, 5'-1-17BPm and 3'-vRNA1-25 led us to optimize the assay
104 revealing a maximal activity with 2 to 5 mM Mg²⁺ while more abortive products and fewer

105 25-mer replication products are visualized in presence of Mn^{2+} (**Supplementary Fig. 1c**). A
106 time course indicates that the reaction is optimal at 4h (**Supplementary Fig. 1d**). We then
107 performed equivalent reaction assays omitting either the 5'- or the 3'-vRNA. These resulted
108 in the absence of 25-mer product formation, clearly indicating the requirement of both
109 promoter ends for replication activity (**Fig. 1b, lanes 3, 4**). The dependence of the product
110 size on the template size was analysed, and reveals that, as expected, a 3'-vRNA1-30
111 template gives rise to products 5 nucleotides longer than a 3'-vRNA1-25 template (**Fig. 1b,**
112 **lanes 2, 5**). Incubation of LACV-L_{Citag_H34K} with 5'-1-17Bpm, 3'-vRNA1-25, ATP, UTP and GTP
113 generates a 9-mer product, consistently with the template sequence that requires CTP
114 incorporation at position 10 (**Fig. 1b, lanes 6, 7**). Interestingly, CTP addition after 1h restores
115 complete product formation, indicating that the reaction performed with ATP, UTP and GTP
116 is stalled in a physiological, early-elongation state (**Fig. 1b, lanes 8, 9**). Higher molecular
117 weight bands appear in addition to the expected products (**Fig. 1b, lanes 2, 5, 8, 9**) and
118 likely correspond to the template/product hybridization upon assay arrest due to their high
119 complementary, as reported in several sNSV *in vitro* activity assays^{14,17}.

120 **Overview of the actively replicating LACV-L structures**

121 Visualisation of replication activity prompted us to collect three large cryo-EM datasets
122 which, coupled with advanced image processing involving extensive 3D classifications,
123 captured LACV-L structures in three key states of replication: initiation, early-elongation and
124 late-elongation at 2.8, 2.9 and 3.9 Å resolution, respectively (**Fig. 1c, d, Supplementary**
125 **Table 1, Supplementary Movie 1**). To stabilize the replication initiation state, LACV-L_{Citag_H34K}
126 was incubated with the 5'-1-17Bpm and the 3'-vRNA1-25 in presence of the initial
127 nucleotides to be incorporated. An initiation-mimicking state, with ATP present at position
128 +1 of the active site, was obtained when the LACV-L initiation complex was formed in
129 presence of ATP and UTP (**Fig. 1c, d, Supplementary Fig. 3**). Although UTP was added in the
130 mix, it is not observed in position -1, which appears to indicate that it is not incorporated at
131 initiation. For the replication early-elongation state, LACV-L_{Citag_H34K} was incubated with the
132 previously used RNAs supplemented with ATP, UTP and GTP resulting in the formation of a
133 9-base pair elongating template-product duplex in the polymerase internal cavity (**Fig. 1c, d,**
134 **Supplementary Fig. 4a**). For the late-elongation state, a 5'-1-17Bpm, 3'-vRNA-1-30 and four
135 nucleotides were used. In this state, the 3'-vRNA is entirely replicated, a 10-base pair
136 template-product duplex is visible in the active site and the 3'-vRNA end, following its exit

137 from the active site as a single stranded RNA, binds in the polymerase 3'-end secondary
138 binding site (**Fig. 1c, d, Supplementary Fig. 4b**).

139 Comparison of the three structures reveals major conformational changes of LACV-L during
140 initiation compared to its stable conformation found in the pre-initiation and the early-/late-
141 elongation states (**Fig. 1d**). A major movement of the ENDO is visible at initiation,
142 coordinated with the movement of the entire CTER, including the CBD, that becomes flexible
143 at this stage. Whereas the ENDO, mid and zinc-binding domain (ZBD) gain back their pre-
144 initiation position at elongation, the CBD remains flexible.

145 **The replication initiation state**

146 The principal novelty in the obtained structures concerns the initiation conformation that
147 provides molecular insights into the mechanics underlying this critical step. The distal
148 duplex, formed between nucleotides 12 to 17 of the 3'/5'-vRNA ends, plays a critical role in
149 triggering the template positioning in the RNA tunnel entrance (**Fig. 2a, Supplementary Fig.**
150 **5a**). The distal duplex is held in place by the 5'-vRNA hook (nucleotides 1-10), the vRNA
151 Binding Lobe domain (vRBL), the clamp, the arch and the α -ribbon. The RNA entrance tunnel
152 is surrounded by the vRBL, the fingers, the thumb ring and the bridge domains. Nucleotides
153 G9 to A11 of the template are stabilized in the tunnel by multiple hydrophobic and polar
154 interactions while nucleotides A6, C7 and A8 are less coordinated and adopt alternative
155 conformations (**Supplementary Fig. 5a**). Their visualized mobility might provide the template
156 with the necessary flexibility for its realignment after priming (**Fig. 2b**). The 3' terminal
157 nucleotides U1 to C5 display clearer density correlated with their higher degree of
158 stabilization by the finger domain and in particular the canonical polymerase motif F (**Fig.**
159 **2c, Supplementary Fig. 5a**). The 3' terminus itself interacts by an unexpected but crucial
160 element: the loop 982-995, that is localized between the fingers and the palm domain, and
161 which, unexpectedly, plays the role of a replication priming loop (**Fig. 2c**). Its dual role in
162 priming and realignment (see below) leads us to name this element the prime-and-realign
163 loop (PR loop).

164 At initiation, the ENDO conformational change brings its residues 172 to 184 (α -helix 7)
165 proximally to the PR loop (**Fig. 2c**). The ENDO residues E177 and K181 would clash with the
166 PR loop in the absence of its rearrangement, suggesting that the ENDO movement is linked
167 with the PR loop extension. In this extended state, the PR loop residue M989 interacts with
168 the 3' terminal nucleotide U1 in position -3 of the active site while the residues 1990 and

169 S991 stabilize the nucleotide C2 in position -2 of the active site (**Fig. 2c**). As the result, the
170 template nucleotides A3 and U4 are localized in position -1 and +1 of the active site (**Figs.**
171 **1c, 2c**). The map displays a clear density for an ATP in position +1 that makes hydrogen
172 bonds with U4 and stacks with A3 (**Supplementary Fig. 5a**). Its ribose interacts with W1064
173 (from motif A) and Q1145 (from motif B), its phosphates interact with D1060, M1061, K1063
174 (from motif A), D1187 (from motif C) and two magnesium ions, themselves located in
175 canonical positions for catalytic magnesium A and B (**Fig. 2c**).

176 The identification of the unexpected PR loop led us to analyze further its role in replication
177 by engineering single-alanine substitutions of its tip residues M989, I990 and S991 (**Fig. 2e**).
178 M989A diminishes the formation of 25-mer product by 95% and of GU/GGU 5' extended
179 replication products by 81%, confirming the importance of this residue in precise template
180 positioning at initiation (**Fig. 2e, lane 2**). I990A mutation does not significantly modify 25-
181 mer product formation but multiplies by 3.88 the formation of GU/GGU 5' extended
182 replication products. (**Fig. 2e, lane 3 vs lane 1**). LACV-L_{Citag_H34K_S991A} produces 10.7 times
183 more GU/GGU 5' extended replication products than LACV-L_{Citag_H34K}. Altogether, these
184 mutations clearly confirm the importance of the PR loop tip in the replication mechanism
185 with a defined role of M989 at initiation and an unexpected role of the mutations I990A and
186 S991A in the formation of GU/GGU 5' extended replication products.

187 **Switch to elongation and related conformational changes**

188 Progression towards elongation implies important remodeling of LACV-L domains. Retraction
189 of the PR loop, that is coupled with ENDO repositioning (**Fig. 2d, 3a**), provide space for the
190 initial formation of the template-product duplex in the active site cavity. The absence of CTP
191 in the mix results in LACV-L stalling with template G10 in position +1 of the active site in a
192 post-translocation pre-incorporation state (**Fig. 2d, Supplementary Fig. 5b**). Whereas the
193 replication initiation was performed internally, realignment must have occurred, as the
194 replication elongation displays an entire product with the 5'-cRNA end that is
195 complementary to the 3'-vRNA nucleotide 1. To accommodate the internal 9-base pair
196 duplex, switch from initiation to elongation state results in the opening of the lid, the thumb
197 and the thumb-ring coupled with the extension of the bridge loop 1423-1441 by unwinding
198 of the helix 1435-1438 (**Fig. 3b**). Together these movements switch the core from a closed
199 to an open state, resulting in the extrusion of the previously defined "priming loop" that
200 induces the opening of the template exit tunnel (**Fig. 3c**). As the "priming loop" remains

201 away from the active site during replication and does not appear to be involved in priming in
202 the visualized states, it is renamed “template exit plug” (**Fig. 3b, c**). Interestingly, reverse
203 movement from an open to a closed core is observed between pre-initiation and initiation
204 (**Fig. 3b, c**) suggesting that the core closure is a rather transient state acquired at initiation
205 only, while the open core is a more stable long-standing state.

206 In addition to the core opening/closing between the replication states, global movements of
207 the vRBL, clamp, arch and α -ribbon are visualised (**Fig. 3b, c**). At pre-initiation, the α -ribbon
208 is ordered and the vRBL is proximal to the core, thereby forming a cleft called the 3'-vRNA
209 secondary binding site to which the 3'-vRNA end binds specifically. At initiation, the acquired
210 α -ribbon flexibility and displacement of the vRBL away from the core disrupt the 3'-vRNA
211 secondary binding site. The 3'-vRNA changes its position to orient itself into the template
212 entry tunnel necessary for initiation. At late-elongation stage, the exit of the 3' template
213 from the template exit channel is coordinated with a repositioning of the vRBL and the
214 clamp close to the core resulting in the formation of the 3' secondary binding site as in pre-
215 initiation. The replicated 3'-vRNA end exits as a single stranded RNA from the active site
216 cavity, binds to a positively charged groove (**Supplementary Fig. 5c, 6**), before reaching the
217 3' secondary binding site, ready for the next replication cycle.

218

219 **LACV-L transcription activity**

220 Transcription activity was analysed using LACV-L_{Citag_H34K} incubated with 14-mer capped RNA
221 primers, the 3'-vRNA1-25/1-30 and the 5'-1-17BPm. The chosen capped RNA primers have a
222 two- (cap14AG) or three-nucleotide (cap14AGU) complementarity with the 3'-vRNA end (5'-
223 ...ACUACU-3') to facilitate priming. Time, divalent ion identity and divalent ion concentration
224 were tested and indicate that the transcription reaction is optimal after 30 min in the
225 presence of 2 mM MgCl₂ (**Supplementary Fig. 7a, b**). Using these conditions, the reactions
226 with cap14AG/cap14AGU lead to the formation of two main transcription products, a 37-
227 /36-mer and a 40-/39-mer (**Fig. 4a, b, lanes 1, 2**). The minority 37-/36-mer products
228 correspond to the capped RNA size elongated by 23/22 nucleotides, considering the
229 hybridization of the last 2 or 3 template nucleotides with the cap14AG/capAGU. Their
230 migration is consistent with a capped molecular weight marker of the same size and
231 composition, and they are identified in NGS (**Fig. 4a, b**). The majority 40-/39-mer products
232 correspond to primed and subsequently realigned transcripts, that result in the addition of

233 an AGU nucleotide triplet at the end of the capped primer before elongation, as detected by
234 NGS analysis (**Supplementary Fig. 2b**). NGS also reveals the presence of products that have
235 been realigned up to three times. This is in accordance with the sequencing of capped RNA
236 produced by *Peribunyaviridae* polymerase *in vivo*^{7,10,18}. As expected, the size of the
237 transcribed RNA depends on the length of the template: replacing the 3'-vRNA1-25 by a 3'-
238 vRNA1-30 gives rise to products that are 5 nucleotides longer (**Fig. 4a, b, lanes 3, 4**). Usage
239 of an NTP subset containing ATP, UTP and GTP, to obtain an early-stalled transcription
240 complex, shows formation of the expected 21/24-nucleotide products with cap14AG and
241 20/23 nucleotide products with cap14AGU (**Fig. 4a, b, lanes 5-8, Supplementary Fig. 7c**).
242 Interestingly, whereas the realignment was almost complete in presence of the 4 NTP, it is
243 only partial in presence of the 3-NTP subset. Subsequent incubation with a CTP restores
244 formation of full-length capped RNA product, but with a majority of non-realigned capped
245 RNA products (**Fig. 4a, b, lanes 9-12**).

246 **Overview of the transcriptionally active LACV-L structures**

247 To visualize the path taken by the RNA during transcription initiation, a large cryo-EM
248 dataset was collected on a transcription reaction mix containing LACV-L_{Citag_H34K}, the
249 cap14AG, the 3'-vRNA1-25, the 5'-1-17BPm, UTP, ATP and MgCl₂. Advanced image
250 processing enabled us to separate and reconstruct several active snapshots (**Supplementary**
251 **Figs. 3 and 8, Supplementary Table 2, Supplementary Movie 2**). A conformation captured at
252 3.9 Å resolution shows the capped RNA bound to the CBD, 45 Å away from the ENDO, a
253 distance compatible with capped RNA primer cleavage after 9 to 17 nucleotides (**Fig. 4c, d**).
254 We called this snapshot the “putative endonuclease cleavage conformation”, although this
255 remains speculative. Only the first two nucleotides of the capped RNA are visible, likely due
256 to the low affinity of the ENDO for RNA, that is related to the diversity of sequences to be
257 cut (**Supplementary Fig. 9a**). The 5'/3' promoters are properly positioned for transcription
258 initiation (**Supplementary Fig. 9a**). The second conformation, obtained at 3.1 Å resolution,
259 shows the cap and the first two nucleotides of the capped RNA protruding towards the
260 active site (**Supplementary Fig. 9b**). It likely corresponds to the conformation just prior
261 transcription initiation and was therefore called “capped primer active site entry”. A
262 “transcription initiation conformation” structure at 3.6 Å resolution is also trapped for a
263 small-particle subset (**Fig. 4c, d**). The capped RNA is visible in the CBD and in the active site
264 with flexible nucleotides in between. It shows incorporation of UTP and ATP in the capped

265 RNA product (**Supplementary Fig. 9c**). Finally, another cryo-EM data collection and image
266 processing led to the determination of a transcription “early-elongation” conformation
267 structure at 3.3 Å resolution. It was obtained by incubating the transcription complex for 30
268 min at 30°C in presence of ATP, UTP, GTP and MgCl₂. It shows an elongated capped RNA that
269 forms a 9-base pair template-product duplex in the active site cavity (**Fig. 4c, d**,
270 **Supplementary Fig. 9d**).

271 **Major functionally relevant conformational changes of LACV-L are visualized between the** 272 **transcription states**

273 The switch from the pre-initiation to the putative endonuclease cleavage conformation is
274 coupled with a large ENDO movement resulting in a 160° rotation that liberates the
275 previously blocked capped RNA tunnel entrance (**Fig. 4e**). The acquired ENDO position is
276 reminiscent to the orientation visualized in the LACV-L ΔCTER X-ray and cryo-EM
277 structures¹¹. In addition to the ENDO movement, the entire CTER rotates by 20° (**Fig. 5a**),
278 bringing the ZBD and the mid domain towards the core, where they interact with the thumb
279 ring (**Fig. 5b**). This conformation of the CTER brings the CBD binding site close to the capped
280 RNA entrance tunnel, contrasting with its exposed position taken at pre-initiation (**Fig. 4e**).
281 The CBD, and in particular the loop 1850-1859 that is involved in cap binding and was
282 disordered at pre-initiation, is stabilized by several domains and becomes visible, explaining
283 the proper capped RNA binding site formation (**Fig. 5b**). The CBD interacts with the palm
284 domain, the mid, the ZBD β-hairpin and the core lobe.

285 Switch from the endonuclease cleavage conformation to the transcription initiation
286 conformation involves a 175° rotation of the ENDO (**Fig. 4d, e**). Interestingly, ENDO charge
287 analysis suggests that, whereas the negatively charged capped RNA might be attracted by
288 the ENDO positively charged catalytic site in the endonuclease cleavage conformation, it
289 would be repelled by the negatively charged ENDO surface in the transcription initiation
290 conformation, thereby facilitating capped RNA entry towards the polymerase active site
291 (**Supplementary Fig. 10**). In addition, to the large ENDO and CTER movements, switch from
292 pre-initiation to endonuclease cleavage conformation/transcription initiation is coupled with
293 core closure, in an analogous manner to what is observed at replication initiation.

294 Nearly all the conformational changes that resulted in transcription initiation are reversed at
295 elongation, bringing back the core in an open conformation, the ENDO and the CTER in the

296 pre-initiation conformation (**Fig. 4d, e**). The only exception concerns the CBD that acquires
297 an orientation differing of 30° compared to pre-initiation.

298

299 **Zoom on contacts between LACV-L and RNA during transcription**

300 The cleavage conformation reveals the peculiar binding of the cap that is stacked between
301 W1847 and R1854 of the CBD, with Q1851 ensuring guanine specificity (**Fig. 6a**). The cap
302 triphosphate moiety makes hydrogen bond interactions with K2011, K2012, R1854 and
303 W1850. The two next nucleotides interact through their bases with Y714 of the core lobe
304 domain, H2014, F2015, K2017 of the mid domain and Y1716 of the thumb ring domain,
305 while the phosphate is stabilized by K2017 (**Fig. 6a**). The cleavage conformation also reveals
306 that the PR loop tip stabilizes the template 3' terminus prior to capped entry into the active
307 site (**Fig. 6a**).

308 Following template binding and capped RNA cleavage, the capped RNA is directed towards
309 the active site (**Fig. 6b**). The cap and the first 2 nucleotides of the primer remain bound in a
310 similar fashion as in the cleavage conformation. The next nucleotides more sparsely bind
311 through their phosphates to charged residues of the finger domain (R820, R824, K983), the
312 thumb-ring (R1614) and the lid (H1620, H1621, D1641) (**Fig. 6b**). Because these residues
313 bind to the phosphate backbone, it is likely that binding is sequence independent and can
314 probably adapt to various sizes of snatched capped RNA primer. The RNA density is more
315 defined towards the active site, indicating tighter binding (**Supplementary Fig. 9c**). Capped
316 RNA bases 13A and 14G that are complementary to 1U, 2A of the 3'-vRNA end form a duplex
317 in position -3 and -2 of the active site (**Fig. 6b**). The UTP and ATP nucleotides added to the
318 reactions were incorporated into the nascent mRNA and are positioned in position -1 and +1
319 of the active site in front of 3'-vRNA A3 and U4.

320 Following the transcription initiation state visualized in the structure, a realignment must
321 occur, as revealed by *in vitro* transcription products that are 3 nucleotides longer than if the
322 capped RNA was simply elongated (**Fig. 4a, b**). We thus investigated the structure at
323 initiation to identify which part of the polymerase might be implicated in realignment.
324 Interestingly, the PR loop is in a retracted position and is proximal to the template position -
325 4. PR loop induced template realignment from position -4 to position -1 would perfectly
326 reposition the template for subsequent elongation (**Fig. 6b**). We thus investigated if
327 mutation of the PR loop tip might affect realignment. While transcription assays performed

328 with LACV-L_{H34KM989A} and LACV-L_{H34KS991A} are equivalent to LACV-L_{H34K}, LACV-L_{H34KI990A} displays
329 a different transcription profile (**Fig. 6c**). NGS analysis reveals that LACV-L_{H34KI990A} generates,
330 in addition to the expected 37- and 40-mer products, a significant number of aberrant
331 products. These latter mainly correspond to the capped primer elongated by repetitive of
332 AGU triplets that are likely added by realignment. Altogether these results confirm the role
333 of the PR loop, and in particular of the residue I990, in realignment during transcription
334 activity.

335 Transcription then proceeds to elongation. While the template behaves like in replication,
336 the capped RNA product exits through a tunnel that is different from the one used for its
337 entry, surrounded by the bridge (K1474), the fingers (R820, N823, R824, R829), the lid
338 (S1622, Y1696) and the thumb ring (H1704) (**Fig. 6d**). The capped RNA reaches the ENDO
339 with which it interacts through residues (R33, K94, F162, H184). The 5'-cap detaches itself
340 from the CBD, permitting the exit of capped RNA through a cleft between the ENDO and the
341 CBD.

342

343 **DISCUSSION**

344 Replication initiation using a prime-and-realign mechanism has been observed in activity
345 assays for decades for several bunyaviruses⁴⁻⁶ and is also used by Influenza for vRNA
346 synthesis from cRNA¹⁹. However it had not yet been structurally characterized in any of
347 these viruses due to the flexibility of the 3'-cRNA end²⁰. Here, we identify the elements
348 stabilizing the 3'-vRNA up to the active site: (i) the distal duplex, corroborating its
349 importance in replication assays²¹⁻²³, (ii) the newly identified PR loop.

350 The PR loop is present in its retracted position in the other structurally determined sNSV
351 polymerases (**Supplementary Fig. 11a**). It remains to be shown whether its extension at
352 initiation also occurs in other sNSV. In terms of sequence, it is not very conserved within
353 Bunyaviruses, except for Hantaviruses L proteins which contains at the loop tip a tyrosine, an
354 isoleucine and a serine in place of M989, I990, S991 in LACV-L (**Supplementary Fig. 11**). As
355 M989 is likely to stabilize the incoming base in position -1, a tyrosine may play an equivalent
356 role in hantaviruses. The exact mechanism of initiation may however slightly differ between
357 LACV-L and HTNV-L. Indeed, the 5' end of HTNV RNA products have been shown to be
358 monophosphate^{4,24} whereas we show here that LACV-L products have triphosphate ends
359 (**Supplementary Fig. 1e**). The model proposed for HTNV-L, that would need to be validated

360 experimentally, is that an extra nucleotide in 5' is added in the prime-and-realign process,
361 that would then be cleaved, leaving a monophosphate at the 5' end. Hantavirus PR loop
362 might thus play an additional role at replication initiation. When comparing with Influenza
363 orthomyxovirus, the PR loop sequence is not conserved, but PB1 residue V273, that
364 corresponds to the PR loop tip, is essential for prime-and-realign both *in vitro* and *in vivo*²⁵
365 (**Supplementary Fig. 11**). In Influenza, the PR loop may somehow complement the role of
366 the priming loop in replication initiation^{23,26,27}. Altogether, these elements strongly suggest
367 the importance of the PR loop for prime-and-realign initiation, with family-related
368 specificities.

369 In LACV-L, the PR loop extension at replication initiation is coupled with a conformational
370 change of the ENDO (**Fig. 3a and Supplementary Fig. 12a**). In the case of influenza
371 polymerase, although template realignment occurs following internal initiation during cRNA
372 to vRNA replication, the situation seems to be quite different. It is thought that two distinct
373 influenza polymerase dimers, between an apo-polymerase and the replicating polymerase,
374 are involved in this process, whereas for LACV-L, no dimers have yet been observed. One
375 influenza dimer has been suggested to facilitate template realignment by modifying the
376 position of influenza priming loop²⁰ (**Supplementary Fig. 12b, c**). The second replicase dimer
377 depends essentially on the host factor ANP32, which mediates formation of dimer between
378 a 'packaging' polymerase and the replicating polymerase²⁸. Influenza polymerases in both
379 these dimers are in the 'replicase' rather than 'transcriptase' conformation, which involves a
380 re-positioning of the ENDO and re-organisation of the PB2-CTER that brings the PB2-cap and
381 the residues 490-493 of the Cap-627 linker close to the putative PR loop (**Supplementary Fig.**
382 **12d**). It remains to be seen if the residues 490-493 of the Cap-627 linker and the CBD may,
383 when acquiring this position, trigger a rearrangement of the PR loop that could complement
384 the role of the priming loop in the prime-and-realign mechanism.

385 Proper transcription initiation, on top of the 3' template proper positioning, depends on
386 capped RNA binding, cleavage and positioning in the active site. In LACV-L, major
387 conformational changes of the ENDO and the CTER are involved in this process. These
388 movements are more pronounced than for influenza polymerase, in which the ENDO
389 remains in the same position between the cleavage and the transcription initiation
390 conformations, the only major movement being done by the PB2-cap that rotates of 70° to
391 bring the capped RNA first to the ENDO active site and then to the polymerase active

392 site^{23,29}. The CBD and ENDO of LACV and influenza polymerases are in different positions
393 compared to the core precluding their exact comparison but the distance separating the cap
394 and the ENDO cleavage site are of 45 Å for LACV and 49/52 Å for Influenza, in accordance
395 with the similar sizes of capped primers (**Supplementary Fig. 13**). Although the CBD
396 positions differ, the capped RNA tunnel entrance is conserved, together with core closing
397 upon initiation. However, the details of transcription initiation differ in the two polymerase
398 due to the exact positioning of the 3'-template in the active site at initiation, with the first
399 nucleotide of the vRNA template being in position -2 and -3 for Influenza and LACV
400 respectively²⁶ (**Supplementary Fig. 13**).

401 Combination of *in vitro* activity assays and structural snapshots of LACV-L in different active
402 conformations lead us to propose a structural model for the prime-and-realign mechanism
403 (**Fig. 7a**). Concerning replication, we hypothesize that proper closing of the active site motifs
404 coupled with the visualized flexibility of the nucleotides 6, 7, 8 of the 3' template enable a
405 transient translocation of both the template and ATP of 1 nucleotide, thereby positioning
406 the ATP in position -1 of the active site, where it would interact with the residue M989 of the
407 PR loop tip (**Fig. 7b**). This would enable positioning of a GTP in position 1 of the active site
408 and formation of a nascent pppApG product. Normal elongation would generate a
409 pppApGpU product concomitantly with translocation of the template with its terminus
410 located in position -5 of the active site. At this stage, the template would be in extreme
411 tension due to (i) the presence of the distal duplex linked to the 5'-hook at the tunnel
412 entrance, (ii) the maximum stretching of template nucleotides 6-8 and (iii) the template
413 positioning constrains by the PR loop. These would act as a spring and break the hydrogen
414 bonds between pppApGpU and the template nucleotide 4 to 6, repositioning the primer in
415 front of template nucleotide 1 to 3, completing the realignment. Further elongation results
416 in proper product formation, exactly complementary to the template.

417 Concerning transcription, it is expected that the template and capped RNA positioning are
418 more permissive than in replication, to adapt to the capped RNA length and composition
419 (**Fig. 7c**). Capped RNA terminating by AG would position in front of the 3' template end in
420 position -2 and -1 of the active site, ready for incorporation of a U in position 1. This would
421 require the retraction by one nucleotide of the template, made possible by the flexibility of
422 its nucleotides 6-8 and capped RNA hybridization. This would permit incorporation of UTP in
423 position +1 of the active site. Subsequent translocation would bring the polymerase in the

424 same state as if a capped RNA terminating by AGU is used as a primer. The 3' template
425 terminus would be in position -3 to -1 of the active site, ready for incorporation of an A in
426 position 1. Following incorporation of the subsequent UTP and GTP, realignment may occur
427 once or several times, accordingly to *in vivo* mRNA composition observation. This may be
428 facilitated by the PR loop, which, by interfering with proper hydrogen bonding between the
429 template end and the product, may reduce the number of hydrogen bonds to be broken
430 upon realignment. Although speculative, this is supported by the absence of interaction of
431 the extreme 3' A nucleotide of the template with its complementary U of the capped RNA in
432 position -3 of the active site (**Fig. 6b**). As for replication, we propose that (i) the
433 destabilization of the distal duplex, (ii) the extreme tension of nucleotides 6-8, (iii) and the
434 PR constrains may provide enough energy to trigger template repositioning. Following
435 realignment(s) (if realignment occurs) the PR loop would retract, thereby switching the
436 polymerase to an elongation mode. The proposed models provide hints on how replication
437 and transcription may occur in other families of the *Bunyavirales* order^{4-6,30,31} and
438 complements the existing knowledge on prime-and-realign mechanisms used by Influenza
439 for the replication of vRNA from a cRNA template^{19,25}, and for transcription³².
440 In conclusion, the combination of *in vitro* assays and cryo-EM structural snapshots of specific
441 LACV-L active states unveils the precise mechanisms of initiation by prime-and-realign and
442 elongation of both replication and transcription. These results are novel for the entire
443 *Bunyavirales* order and shed light on processes conserved in all sNSV. They will be key to
444 design structure-based drugs to counteract these life-threatening viruses that would target
445 essential elements of replication and transcription activity or prevent essential
446 conformational changes.

447

448 **METHODS**

449 **LACV minigenome assay**

450 Sub-confluent monolayers of HEK-293 cells seeded in 6-well dishes were transfected with
451 250 ng each of plasmid pHH21-LACV-vM_{Ren}, pTM-LACV_L (or insertion mutants), pTM-
452 LACV_N, 500 ng pCAGGS-T7, and 100 ng pTM1-FF-Luc (described in^{33,34} using Nanofectin
453 transfection reagent in 200 μ l serum-free medium (OptiMEM, Gibco-BRL). In addition, 250
454 ng of pl.18-HA-PKR was added to the plasmid mix³³. At 72 h post-transfection, cells were
455 lysed in 200 μ l Dual Luciferase Passive Lysis Buffer (Kit Promega) and both Luciferase (FF-

456 Luc) and Renilla luciferase (REN-Luc) activities were measured as described by the
457 manufacturer (Promega).

458 As pTM-LACV_L, we used N-terminal (Nter) His-tag LACV-L full-length construct (strain
459 LACV/mosquito/1978, GenBank: EF485038.1, UniProt: A5HC98) previously cloned into a
460 pFastBac1 vector between NdeI and NotI restriction sites (Arragain et al, 2020), Cterminal
461 (Cter) His-tag LACV-L full length construct. An insertion mutant of LACV-L full length
462 construct (insertions of 2 internal Strep-Tag II “WSHPQFEK” flanked by a SG/GSG linkers) was
463 as well generated between residues G1034-Y1035.

464

465 **Cloning, expression and purification**

466 The LACV-L full length construct with an addition of one internal Strep-Tag II “WSHPQFEK”
467 flanked by a SG/GSG linker between residues G1034 and Y1035 (G1034-SG-StrepTag II-GSG-
468 Y1035) was chosen and recloned. An additional H34K mutation was inserted leading to the
469 LACV-L_{CItag_H34K} construct used for cryo-EM analysis and activity assays. Mutations
470 M989A/I990A/S991A (LACV-L_{CItag_H34K_M989A}/LACV-L_{CItag_H34K_I990A}/LACV-L_{CItag_H34K_S991A}) were
471 inserted on LACV-L_{CItag_H34K} construct and used for activity assays testing the role of the PR
472 loop. All LACV-L clones were made using combination of Polymerase Chain Reactions (PCR),
473 agarose gel purification, DNA extraction using PCR clean-up kit (Machery-Nagel), Gibson
474 assembly (NEB) and finally sequenced (Genewiz) before proceeding to further experiments.

475 For all LACV-L_{CItag_H34K} constructs, expressing baculoviruses were made using the Bac-to-Bac
476 method (Invitrogen)³⁵. For protein expression, High 5 (Hi5) cells at 0.5×10^6 cells/mL
477 concentration were infected by adding 0.1% of virus and collected 72 h to 96 h after the day
478 of proliferation arrest.

479 Hi5 cells were resuspended in lysis buffer (50 mM Tris-HCl pH 8, 500 mM NaCl, 2 mM β -
480 mercaptoethanol (BME), 5% glycerol) with cOmplete EDTA-free protease inhibitor cocktail
481 (Roche) and disrupted by sonication for 3 min (10 s ON, 20 s OFF, 50% amplitude) on ice.
482 Lysate was clarified by centrifugation at 48,000 g during 45 min at 4 °C and filtered. Soluble
483 fraction was loaded on pre-equilibrated StrepTrap HP column (Sigma-Aldrich) and eluted
484 using initial lysis buffer supplemented by 2 mM d-Desthiobiotin (Sigma-Aldrich). LACV-
485 L_{CItag_H34K} fractions were subsequently pooled, dialyzed 1h at 4 °C in heparin buffer (50 mM
486 Tris-HCl pH 8, 250 mM NaCl, 2 mM BME, 5% glycerol) and loaded on HiTrap Heparin HP
487 column (Sigma-Aldrich). Elution was performed using 50 mM Tris-HCl pH 8, 1 M NaCl, 5 mM

488 BME, 5% glycerol. LACV-L_{Citag_H34K} protein was finally loaded on a pre-equilibrated Superdex
489 200 Increase 10/300 GL (Sigma-Aldrich) size exclusion chromatography column in 50 mM
490 Tris-HCl pH 8, 400 mM NaCl, 5 mM BME. Best fractions were pulled, concentrated using
491 Amicon Ultra 10 kDa (Merck Millipore), flash frozen in liquid nitrogen and conserved at -80°C
492 for future experiments.

493

494 ***In vitro* transcription, capping and RNase T1 cleavage**

495 *In vitro* transcription was realized using a DNA oligo enclosing the T7 promoter sequence
496 with an additional G at the 3' end (5'-TAA TAC GAC TCA CTA TAG G-3') and a template DNA
497 oligo enclosing both complementary (i) T7 promoter sequence and (ii) 3'-5' DNA sequence of
498 the desired 5'-3' RNA. The two nucleotides in 5' of the primers were 2'-O methylated in
499 order to reduce nontemplated nucleotide addition at the 3'terminus of RNAs transcribed
500 with T7 RNA polymerase.

501 Different templates were used to produce a 37-mer RNA (3'-ATT ATG CTG AGT GAT ATC CTA
502 CGA TAT TAT CAT CAC ATG ATG GTT CAT ATC TAT-5' giving 5'-GGA UGC UAU AAU AGU AGU
503 GUA CUA CCA AGU AUA GAU A-3'), a 21-mer RNA (3'-ATT ATG CTG AGT GAT ATC CTA CGA
504 TAT TAT CAT CAC AU-5' giving 5'-GGA UGC UAU AAU AGU AGU GUA-3'), a 14-mer RNA (3'-
505 ATT ATG CTG AGT GAT ATC CTA CGA TAT ATC A-5' giving 5'-GGA UGC UAU AUA GU-3').

506 T7 promoter (15 μM) and template DNA (10 μM) oligos were mixed in transcription buffer
507 (30 mM TRIS-HCl pH 8, 20 mM MgCl₂, 0.01% Triton X-100, 2 mM Spermidine), annealed at
508 80°C for 2 min and cooled down to room temperature (RT). Hybridized DNA oligos (1 μM)
509 were then added to transcription buffer supplemented by 5 mM ATP, UTP, GTP and CTP
510 (Sigma-Aldrich), 10 mM dithiothreitol (DTT), 1% PEG8000, 50 μg/mL T7 RNA polymerase and
511 incubated overnight at 37°C. *In vitro* transcription reactions were stopped by adding 2X RNA
512 loading dye (95% formamide, 1 mM EDTA, 0.025% SDS, 0.025% bromophenol blue, 0.01%
513 xylene cyanol), heated 5 min at 95°C and loaded on a 20% Tris-Borate-EDTA-7M urea-
514 polyacrylamide gel. The corresponding products bands were cut, soaked in 0.3 M NaOAc pH
515 5.0-5.2 and put at rocking overnight at 4°C. After filtration, RNAs solutions were
516 supplemented with 3V of 100% ethanol, vortexed and put 2h at -80°C. Precipitated RNAs
517 were pelleted by centrifugation at 48,000 g during 20 min at 4°C. The supernatant was
518 discarded and RNAs pellets were carefully washed with 70% ethanol before another
519 centrifugation at 48,000 g during 20 min at 4°C. After two consecutive washes, 70% ethanol

520 was removed, RNAs pellets were dried 10 to 15 min before being resuspended in 50 mM
521 TRIS-HCl pH 8 and stored at -20°C.

522 *In vitro* RNA capping was achieved using RNAs produced by *in vitro* transcription as substrate
523 and the Vaccinia capping system (NEB). The 37-mer (50 μM), the 21-mer (50 μM) and the 14-
524 mer (20 μM) RNAs were mixed with Vaccinia Capping Enzyme buffer (NEB), 2 mM SAM, 3
525 pM GTP α-³²P (PerkinElmer), 10 Units of VCE and incubated 30 min at 37°C.

526 To produce capped molecular markers, 2X RNA loading dye was added to capped RNA and
527 solution was heated 5 min at 95°C before loading on a 20% TBE-Urea-Polyacrylamide gel.
528 Visible band corresponds to capped 37-mer (5'-m7Gppp GGA UGC UAU AAU AGU AGU GUA
529 CUA CCA AGU AUA GAU A-3') a capped 21-mer (5'-m7Gppp GGA UGC UAU AAU AGU AGU
530 GUA-3') and a capped 14-mer (5'-m7Gppp GGA UGC UAU AUA GU-3'). Additionally, capped
531 37-mer was digested using RNase T1 cleavage kit (Ambion). Cleavage reaction was
532 performed for 15 min at 37°C, stopped by adding 2X RNA loading dye and heating 5 min at
533 95°C. Observed bands corresponds to a 34-mer, a 29-mer, a 19-mer, a 17-mer and a 14-mer
534 RNA (5'-m7Gppp GGA UGC UAU AAU AG/U AG/U G/UA CUA CCA AG/U AUA G/AU A-3').

535 **Replication and transcription activity assay**

536 Synthetic RNAs (Microsynth) 5'-vRNA1-17 (5'-AGU AGU GUG CUA CCA AG-3'), 5'-1-17Bpm
537 (5'-ACG AGU GUC GUA CCA AG-3'), 3'-vRNA1-25 (5'-UAU CUA UAC UUG GUA GUA CAC UAC
538 U-3'), 3'-vRNA1-30 (5'- AAC GUU AUC UAU ACU UGG UAG UAC ACU AC U-3') were used as
539 viral RNAs (vRNA). Capped 14-mer ending by "AG" (5'-m7Gppp GAA UGC UAU AAU AG-3')
540 (TriLink Biotechnologies) and *in vitro* produced capped 14-mer ending by "AGU" (5'-m7Gppp
541 GGA UGC UAU AUA GU-3') were used as capped primers.

542 For replication activity assays, 0.6 μM LACV-L_{CItag_H34K} were mixed with (i) 0.9 μM 5'-vRNA
543 for 1h at 4°C, then 0.9 μM 3'-vRNA for overnight incubation at 4°C.

544 Reactions were launched at RT, 27°C, 30°C, 37°C for 0.5, 1, 2, 3 and 4h by adding all NTPs or
545 only ATP/GTP/UTP at 100 μM/NTP, 0.75 μCi/ml α-³²P GTP and MgCl₂/MnCl₂ in a final assay
546 buffer containing 50 mM TRIS-HCl pH 8, 150 mM NaCl, 5 mM BME, 100 μM/NTP, 0-1-2-3-4-5
547 mM MgCl₂/MnCl₂. Decade markers system (Ambion) was used as molecular weight ladder.

548 For transcription activity assays, same conditions were used as in replication activity assay
549 except that 0.9 μM capped RNA primer was added at the same time as the 5'-1-17Bpm.

550 The 37-mer capped primer, before and after RNase T1 cleavage, was used as molecular
551 weight ladder.

552 Reactions were stopped by adding 2X RNA loading dye, heating 5 min at 95°C and
553 immediately loaded on a 20% TBE-7M urea-polyacrylamide gel that was run 2h at 20 W. The
554 gel was exposed on a storage phosphor screen and read with an Amersham Typhoon
555 scanner.

556 In order to see if LACV-L_{Citag_H34K} replication products are 5' triphosphate or 5'
557 monophosphate, a replication reaction performed as described above at 30°C for 4h was
558 subjected to an incubation with 1U of Terminator Exonuclease (Lucigen) during 1h at 30°C in
559 the Terminator Exonuclease buffer. The resulting product was run next to the equivalent
560 reaction untreated with Terminator Exonuclease on a 20% Urea-PAGE gel (**Supplementary**
561 **Fig. 1e**).

562 Quantification of **Fig. 2e** was performed based on three gels done with distinct samples.
563 Measurements were done using ImageJ³⁶ and statistical analysis was done using R³⁷. A two-
564 tailed unpaired t-test was performed. P-value under 0.05 are considered significant.

565 **RNA-sequencing**

566 To confirm the identity of the products, RNA sequencing was performed. First, *in vitro*
567 replication and transcription assays were performed by incubating at 30°C 0.6 μM of LACV-
568 L_{Citag_H34K}, 0.9 μM of capped 14-mer "AG" (for transcription only), 0.9 μM 5'-1-17BPm, 0.9
569 μM 3'-vRNA1-25, 100 μM of ATP/ GTP/UTP/CTP in a buffer containing 50 mM TRIS-HCl pH 8,
570 150 mM NaCl, 5 mM BME. Reaction time was 30 min for transcription and 4h for replication.
571 After completion of the reaction, proteins were removed by Monarch RNA clean-up kit
572 (NEB). Replication reactions were subjected to RNA 5' pyrophosphohydrolase (RppH, NEB)
573 and to a second purification with the Monarch RNA clean-up kit (NEB) to obtain purified 5'-
574 monophosphate replication products.

575 RNA integrity and concentration were checked using the RNA Pico 6000 Assay Kit of the
576 Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA). Small RNA-Seq libraries
577 were prepared manually from 10 ng of RNA as input using the SMARTer[®] smRNA-Seq kit for
578 Illumina[®] Platforms (Takara Bio). Library preparation was done following the manufacturer's
579 instructions. The size distribution of the libraries was assessed on Bioanalyzer with a DNA
580 High Sensitivity kit (Agilent Technologies), and concentration was measured with Qubit[®]
581 DNA High Sensitivity kit in Qubit[®] 2.0 Fluorometer (Life Technologies). Libraries that passed
582 the QC step were pooled in equimolar amounts and final pool was purified with SPRI select
583 beads (Beckman Coulter) in a 1.3x ratio. 10 pM solution of the pool of libraries was loaded

584 on the Illumina sequencer MiSeq and sequenced uni-directionally, generating ~17 million
585 reads, each 60 bases long.

586 Raw sequencing data were processed by BCL-Convert (version 3.7.5, Illumina, San Diego,
587 USA). After basecalling the reads were trimmed using Cutadapt (version 3.4³⁸) using
588 parameters: '-u 3 -a "AAAAAAAAAA" -e 0.3 -n 3'. Trimmed reads were then subjected to
589 series of pattern-based searches using NR-grep (version 1.1.2³⁹), GNU grep (version 2.20),
590 and GNU AWK (version 4.0.2). Reads were grouped into categories based on the pattern
591 found in the sequence and counted. Detailed description of the patterns can be found in
592 **Supplementary Table 3.**

593

594 **Electron microscopy**

595 To catch structural snapshots of LACV-L_{CItag_H34K} in replication early-elongation state, 1.3 μM
596 LACV-L_{CItag_H34K} were sequentially incubated for 1h at 4°C with (i) 2.6 μM 5'-1-17BPm, (ii) 2.6
597 μM 3'-vRNA1-25. LACV-L_{CItag_H34K} bound to vRNAs was subsequently incubated for 4h at
598 30°C in a final buffer containing 50 mM TRIS-HCl pH 8, 150 mM NaCl, 5 mM MgCl₂, 5 mM
599 BME and 100 μM of each NTP (ATP/GTP/UTP).

600 To catch structural snapshots of LACV-L_{CItag_H34K} in replication late-elongation, 1.7 μM LACV-
601 L_{CItag_H34K} were sequentially incubated for 1h at 4°C with (i) 1.9 μM 5'-1-17BPm, (ii) 1.9 μM 3'-
602 vRNA1-30. LACV-L_{CItag_H34K} bound to vRNAs was subsequently incubated for 4h at 30°C in a
603 final buffer containing 50 mM TRIS-HCl pH 8, 150 mM NaCl, 5 mM MgCl₂, 5 mM BME and
604 100 μM of each NTP (ATP/GTP/UTP/CTP).

605 To catch structural snapshots of LACV-L_{CItag_H34K} in transcription, 1.3 μM LACV-L_{CItag_H34K}
606 were sequentially incubated for 1h at 4°C with (i) 1.9 μM 5'-1-17BPm and 3.9 μM
607 commercial 14-mer capped primer finishing by "AG", (ii) 1.9 μM 3'-vRNA1-25. LACV-
608 L_{CItag_H34K} bound to vRNAs and capped primer was subsequently incubated for 1h at 30°C in
609 a final buffer containing 50 mM TRIS-HCl pH 8, 150 mM NaCl, 2 mM MgCl₂, 5 mM BME and
610 100 μM of NTPs (ATP/UTP) for transcription cleavage/ capped primer entry/ initiation states
611 or 100 μM of each NTP (ATP/GTP/UTP) for transcription early-elongation state.

612 Both replication and transcription reaction were put at 4°C before grids freezing. For each
613 reaction, 3.5 μl of sample were deposited on negatively glow-discharged (25 mA, 45 sec)
614 UltraAuFoil gold grids 300 mesh, R 1.2/1.3. Excess solution was blotted for 2 sec, blot force

615 1, 100% humidity at 20°C with a Vitrobot Mark IV (Thermo Fischer Scientific) before plunge-
616 freezing in liquid ethane.

617 Automated data collections of (i) the replication early-elongation state, (ii) the replication
618 late-elongation state, (iii) the transcription initiation state and (iv) the transcription early-
619 elongation state were performed on a 200 kV Glacios cryo-TEM microscope (Thermo Fischer
620 Scientific) equipped with a K2 direct electron detector (Gatan) using SerialEM⁴⁰. Coma and
621 astigmatism correction were performed on a carbon quantifoil grid. Movies of 60 frames
622 were recorded in counting mode at a 36,000× magnification giving a pixel size of 1.145 Å
623 with defocus ranging from -0.8 to -2.0 μm. Total exposure dose was 60 e⁻/Å². The number of
624 movies per experiment are indicated in **Supplementary Figs. 3, 4 and 8** and in
625 **Supplementary Tables 1 and 2**.

626 To improve the resolution of the (i) replication initiation state, the (ii) transcription capped
627 primer entry state and the (iii) transcription initiation state, an automated data collection
628 was performed on a Titan Krios G3 (Thermo Fischer Scientific) operated at 300 kV equipped
629 with a K3 (Gatan) direct electron detector camera and a BioQuantum energy filter using
630 SerialEM. Micrographs were recorded in counting mode at a 130,000× magnification giving a
631 pixel size of 0.645 Å with defocus ranging from -0.8 to -1.8 μm. Movies of 40 frames were
632 collected with a total exposure of 50 e⁻/Å². 15573 movies were collected.

633

634 **Image processing**

635 For each collected dataset, movie drift correction was performed in Motioncor2⁴¹. For
636 images collected on the Thermo Fischer Scientific Glacios TEM, the two first frames and the
637 last ten were removed. For images collected on the Thermo Fischer Scientific Titan Krios
638 TEM, only the first two first frames were removed. Both gain reference and camera defect
639 corrections were applied. Further initial image processing steps were performed in
640 cryoSPARC v3.2.0⁴². CTF parameters were determined using “Patch CTF estimation” on non-
641 dose weighted micrographs. Realigned micrographs were then inspected and low-quality
642 micrographs displaying crystalline ice, ice contamination or aggregates for example were
643 manually discarded for further image processing. LACV-L_{Citag_H34K} particles were picked using
644 a circular blob with a diameter ranging from 90 to 170 Å. The number of particles picked is
645 indicated in **Supplementary Figs. 3, 4 and 8** and in **Supplementary Tables 1 and 2**. For
646 Thermo Fischer Scientific Glacios datasets, particles were extracted from dose-weighted

647 micrographs using a box size of 260×260 pixels². For the Thermo Fischer Scientific Titan
648 Krios dataset, the box size was 440×440 pixels². For each dataset, the same image
649 processing approach was used to separate the different LACV-L conformation and get the
650 best map quality. First, successive 2D classifications were used to eliminate bad quality
651 particles displaying poor structural features. LACV-L_{Cltag_H34K} initial 3D reconstruction was
652 generated with “Ab-initio reconstruction” in cryoSPARC using a small subset of particles. All
653 selected particles were subjected to 3D refinement with per-particle CTF estimation. The
654 rest of the image processing was done in RELION 3.1^{43,44}. For each dataset, particles were
655 divided in equal subset of 250k particles and subjected to multiple 3D classification (10
656 classes for each subset) with coarse image-alignment sampling using a circular mask of
657 170 Å. For each identical LACV-L conformation, particles were grouped and subjected to a 3D
658 refinement with a circular mask of 170 Å followed by 3D classification using local angular
659 searches based on previously determined orientations. High-resolution classes and particles
660 belonging to the same conformation were finally grouped and subjected to 3D masked
661 refinement with local angular searches to obtain high-resolution structure. Masks to
662 perform sharpening were generated using last 3D refined maps which has been low-pass
663 filtered at 10 Å, extended by 4 pixels with 8 pixels of soft-edge. Post-processing was done
664 using a manually determined *B*-factor. For each final map, reported global resolution is
665 based on the FSC 0.143 cut-off criteria and local resolution variations were also estimated in
666 RELION 3.1 (**Supplementary Figs. 3, 4 and 8**).

667 **Model building in the cryo-EM maps**

668 Both LACV-L in pre-initiation state (PDB: 6Z6G) and LACV-L in elongation mimicking state
669 (PDB: 6Z8K) models¹² were used as a starting point to manually build into the different cryo-
670 EM maps in either replication/transcription initiation/elongation states using COOT⁴⁵. All the
671 RNAs (5'-1-17BPm, 3'-vRNA1-25, 3'-vRNA1-30, the capped primer and both replication and
672 transcription products) were manually built using COOT. Each model was refined using
673 Phenix-real space refinement⁴⁶. Atomic model validation was performed using the Phenix
674 validation tool and the PDB validation server. Model resolution according to cryo-EM map
675 was estimated at the 0.5 FSC cutoff. Figures were generated using ChimeraX⁴⁷. Electrostatic
676 potential was calculated using PDB2PQR⁴⁸ and APBS⁴⁹.

677 **Multiple alignment**

678 Multiple alignment was performed using Muscle⁵⁰ and is displayed using ESPript⁵¹.

679

680 **DATA AVAILABILITY**

681 Coordinates and structure factor have been deposited in the Protein Data Bank and the
682 Electron Microscopy Data Bank.

683 LACV-L replication initiation state PDB 7ORN EMD EMD-13043

684 LACV-L replication early-elongation state PDB 7ORO EMD EMD-13044

685 LACV-L replication late-elongation state PDB 7ORI EMD EMD-13038

686 LACV-L transcription capped primer cleavage state PDB 7ORJ EMD EMD-13039

687 LACV-L transcription capped primer active site entry state PDB 7ORK EMD EMD-13040

688 LACV-L transcription initiation state PDB 7ORL EMD EMD-13041

689 LACV-L transcription early-elongation state PDB 7ORM EMD EMD-13042

690 NGS raw reads have been deposited in the European nucleotide archive under study
691 ERP132950.

692 Source data are provided with this paper.

693 **REFERENCES**

- 694 1. Maes, P. *et al.* Taxonomy of the order Bunyvirales: second update 2018. *Arch. Virol.*
695 **164**, 927–941 (2019).
- 696 2. Sun, Y., Li, J., Gao, G. F., Tien, P. & Liu, W. Bunyavirales ribonucleoproteins: the viral
697 replication and transcription machinery. *Critical Reviews in Microbiology* **44**, 522–540 (2018).
- 698 3. Reguera, J., Gerlach, P. & Cusack, S. Towards a structural understanding of RNA
699 synthesis by negative strand RNA viral polymerases. *Curr. Opin. Struct. Biol.* **36**, 75–84
700 (2016).
- 701 4. Garcin, D. *et al.* The 5' ends of Hantaan virus (Bunyaviridae) RNAs suggest a prime-
702 and-realign mechanism for the initiation of RNA synthesis. *J. Virol.* **69**, 5754–5762 (1995).
- 703 5. Polyak, S. J., Zheng, S. & Harnish, D. G. 5' termini of Pichinde arenavirus S RNAs and
704 mRNAs contain nontemplated nucleotides. *Journal of Virology* **69**, 3211–3215 (1995).
- 705 6. Raju, R. *et al.* Nontemplated bases at the 5' ends of Tacaribe virus mRNAs. *Virology*
706 **174**, 53–59 (1990).
- 707 7. Coupeau, D. *et al.* Characterization of messenger RNA termini in Schmallenberg virus
708 and related Simbuviruses. *J Gen Virol* **94**, 2399–2405 (2013).
- 709 8. Patterson, J. L. & Kolakofsky, D. Characterization of La Crosse virus small-genome
710 transcripts. *J. Virol.* **49**, 680–685 (1984).
- 711 9. Bouloy, M., Pardigon, N., Vialat, P., Gerbaud, S. & Girard, M. Characterization of the
712 5' and 3' ends of viral messenger RNAs isolated from BHK21 cells infected with germiston
713 virus (bunyavirus). *Virology* **175**, 50–58 (1990).
- 714 10. Jin, H. & Elliott, R. M. Characterization of Bunyamwera virus S RNA that is transcribed
715 and replicated by the L protein expressed from recombinant vaccinia virus. *J. Virol.* **67**, 1396–
716 1404 (1993).
- 717 11. Gerlach, P., Malet, H., Cusack, S. & Reguera, J. Structural Insights into Bunyavirus
718 Replication and Its Regulation by the vRNA Promoter. *Cell* **161**, 1267–1279 (2015).

- 719 12. Arragain, B. *et al.* Pre-initiation and elongation structures of full-length La Crosse
720 virus polymerase reveal functionally important conformational changes. *Nat Commun* **11**,
721 3590 (2020).
- 722 13. Peng, R. *et al.* Structural insight into arenavirus replication machinery. *Nature* **579**,
723 615–619 (2020).
- 724 14. Vogel, D. *et al.* Structural and functional characterization of the severe fever with
725 thrombocytopenia syndrome virus L protein. *Nucleic Acids Res.* **48**, 5749–5765 (2020).
- 726 15. Wang, P. *et al.* Structure of severe fever with thrombocytopenia syndrome virus L
727 protein elucidates the mechanisms of viral transcription initiation. *Nat Microbiol* **5**, 864–871
728 (2020).
- 729 16. Reguera, J., Weber, F. & Cusack, S. Bunyaviridae RNA polymerases (L-protein) have an
730 N-terminal, influenza-like endonuclease domain, essential for viral cap-dependent
731 transcription. *PLoS Pathog.* **6**, e1001101 (2010).
- 732 17. Wandzik, J. M. *et al.* A Structure-Based Model for the Complete Transcription Cycle of
733 Influenza Polymerase. *Cell* **181**, 877–893.e21 (2020).
- 734 18. Vialat, P. & Bouloy, M. Germiston virus transcriptase requires active 40S ribosomal
735 subunits and utilizes capped cellular RNAs. *J Virol* **66**, 685–693 (1992).
- 736 19. Deng, T., Vreede, F. T. & Brownlee, G. G. Different De Novo Initiation Strategies Are
737 Used by Influenza Virus RNA Polymerase on Its cRNA and Viral RNA Promoters during Viral
738 RNA Replication. *J. Virol.* **80**, 2337–2348 (2006).
- 739 20. Fan, H. *et al.* Structures of influenza A virus RNA polymerase offer insight into viral
740 genome replication. *Nature* **573**, 287–290 (2019).
- 741 21. Kohl, A., Dunn, E. F., Lowen, A. C. & Elliott, R. M. Complementarity, sequence and
742 structural elements within the 3' and 5' non-coding regions of the Bunyamwera
743 orthobunyavirus S segment determine promoter strength. *J Gen Virol* **85**, 3269–3278 (2004).
- 744 22. Kohl, A., Bridgen, A., Dunn, E., Barr, J. N. & Elliott, R. M. Effects of a point mutation in
745 the 3' end of the S genome segment of naturally occurring and engineered Bunyamwera
746 viruses. *J Gen Virol* **84**, 789–793 (2003).
- 747 23. Reich, S. *et al.* Structural insight into cap-snatching and RNA synthesis by influenza
748 polymerase. *Nature* **516**, 361–6 (2014).
- 749 24. Habjan, M. *et al.* Processing of Genome 5' Termini as a Strategy of Negative-Strand
750 RNA Viruses to Avoid RIG-I-Dependent Interferon Induction. *PLOS ONE* **3**, e2032 (2008).
- 751 25. Oymans, J. & Te Velthuis, A. J. W. A Mechanism for Priming and Realignment during
752 Influenza A Virus Replication. *J Virol* **92**, (2018).
- 753 26. Kouba, T., Drncová, P. & Cusack, S. Structural snapshots of actively transcribing
754 influenza polymerase. *Nat. Struct. Mol. Biol.* **26**, 460–470 (2019).
- 755 27. Te Velthuis, A. J. W., Robb, N. C., Kapanidis, A. N. & Fodor, E. The role of the priming
756 loop in influenza A virus RNA synthesis. *Nat Microbiol* **1**, 16029 (2016).
- 757 28. Carrique, L. *et al.* Host ANP32A mediates the assembly of the influenza virus
758 replicase. *Nature* **587**, 638–643 (2020).
- 759 29. Pflug, A., Guilligay, D., Reich, S. & Cusack, S. Structure of influenza A polymerase
760 bound to the viral RNA promoter. *Nature* **516**, 355–60 (2014).
- 761 30. Amroun, A., Priet, S. & Querat, G. Toscana virus cap-snatching and initiation of
762 transcription. *J Gen Virol* **98**, 2676–2688 (2017).
- 763 31. Liu, X. *et al.* Rice Stripe Tenuivirus Has a Greater Tendency To Use the Prime-and-
764 Realign Mechanism in Transcription of Genomic than in Transcription of Antigenomic
765 Template RNAs. *Journal of Virology* **92**, e01414-17 (2018).

- 766 32. Te Velthuis, A. J. W. & Oymans, J. Initiation, Elongation, and Realignment during
767 Influenza Virus mRNA Synthesis. *J Virol* **92**, e01775-17 (2018).
- 768 33. Blakqori, G., Kochs, G., Haller, O. & Weber, F. Functional L polymerase of La Crosse
769 virus allows in vivo reconstitution of recombinant nucleocapsids. *J Gen Virol* **84**, 1207–1214
770 (2003).
- 771 34. Klemm, C. *et al.* Systems To Establish Bunyavirus Genome Replication in the Absence
772 of Transcription. *J Virol* **87**, 8205–8212 (2013).
- 773 35. Nie, Y., Bellon-Echeverria, I., Trowitzsch, S., Bieniossek, C. & Berger, I. Multiprotein
774 complex production in insect cells by using polyproteins. *Methods Mol. Biol.* **1091**, 131–141
775 (2014).
- 776 36. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of
777 image analysis. *Nat Methods* **9**, 671–675 (2012).
- 778 37. R Core Team. (2016). R: A Language and Environment for Statistical Computing.
779 Vienna, Austria. Retrieved from <https://www.R-project.org/>. in.
- 780 38. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing
781 reads. *EMBnet.journal* **17**, 10–12 (2011).
- 782 39. NR-grep: a fast and flexible pattern-matching tool - Navarro - 2001 - Software:
783 Practice and Experience - Wiley Online Library.
784 <https://onlinelibrary.wiley.com/doi/10.1002/spe.411>.
- 785 40. Schorb, M., Haberbosch, I., Hagen, W. J. H., Schwab, Y. & Mastronarde, D. N.
786 Software tools for automated transmission electron microscopy. *Nat. Methods* **16**, 471–477
787 (2019).
- 788 41. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-
789 atomic-resolution single-particle cryo-EM. *Nat Meth* **10**, 584–590 (2013).
- 790 42. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for
791 rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
- 792 43. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM
793 structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
- 794 44. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure
795 determination in RELION-3. *Elife* **7**, (2018).
- 796 45. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta*
797 *Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
- 798 46. Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and
799 crystallography. *Acta Crystallogr D Struct Biol* **74**, 531–544 (2018).
- 800 47. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and
801 analysis. *Protein Sci.* **27**, 14–25 (2018).
- 802 48. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an automated
803 pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* **32**,
804 W665-667 (2004).
- 805 49. Jurrus, E. *et al.* Improvements to the APBS biomolecular solvation software suite.
806 *Protein Sci* **27**, 112–128 (2018).
- 807 50. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high
808 throughput. *Nucleic Acids Res* **32**, 1792–1797 (2004).
- 809 51. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new
810 ENDscript server. *Nucleic Acids Res* **42**, W320–W324 (2014).

811

812 **ACKNOWLEDGEMENTS**

813 We thank Karine Huard, Angélique Fraudeau, Petra Drncová, Alice Aubert and Martin
814 Pelosse for technical advices on expression and purification; Friedemann Weber for
815 providing the LACV mini-replicon to S.C.; Sissy Kalayil, Tomas Kouba, Anna Dubankova and
816 Joanna Wandzik for technical advices on radioactivity experiments; Vladimir Benes from
817 EMBL Heidelberg GeneCore for NGS; Lionel Imbert and Alice Stelfox for advices on *in vitro*
818 transcription; Simon Fromm for data collection on the Heidelberg Titan Krios; Aymeric Peuch
819 for setting up and maintaining the EM computing cluster; Piotr Gerlach and Juan Reguera for
820 useful discussions and Daphna Fenel for technical support.

821 This work used the platforms of the Grenoble Instruct-ERIC center (ISBG; UAR 3518 CNRS-
822 CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology (PSB), supported by
823 FRISBI (ANR-10-INBS-05-02) and GRAL, financed within the University Grenoble Alpes
824 graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). The
825 electron microscope facility is supported by the Auvergne-Rhône-Alpes Region, the
826 Fondation pour la Recherche Médicale (FRM), the fonds FEDER and the GIS-Infrastructures
827 en Biologie Santé et Agronomie (IBiSA). This work benefited from access to the cryo-electron
828 microscopy platform of the European Molecular Biology Laboratory (EMBL) in Heidelberg
829 and has been supported by iNEXT-Discovery, project number 871037, funded by the Horizon
830 2020 program of the European Commission. We thank all platform staff that enabled us to
831 perform these analyses. IBS acknowledges integration into the Interdisciplinary Research
832 Institute of Grenoble (IRIG, CEA).

833 This work was supported by the ANR-19-CE11-0024-02 and the Institut Universitaire de
834 France endowment to H.M., the FRM grant number FDT202012010396 to B.A.

835 **AUTHOR CONTRIBUTIONS**

836 F.B. and B.A. cloned LACV-L FL with internal tags. F.B. performed mini-replicon assays. B.A.
837 and Q.D.T. cloned mutant constructs. B.A. and Q.D.T. expressed and purified LACV-L. B.A.
838 performed *in vitro* activity assays. B.A. and H.M. prepared cryo-EM grids. H.M. and B.A.
839 collected cryo-EM data on a Thermo Fischer Scientific Glacios EM thanks to advices and
840 training from G.S. who set up and maintains the IBS-ISBG EM platform. B.A. performed cryo-

841 EM image processing. B.A and H.M. built the models based on the cryo-EM maps with inputs
842 from S.C. B.A. and H.M. performed structural analysis. N.A. performed RNA library
843 preparation and NGS. J.P. performed bioinformatic analysis of NGS. H.M. and G.S. co-
844 supervise B.A. and Q.D.T. The project was conceived by H.M. with inputs from S.C. and G.S.
845 This project used funding obtained by H.M. and G.S. The manuscript was written by H.M.
846 and B.A. with input from all authors.

847

848 **COMPETING INTERESTS STATEMENT**

849 The authors declare no competing interests.

850

851 **FIGURE LEGENDS**

852 **Fig. 1. Overview of LACV-L activity and cryo-EM structures stalled in specific replication**
853 **states**

854 **a**, Histogram of LACV-L mini-replicon activity. The percentage of activity in presence of
855 different tags is compared to LACV-L activity without tag (WT). Tag positions are shown on
856 LACV-L_{Nter-Histag} surface (PDB: 6Z6G) as dotted lines. The polymerase core is colored in light
857 grey with the California insertion in orange, the CBD in gold, the ENDO in green, the mid
858 domain in beige and the ZBD in brown.

859 **b**, *In vitro* replication activity of LACV-L_{Citag_H34K} using different combinations of 5'- and 3'-
860 vRNAs. The assays are done in presence of either the 4 NTPs (AUGC), 3 NTPs (AUG) or 3 NTPs
861 during 4h, subsequently supplemented with CTP for 30 min (AUG+C). Replication products
862 and their respective lengths are displayed on the right side of the gel. The decade molecular
863 weight marker (MW) is shown.

864 **c**, Sequence and secondary structures of the RNA bound to LACV-L_{Citag_H34K} for each
865 replication state. 5'-vRNAs, 3'-vRNAs and replication products are respectively colored in
866 pink/grey, gold and cyan. Bases present in the sequences but not seen in the structures are
867 shown in transparent.

868 **d**, Cartoon representation of each LACV-L_{Citag_H34K} replication structure. LACV-L domains are
869 colored as in **(a)**. RNAs displayed as surfaces are colored as in **(a,c)**. The pre-initiation state
870 corresponds to the PDB 6Z6G.

871

872 **Fig. 2. Interactions between LACV-L_{Citag_H34K} and 5'/ 3'-vRNA during replication**

873 **a**, Interaction between LACV-L_{Citag_H34K} and the distal duplex formed by 5'-1-17BPm and 3'-
874 vRNA1-25. Interacting domains, namely the arch, the vRBL, the α -ribbon, the fingers, the
875 bridge and the thumb ring domains are respectively colored in magenta, violet red, steel
876 blue, cyan, purple and blue. The RNAs are colored as in **Fig. 1c,d**. ATP in position +1 of the
877 active site is surrounded by a dotted line.

878 **b**, Zoom on the interaction between LACV-L_{Citag_H34K} and the 3'-vRNA1-25. Surface of
879 interacting domains are displayed and colored according to **(a)**. Residues that interact with
880 the 3'-vRNA template are labelled. Flexible nucleotides A6/C7/A8 of the 3'-vRNA1-25 are
881 surrounded by a dotted line. The two visible conformations of A8 are shown.

882 **c**, Active site of LACV-L_{Citag_H34K} in replication initiation state. RdRp motifs A, B, C, D and F are
883 respectively colored in dark turquoise, purple, grey, light green and blue. The extended PR
884 loop and ENDO are colored in orchid and green. Residues that interact with the 3'-vRNA1-25
885 are displayed. The two magnesium ions are colored in light green.

886 **d**, Active site of LACV-L_{Citag_H34K} in replication elongation state. LACV-L domains, RdRp motifs
887 and vRNAs are colored as in **(c)**. The newly synthesized product forming a 9 base-pair duplex
888 with the 3'-vRNA1-25 template is colored in cyan. The +1 active site position is surrounded
889 by a dotted line.

890 **e**, *In vitro* replication activity of LACV-L_{Citag-H34K}, LACV-L_{Citag-H34K-M989A} (M989A), LACV-L_{Citag-H34K-}
891 _{I990A} (I990A), LACV-L_{Citag-H34K-S991A} (S991A) with 5'-1-17BPm, 3'-vRNA1-25 and 4 NTPs.
892 Replication products and their respective lengths are displayed on the right side of the gel.
893 Histogram of product formation. Intensities are mean \pm s.d. n=3 independent reactions. P-
894 value for two-tailed unpaired t-test between of LACV-L_{Citag-H34K} and each LACV-L PR loop
895 mutant are shown. P <0.05 is considered significant. Intensity ratios (LACV-L PR loop
896 mutant/LACV-L_{Citag-H34K}) are displayed.

897

898 **Fig. 3. Replication-induced conformational changes of LACV-L_{Citag_H34K}**

899 **a**, ENDO and PR loop movements between the replication states. ENDO N-ter and C-ter are
900 indicated. Arrows indicate movements of the ENDO, the α 7 helix and the PR loop.

901 **b**, LACV-L domain movements and RNA position variations between the replication states.
902 For clarity, only domains that undergo conformational changes are displayed. The template
903 exit plug, colored in red, is indicated with a dotted line if flexible. RNAs are shown as

904 surfaces and colored as in **Fig. 1c,d**. Bridge residues 1423-1441 are surrounded by a dotted
905 line.

906 **c**, Template exit plug, vRBL, clamp and α -ribbon conformational changes between the
907 replication states. LACV-L surface is displayed and each domain are colored as in **Fig. 2a,b**
908 and **3b**. The template exit tunnel is surrounded by a dotted line. Template entry/exit tunnels
909 are indicated in the early-elongation state with gold arrows.

910

911 **Fig. 4. Overview of LACV-L_{Citag_H34K} activity and cryo-EM structures in transcription**

912 **a**, *In vitro* transcription activity of LACV-L_{Citag_H34K} using different combination of 5'/3'-vRNAs,
913 NTPs and capped primers (capped 14-mer finishing by "AG" (cap14AG) and *in vitro* produced
914 capped 14-mer finishing by "AGU" (IVT cap14AGU)). The cap37-mer lane corresponds to a
915 capped 37-mer identical in sequence to the theoretical transcription product of LACV-
916 L_{Citag_H34K} with the cap14AG and the 3'-vRNA1-25 without prime-and-realign. The lane
917 indicated as cap37-mer T1 corresponds to a RNase T1 cleavage of the cap37. Dotted squares
918 indicate transcription products of interest that are reported in **b**.

919 **b**, Zoom on LACV-L_{Citag_H34K} transcription products. Lane numbers are referring to **a**.
920 Transcription products length are indicated on the right side of the cropped gels with the
921 corresponding molecular weight ladder.

922 **c**, Sequence and secondary structures of 5'/3'-vRNAs, capped RNA primer and transcription
923 products. 5'-vRNAs, 3'-vRNAs and capped primer/product are respectively colored in
924 pink/grey, gold and cyan. Bases present in the sequences but not seen in the structures are
925 shown in transparent.

926 **d**, Cartoon representation of each LACV-L_{Citag_H34K} transcription structure. Note that LACV-L in
927 "capped primer active site entry" state has the same conformation as in "initiation" state.
928 LACV-L domains are colored as in **Fig. 1**. RNAs are displayed as surfaces and colored as in **c**.

929 **e**, ENDO and CBD movements during transcription. The ENDO, CBD, Mid/ZBD, RNAs are
930 displayed as surfaces and colored as in **d**. The ENDO active site and CBD cap-binding site are
931 colored in red and blue. The closing/opening of product exit tunnel is highlighted by white
932 arrows.

933

934 **Fig. 5. LACV-L_{Citag_H34K} conformational changes induced by transcription initiation**

935 **a**, Mid and ZBD movements from pre-initiation to transcription initiation states. LACV-L core
936 is transparent. Mid and ZBD are colored in grey in the pre-initiation state, in beige and
937 brown in the transcription initiation state and their rotation highlighted with an arrow.
938 **b**, CTER interaction with the polymerase core and the ENDO during transcription initiation.
939 Both LACV-L core and ENDO are displayed as grey surface. The CTER domains (mid, CBD,
940 ZBD) are displayed as cartoon. All interacting regions are displayed, colored and numbered.
941 The CBD loop close to the CBD-binding site is surrounded by a dotted line.

942

943 **Fig. 6. Zoom on essential RNA-LACV-L contacts during transcription**

944 **a**, 5'-cap interaction with LACV-L_{Citag_H34K} in the transcription cleavage conformation. The
945 ENDO active site residues are colored in red. The capped primer is colored in cyan and the
946 dotted line stands for the potential cleavage path. On the right panel, residues implicated in
947 the 5'-capped stabilization are displayed.

948 **b**, Capped primer path from the CBD active site to the RdRp active site during transcription
949 initiation. On the right panel, residues that are proximal to the capped primer are indicated.

950 **c**, *In vitro* transcription activity of LACV-L_{Citag_H34K}, LACV-L_{Citag_H34K-M989A} (M989A), LACV-L_{Citag-}
951 _{H34K-I990A} (I990A), LACV-L_{Citag_H34K-S991A} (S991A) with 5'-1-17Bpm, cap14AG, 3'-vRNA1-25 and
952 four NTPs. Transcription products are labelled. Cap37-mer and cap37-mer T1 lanes are
953 equivalent to **Fig. 4a** description.

954 **d**, Interaction between the capped RNA product and LACV-L_{Citag_H34K} in transcription
955 elongation state. Residues stabilizing it are indicated. Flexible nucleotides are represented as
956 a dotted line.

957

958 **Fig. 7. LACV-L replication and transcription cycle models**

959 **a**, Structure-based model of the mechanisms underlying LACV-L replication and
960 transcription. All structures are displayed as surfaces with specific domains and features
961 colored as in **Fig. 1 and 2**. The vRNA is shown as a line with the 5'/3' ends colored in
962 pink/yellow. At initiation, the 3'-vRNA terminus is released from the secondary binding site
963 and brought into the RdRp active site where it is stabilized by the PR loop, that extends due
964 to the ENDO movement. If the flexible CBD snatches a 5'-cap, important conformational
965 changes of the CBD, the ENDO, the mid and the ZBD occur, triggering capped primer
966 translocation into the ENDO active site (in red). Once cleaved (yellow star), an ENDO

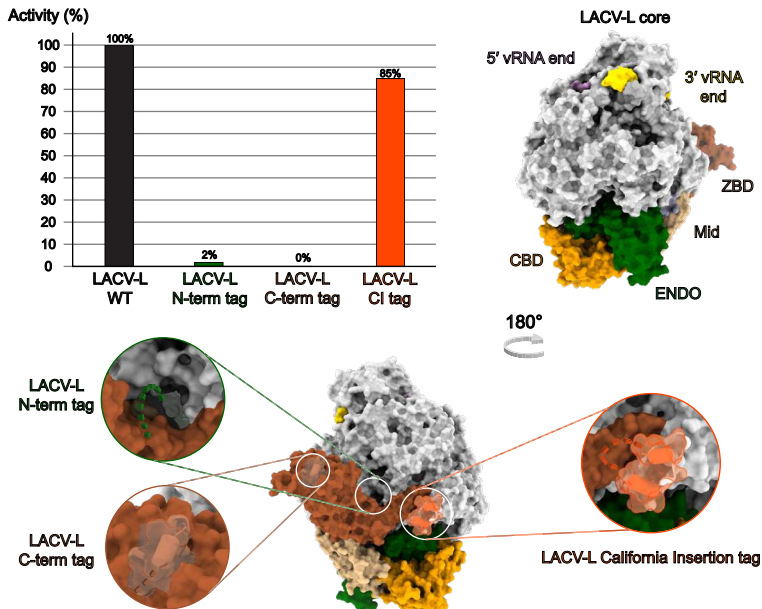
967 movement triggers the capped primer repositioning towards the RdRp active site. Early
968 elongation induces extrusion of the template exit plug resulting in the template exit opening
969 (yellow dotted circle). In late-elongation, the 3'-vRNA template end goes back to the
970 secondary binding site.

971 **b,c**, Structure-based model of initiation by prime-and-realign for replication **(b)** and
972 transcription **(c)**. The 5'/3'-vRNA ends are respectively colored in pink and gold. The 5'-hook
973 structure is represented as a dotted line. Incorporated nucleotides are colored in blue. The
974 PR loop is shown in pink. The binding pocket of nucleotides A6, C7 and A8 at priming stage is
975 surrounded by a dotted square. The proposed successive steps of the models are presented
976 from left to right. The main role of each step is indicated below each schematic.

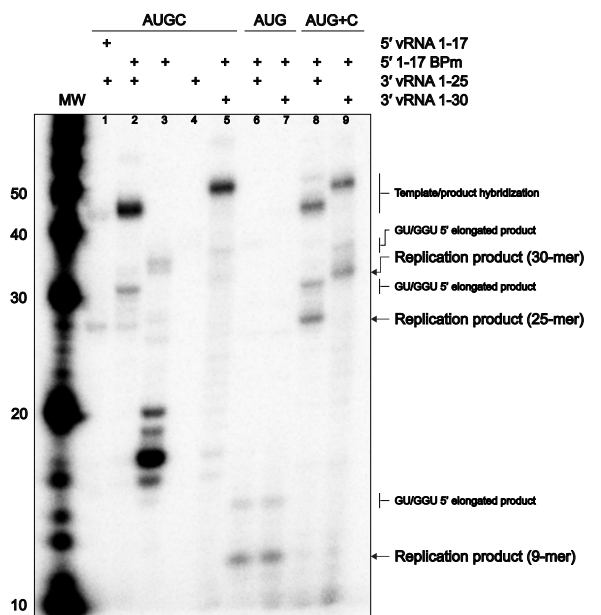
977
978

FIGURE 1

a

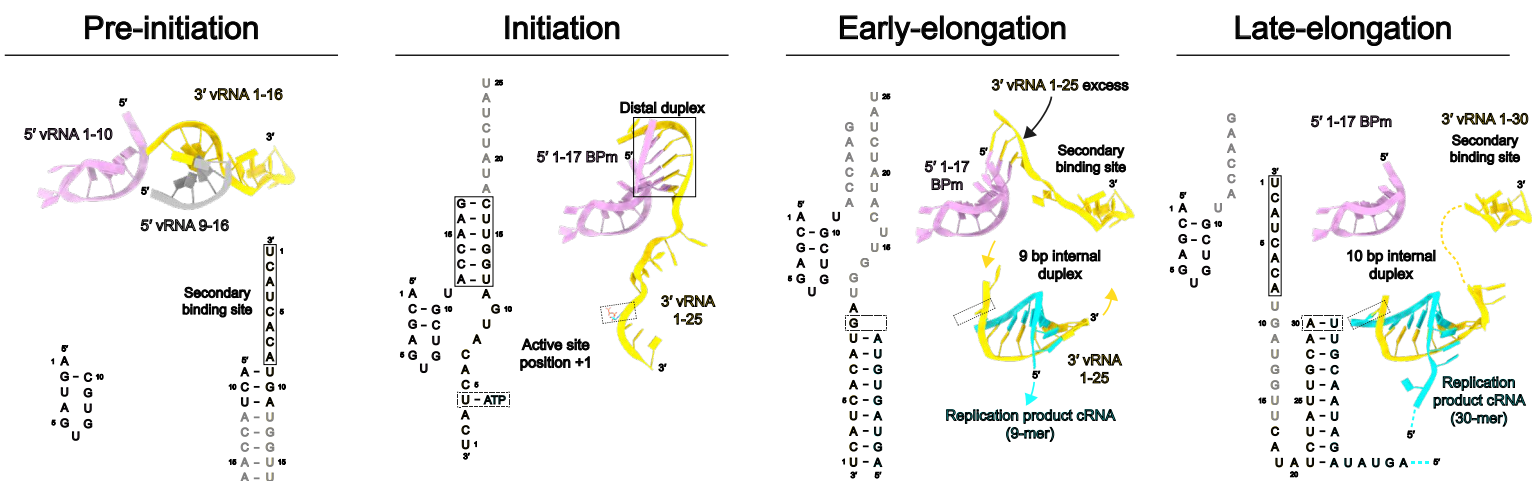


b



c

Replication



d

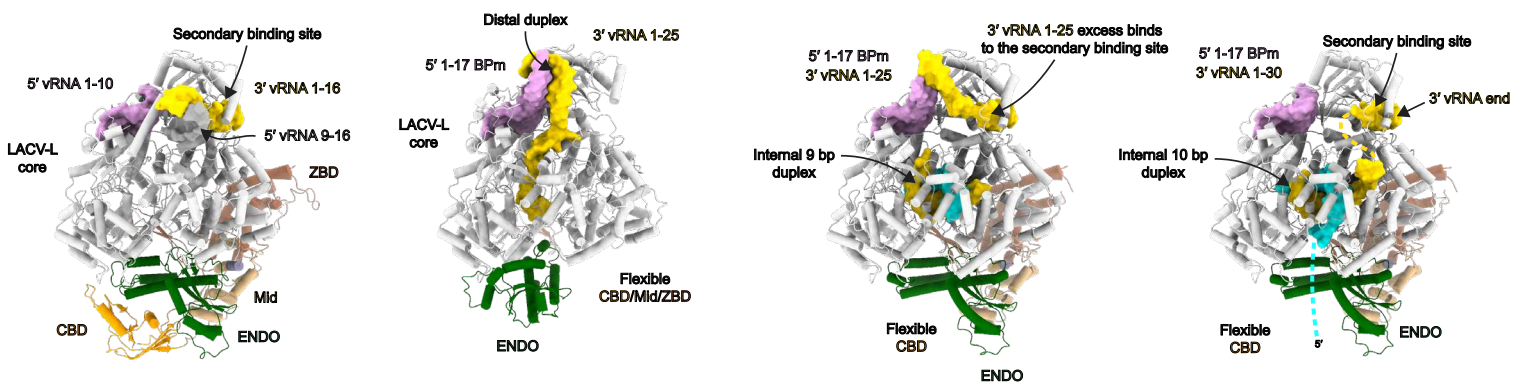


Fig. 1. Overview of LACV-L activity and cryo-EM structures stilled in specific replication states

a, Histogram of LACV-L mini-replicon activity. The percentage of activity in presence of different tags is compared to LACV-L activity without tag (WT). Tag positions are shown on LACV-L_{Nter-Histag} surface (PDB: 6Z6G) as dotted lines. The polymerase core is colored in light grey with the California insertion in orange, the CBD in gold, the ENDO in green, the mid domain in beige and the ZBD in brown.

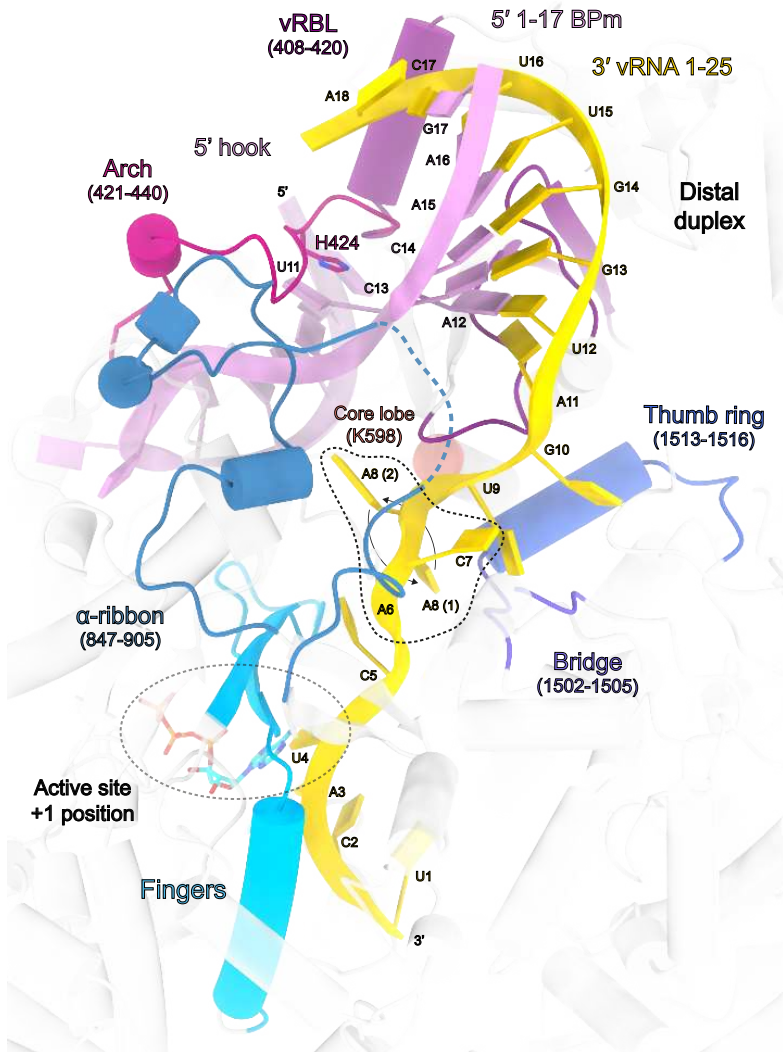
b, In vitro replication activity of LACV-L_{Cltag_H34K} using different combinations of 5'- and 3'-vRNAs. The assays are done in presence of either the 4 NTPs (AUGC), 3 NTPs (AUG) or 3 NTPs (AUG) during 4h, subsequently supplemented with CTP for 30 min (AUG+C). Replication products and their respective lengths are displayed on the right side of the gel. The decade molecular weight marker (MW) is shown.

c, Sequence and secondary structures of the RNA bound to LACV-L_{Cltag_H34K} for each replication state. 5'-vRNAs, 3'-vRNAs and replication products are respectively colored in pink/grey, gold and cyan. Bases present in the sequences but not seen in the structures are shown in transparent.

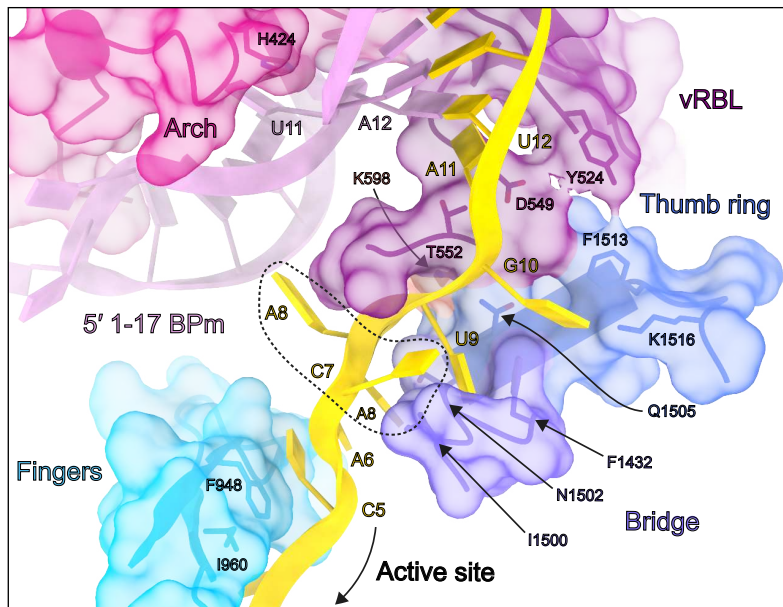
d, Cartoon representation of each LACV-L_{Cltag_H34K} replication structure. LACV-L domains are colored as in (a). RNAs displayed as surfaces are colored as in (a,c). The pre-initiation state corresponds to the PDB 6Z6G.

FIGURE 2

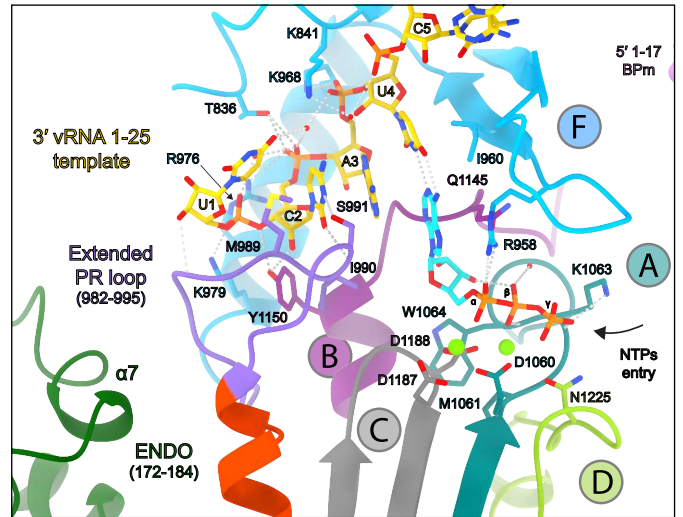
a



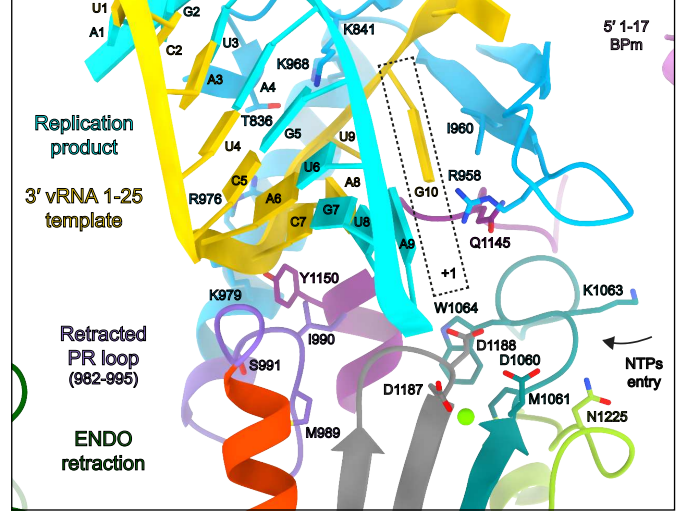
b



c



d



e

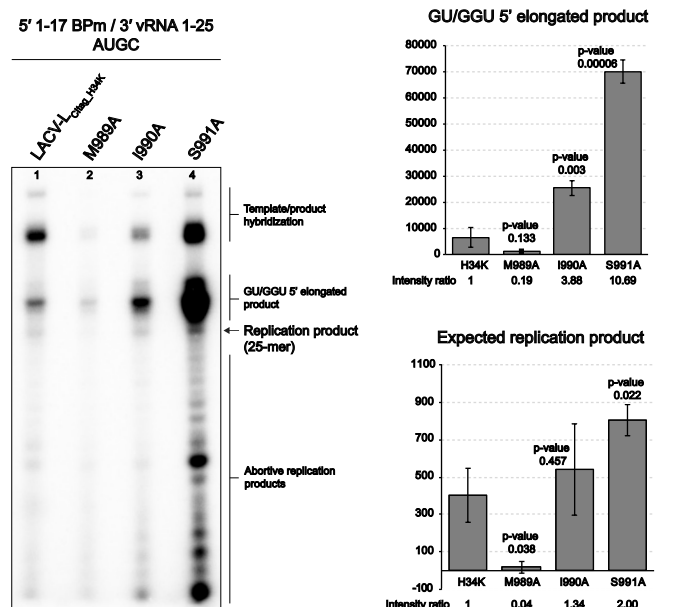


Fig. 2. Interactions between LACV-L_{Clitag_H34K} and 5' 3'-vRNA during replication

a, Interaction between LACV-L_{Clitag_H34K} and the distal duplex formed by 5'-1-17BPM and 3'-vRNA1-25. Interacting domains, namely the arch, the vRBL, the α -ribbon, the fingers, the bridge and the thumb ring domains are respectively colored in magenta, violet red, steel blue, cyan, purple and blue. The RNAs are colored as in Fig. 1c,d. ATP in position +1 of the active site is surrounded by a dotted line.

b, Zoom on the interaction between LACV-L_{Clitag_H34K} and the 3'-vRNA1-25. Surface of interacting domains are displayed and colored according to (a). Residues that interact with the 3'-vRNA template are labeled. Flexible nucleotides A6/C7/A8 of the 3'-vRNA1-25 are surrounded by a dotted line. The two visible conformations of A8 are shown.

c, Active site of LACV-L_{Clitag_H34K} in replication initiation state. RdRp motifs A, B, C, D and F are respectively colored in dark turquoise, purple, grey, light green and blue. The extended PR loop and ENDO are colored in orchid and green. Residues that interact with the 3'-vRNA1-25 are displayed. The two magnesium ions are colored in light green.

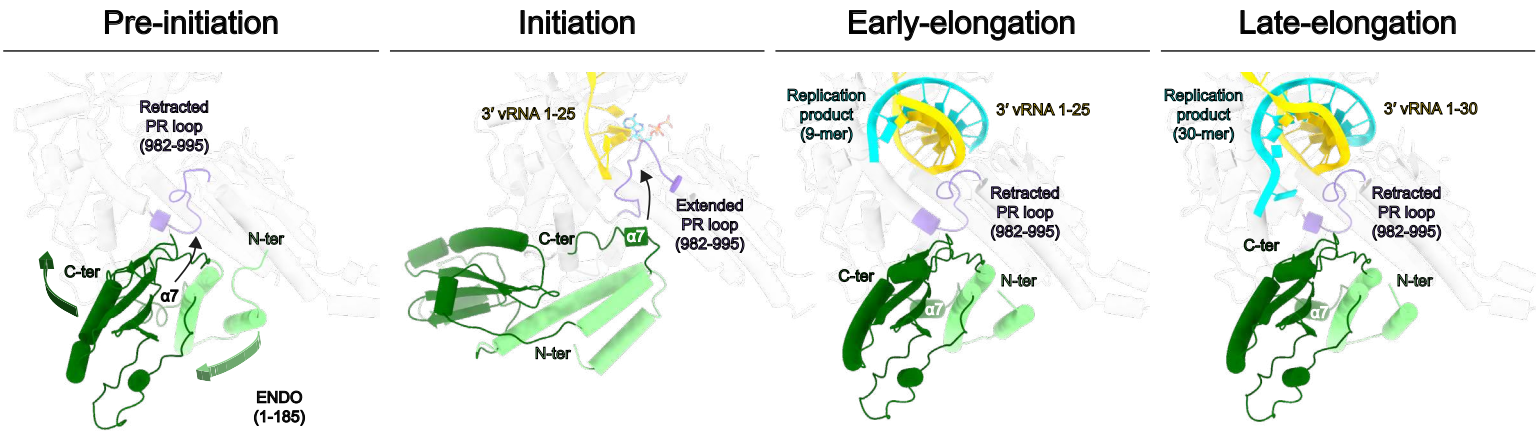
d, Active site of LACV-L_{Clitag_H34K} in replication elongation state. LACV-L domains, RdRp motifs and vRNAs are colored as in (c). The newly synthesized product forming a 9 base-pair duplex with the 3'-vRNA1-25 template is colored in cyan. The +1 active site position is surrounded by a dotted line.

e, *In vitro* replication activity of LACV-L_{Clitag_H34K}, LACV-L_{Clitag_H34K-M989A} (M989A), LACV-L_{Clitag_H34K-I990A} (I990A), LACV-L_{Clitag_H34K-S991A} (S991A) with 5'-1-17BPM, 3'-vRNA1-25 and 4 NTPs. Replication products and their respective lengths are displayed on the right side of the gel. Histogram of product formation. Intensities are mean \pm s.d. n=3 independent reactions. P-value for two-tailed unpaired t-test between of LACV-L_{Clitag_H34K} and each LACV-L PR loop mutant are shown. P < 0.05 is considered significant. Intensity ratios (LACV-L PR loop mutant/LACV-L_{Clitag_H34K}) are displayed.

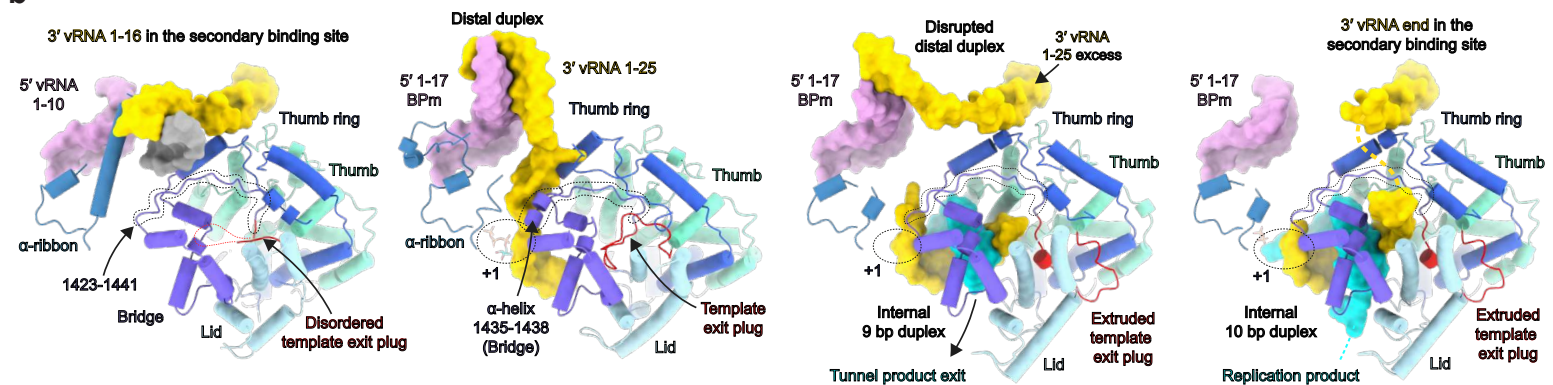
FIGURE 3

Replication

a



b



c

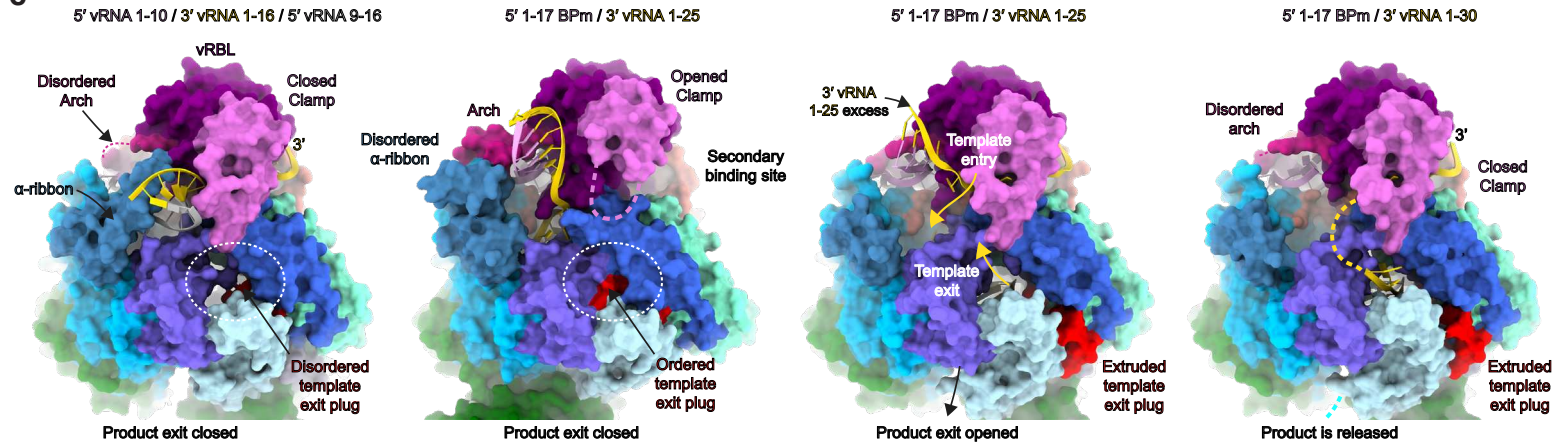


Fig. 3. Replication-induced conformational changes of LACV-L_{C118Q_H34K}

a, ENDO and PR loop movements between the replication states. ENDO N-ter and C-ter are indicated. Arrows indicate movements of the ENDO, the $\alpha 7$ helix and the PR loop.

b, LACV-L domain movements and RNA position variations between the replication states. For clarity, only domains that undergo conformational changes are displayed. The template exit plug, colored in red, is indicated with a dotted line if flexible. RNAs are shown as surfaces and colored as in Fig. 1c,d. Bridge residues 1423-1441 are surrounded by a dotted line.

c, Template exit plug, vRBL, clamp and α -ribbon conformational changes between the replication states. LACV-L surface is displayed and each domain are colored as in Fig. 2a,b and 3b. The template exit tunnel is surrounded by a dotted line. Template entry/exit tunnels are indicated in the early-elongation state with gold arrows.

FIGURE 4

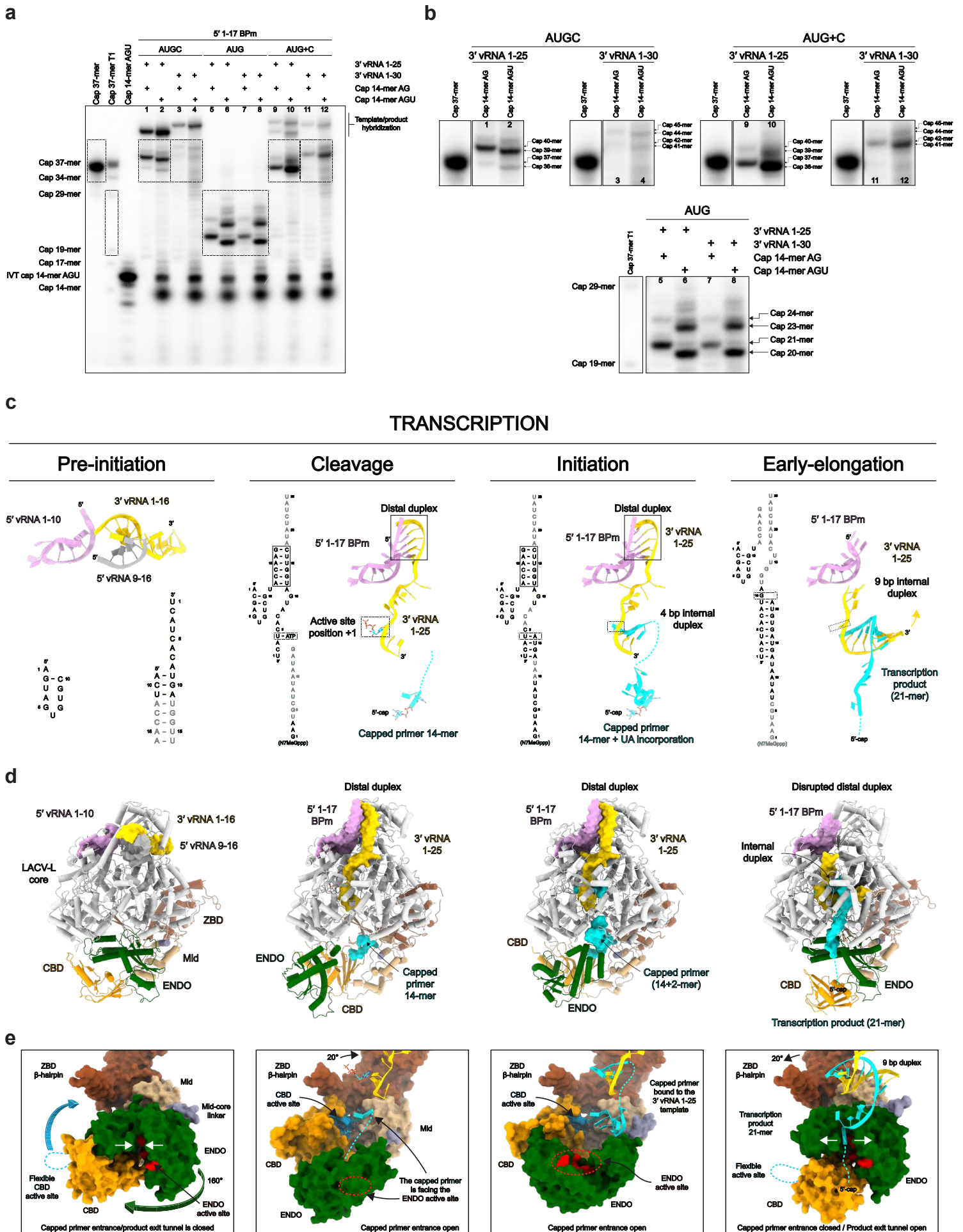


Fig. 4. Overview of LACV-L_{C1tag_H34K} activity and cryo-EM structures in transcription
a. *In vitro* transcription activity of LACV-L_{C1tag_H34K} using different combination of 5'/3'-vRNAs, NTPs and capped primers (capped 14-mer finishing by "AG" (cap14AG) and *in vitro* produced capped 14-mer finishing by "AGU" (IVT cap14AGU)). The cap37-mer lane corresponds to a capped 37-mer identical in sequence to the theoretical transcription product of LACV-L_{C1tag_H34K} with the cap14AG and the 3'-vRNA1-25 without prime-and-realign. The lane indicated as cap37-mer T1 corresponds to a RNase T1 cleavage of the cap37. Dotted squares indicate transcription products of interest that are reported in **b**.
b. Zoom on LACV-L_{C1tag_H34K} transcription products. Lane numbers are referring to **a**. Transcription products length are indicated on the right side of the cropped gels with the corresponding molecular weight ladder.
c. Sequence and secondary structures of 5'/3'-vRNAs, capped RNA primer and transcription products. 5'-vRNAs, 3'-vRNAs and capped primer/product are respectively colored in pink/grey, gold and cyan. Bases present in the sequences but not seen in the structures are shown in transparent.
d. Cartoon representation of each LACV-L_{C1tag_H34K} transcription structure. Note that LACV-L in "capped primer active site entry" state has the same conformation as in "initiation" state. LACV-L domains are colored as in Fig. 1. RNAs are displayed as surfaces and colored as in **c**.
e. ENDO and CBD movements during transcription. The ENDO, CBD, Mid/ZBD, RNAs are displayed as surfaces and colored as in **d**. The ENDO active site and CBD cap-binding site are colored in red and blue. The closing/opening of product exit tunnel is highlighted by white arrows.

FIGURE 5

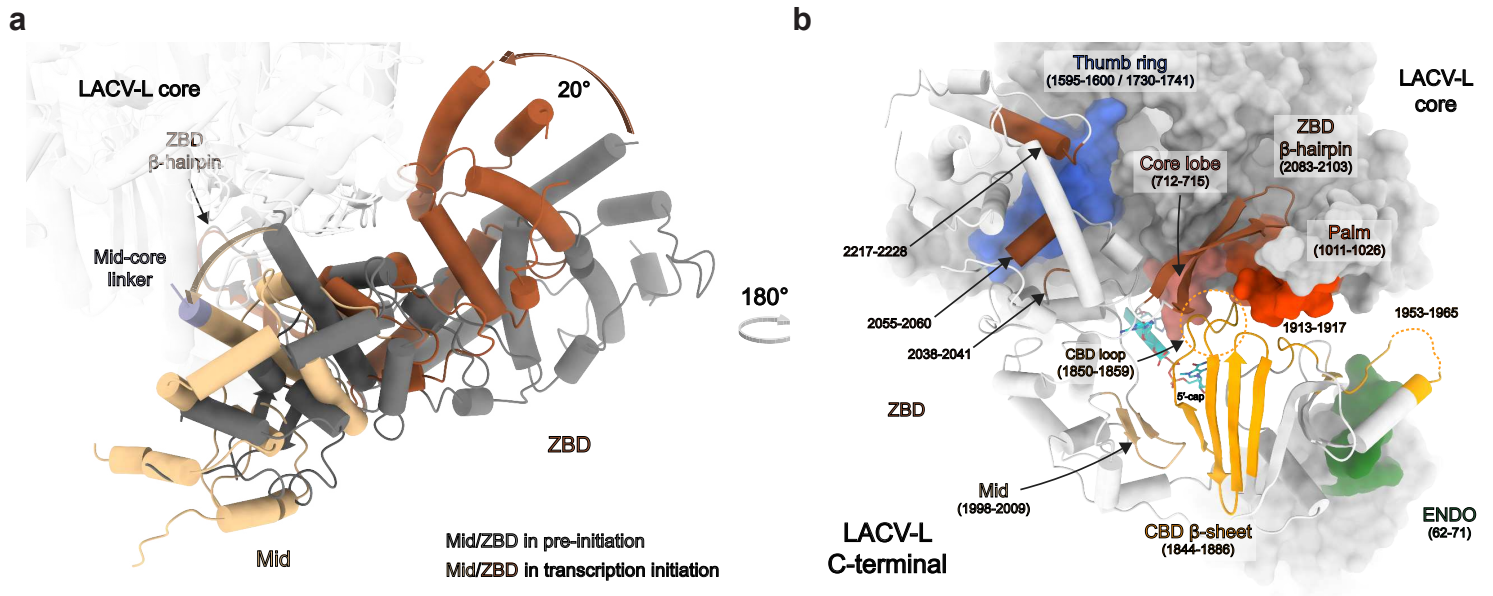


Fig. 5. LACV-L_{C1tag_H34K} conformational changes induced by transcription initiation

a, Mid and ZBD movements from pre-initiation to transcription initiation states. LACV-L core is transparent. Mid and ZBD are colored in grey in the pre-initiation state, in beige and brown in the transcription initiation state and their rotation highlighted with an arrow.

b, CTER interaction with the polymerase core and the ENDO during transcription initiation. Both LACV-L core and ENDO are displayed as grey surface. The CTER domains (mid, CBD, ZBD) are displayed as cartoon. All interacting regions are displayed, colored and numbered. The CBD loop close to the CBD-binding site is surrounded by a dotted line.

FIGURE 6

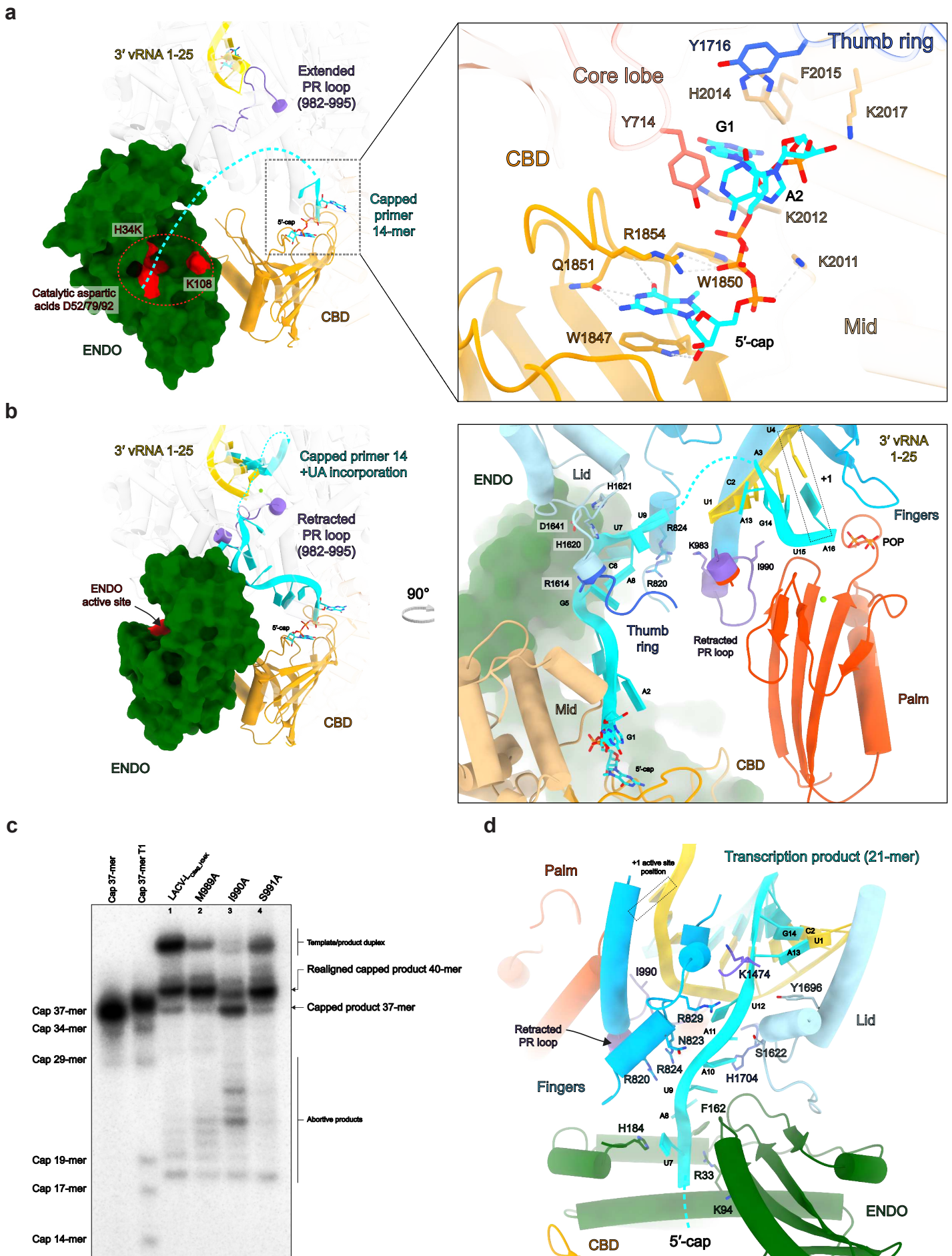


Fig. 6. Zoom on essential RNA-LACV-L contacts during transcription

a, 5'-cap interaction with LACV-L_{Clitag_H34K} in the transcription cleavage conformation. The ENDO active site residues are colored in red. The capped primer is colored in cyan and the dotted line stands for the potential cleavage path. On the right panel, residues implicated in the 5'-capped stabilization are displayed.

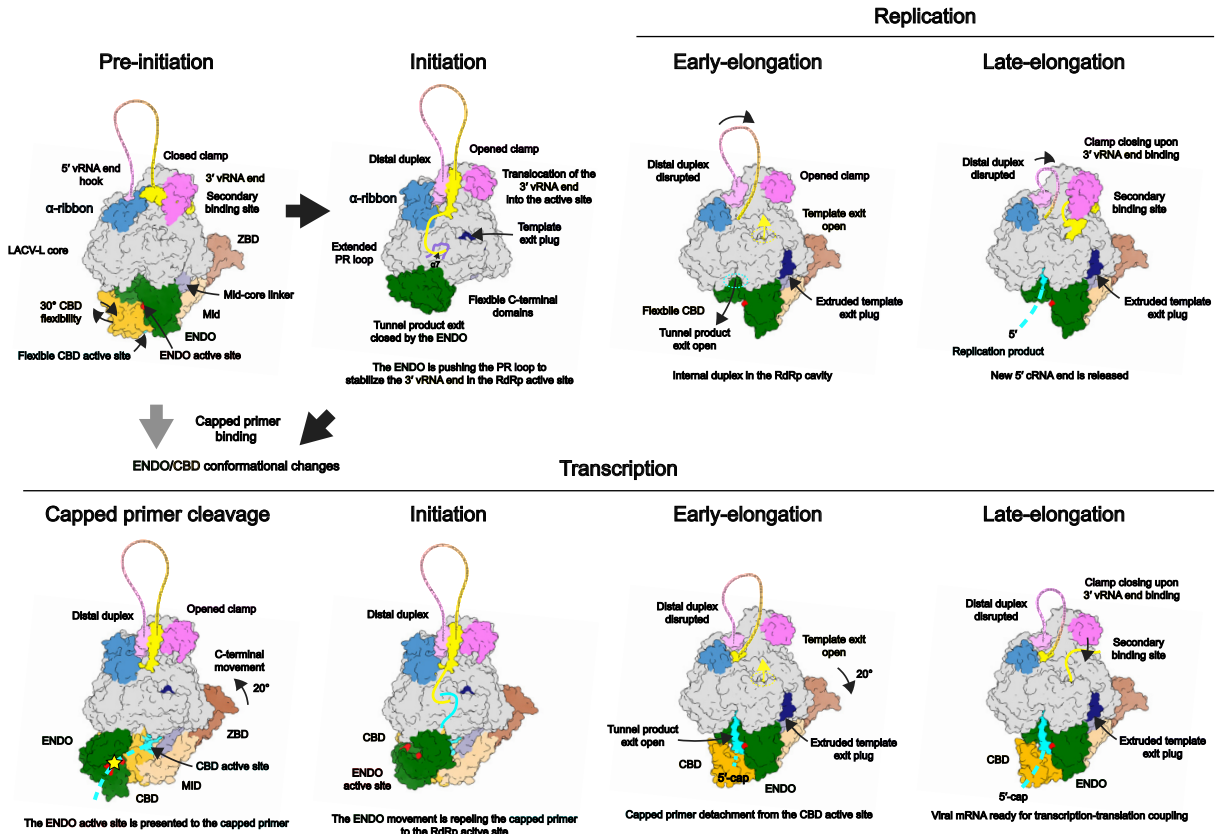
b, Capped primer path from the CBD active site to the RdRp active site during transcription initiation. On the right panel, residues that are proximal to the capped primer are indicated.

c, In vitro transcription activity of LACV-L_{Clitag_H34K}, LACV-L_{Clitag_H34K-M989A} (M989A), LACV-L_{Clitag_H34K-I990A} (I990A), LACV-L_{Clitag_H34K-S991A} (S991A) with 5'-1-17BPm, cap14AG, 3'-vRNA1-25 and four NTPs. Transcription products are labeled. Cap37-mer and cap37-mer T1 lanes are equivalent to **Fig. 4a** description.

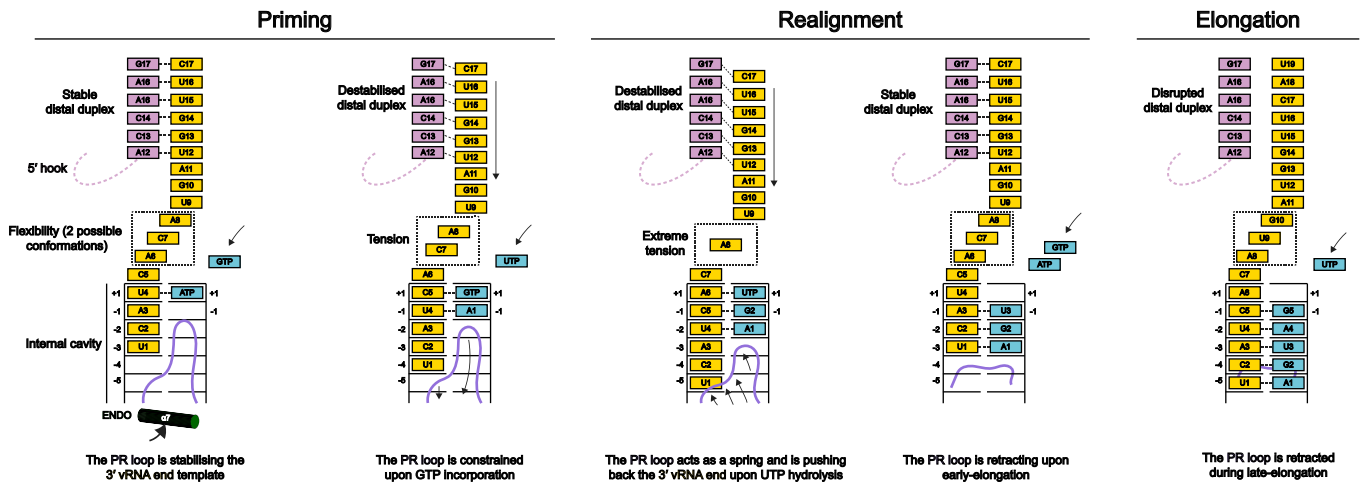
d, Interaction between the capped RNA product and LACV-L_{Clitag_H34K} in transcription elongation state. Residues stabilizing it are indicated. Flexible nucleotides are represented as a dotted line.

FIGURE 7

a



b



c

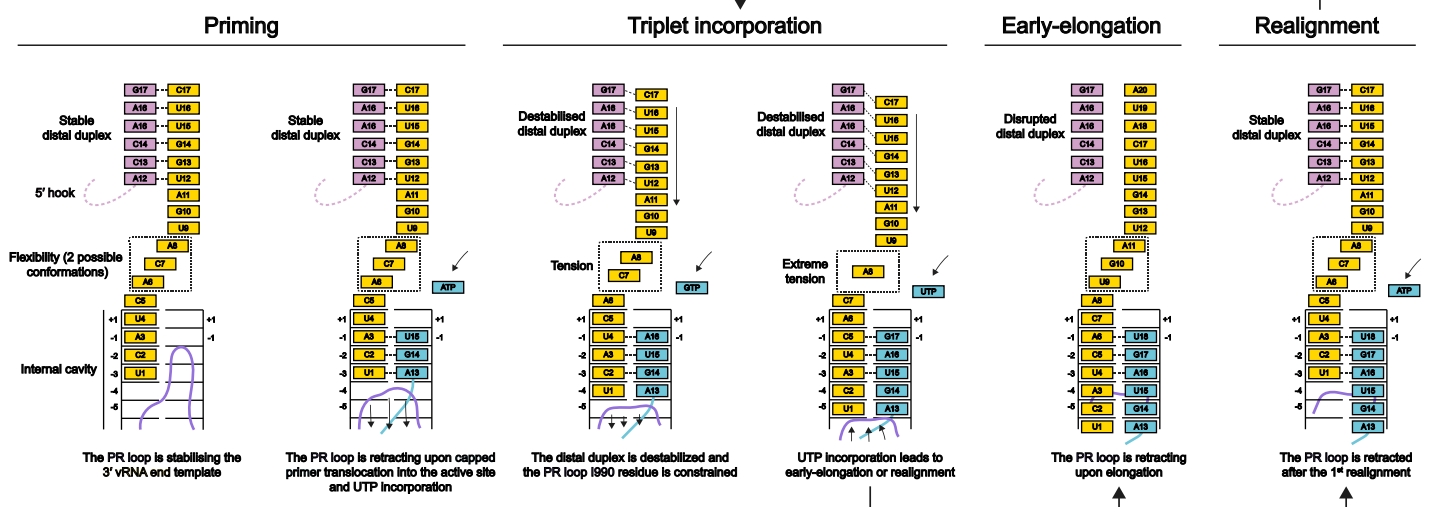


Fig. 7. LACV-L replication and transcription cycle models

a, Structure-based model of the mechanisms underlying LACV-L replication and transcription. All structures are displayed as surfaces with specific domains and features colored as in Fig. 1 and 2. The vRNA is shown as a line with the 5'/3' ends colored in pink/yellow. At initiation, the 3'-vRNA terminus is released from the secondary binding site and brought into the RdRp active site where it is stabilized by the PR loop, that extends due to the ENDO movement. If the flexible CBD snatches a 5'-cap, important conformational changes of the CBD, the ENDO, the mid and the ZBD occur, triggering capped primer translocation into the ENDO active site (in red). Once cleaved (yellow star), an ENDO movement triggers the capped primer repositioning towards the RdRp active site. Early elongation induces extrusion of the template exit plug resulting in the template exit opening (yellow dotted circle). In late-elongation, the 3'-vRNA template end goes back to the secondary binding site.

b, c, Structure-based model of initiation by prime-and-realign for replication (b) and transcription (c). The 5'/3'-vRNA ends are respectively colored in pink and gold. The 5'-hook structure is represented as a dotted line. Incorporated nucleotides are colored in blue. The PR loop is shown in pink. The binding pocket of nucleotides A6, C7 and A8 at priming stage is surrounded by a dotted square. The proposed successive steps of the models are presented from left to right. The main role of each step is indicated below each schematic.

SUPPLEMENTARY INFORMATIONS

Structural snapshots of La Crosse virus polymerase reveal the mechanisms underlying *Peribunyaviridae* replication and transcription

Benoît Arragain *et al.*

Supplementary Table 1. Cryo-EM data collection, refinement and validation statistics of LACV-L replication snapshots

	LACV L Replication initiation	LACV L Replication early elongation	LACV L Replication late-elongation
	(PDB 7ORN) (EMD-13043)	(PDB 7ORO) (EMD-13044)	(PDB 7ORI) (EMD-13038)
Data collection and processing			
Microscope	Thermo Fisher Scientific Titan Krios	Thermo Fisher Scientific Glacios	Thermo Fisher Scientific Glacios
Camera	Gatan K3	Gatan K2 Summit	Gatan K2 Summit
Magnification	130000	36000	36000
Voltage (kV)	300	200	200
Number of frames	40	60	60
Electron exposure (e ⁻ /Å ²)	50	60	60
Defocus range (µm)	-0.8 to -1.8	-0.8 to -2.0	-0.8 to -2.0
Pixel size (Å)	0.645	1.145	1.145
Symmetry imposed	C1	C1	C1
Initial/Final micrographs (no.)	15573/15444	2927/2923	1848/1841
Final particles (no.)	641008	417757	24151
Map resolution (Å) 0.143	2.8	2.9	3.9
FSC threshold			
Map resolution range (Å)	2.6-3.5	2.7-4	3.6-6
Refinement			
Initial model used	6Z6G	6Z8K	6Z8K
Model resolution (Å) 0.5 FSC threshold	2.8	3.0	3.9
Map sharpening B factor (Å ²)	-50	-50	-50
Model composition			
Protein residues	1700	2008	2007
Nucleotide residues	35	48	44
Ligands	3	2	3
Water	78	0	0
B-factor (Å ² , min-max (mean))			
Protein	38.17-142.20 (64.15)	37.78-160.62 (74.97)	68.43-192.36 (109.75)
Nucleotides	44.49-108.37 (67.92)	51.26-142.63 (75.16)	97.23-188.85 (112.03)
Ligands	51.03-53.69 (51.85)	46.51-133.86 (90.19)	88.19-172.55 (112.03)
R.m.s deviations			
Bond lengths (Å)	0.003	0.004	0.003
Bond angles (°)	0.508	0.531	0.501
Validation			
MolProbity score	1.37	1.52	1.69
Clashscore	5.27	6.40	9.32
Poor rotamers (%)	0.06	0.05	0.00
Ramachandran plot			
Favored (%)	97.57	97.04	96.79
Allowed (%)	2.43	2.96	3.21
Disallowed (%)	0	0	0

Supplementary Table 2. Cryo-EM data collection, refinement and validation statistics of LACV-L transcription snapshots

	LACV L Transcription cleavage conformation	LACV L Capped primer active site entry	LACV L Transcription initiation	LACV L Transcription early elongation
	(PDB 7ORJ) (EMD-13039)	(PDB 7ORK) (EMD-13040)	(PDB 7ORL) (EMD-13041)	(PDB 7ORM) (EMD-13042)
Data collection and processing				
Microscope	Thermo Fisher Scientific Glacios	Thermo Fisher Scientific Titan Krios	Thermo Fisher Scientific Titan Krios	Thermo Fisher Scientific Glacios
Camera	Gatan K2 Summit	Gatan K3	Gatan K3	Gatan K2 Summit
Magnification	36000	130000	130000	36000
Voltage (kV)	200	300	300	200
Number of frames	60	40	40	60
Electron exposure (e ⁻ /Å ²)	60	50	50	60
Defocus range (µm)	-0.8 to -2.0	-0.8 to -1.8	-0.8 to -1.8	-0.8 to -2.0
Pixel size (Å)	1.145	0.645	0.645	1.145
Symmetry imposed	C1	C1	C1	C1
Initial/Final micrographs (no.)	3270/3149	15573/15444	15573/15444	2524/2363
Final particles (no.)	29065	76579	19112	229480
Map resolution (Å) 0.143 FSC threshold	3.9	3.1	3.6	3.3
Map resolution range (Å)	3.6-6.0	2.8-5	3.3-5	3.1-5.5
Refinement				
Initial model used	6Z6G	6Z6G	6Z6G	6Z8K
Model resolution (Å) 0.5 FSC threshold	3.9	3.0	3.6	3.4
Map sharpening B factor (Å ²)	-60	-50	-50	-80
Model composition				
Protein residues	2129	2182	2183	2145
Nucleotide residues	36	37	47	40
Ligands	2	4	3	2
Water	0	48	0	0
B-factor (Å ² , min-max (mean))				
Protein	18.86-240.60 (60.25)	12.50-164.40 (60.74)	54.53-210.52 (97.25)	19.07-204.31 (61.26)
Nucleotides	29.50-93.79 (55.70)	34.94-154.19 (66.57)	76.53-172.59 (124.10)	29.40-152.41 (66.03)
Ligands	37.90-99.40 (49.45)	31.21-140.20 (49.15)	63.41-98.72 (97.56)	25.00-112.91 (68.95)
R.m.s deviations				
Bond lengths (Å)	0.003	0.004	0.002	0.003
Bond angles (°)	0.579	0.533	0.560	0.533
Validation				
MolProbity score	1.72	1.59	1.72	1.61
Clashscore	9.89	6.94	6.48	7.63
Poor rotamers (%)	0.00	0.01	0.00	0.00
Ramachandran plot				
Favored (%)	96.73	96.73	96.18	96.85
Allowed (%)	3.27	3.27	3.73	3.10
Disallowed (%)	0	0	0	0

Supplementary Table 3. Search patterns used to calculate the read numbers in Supplementary Fig. 2

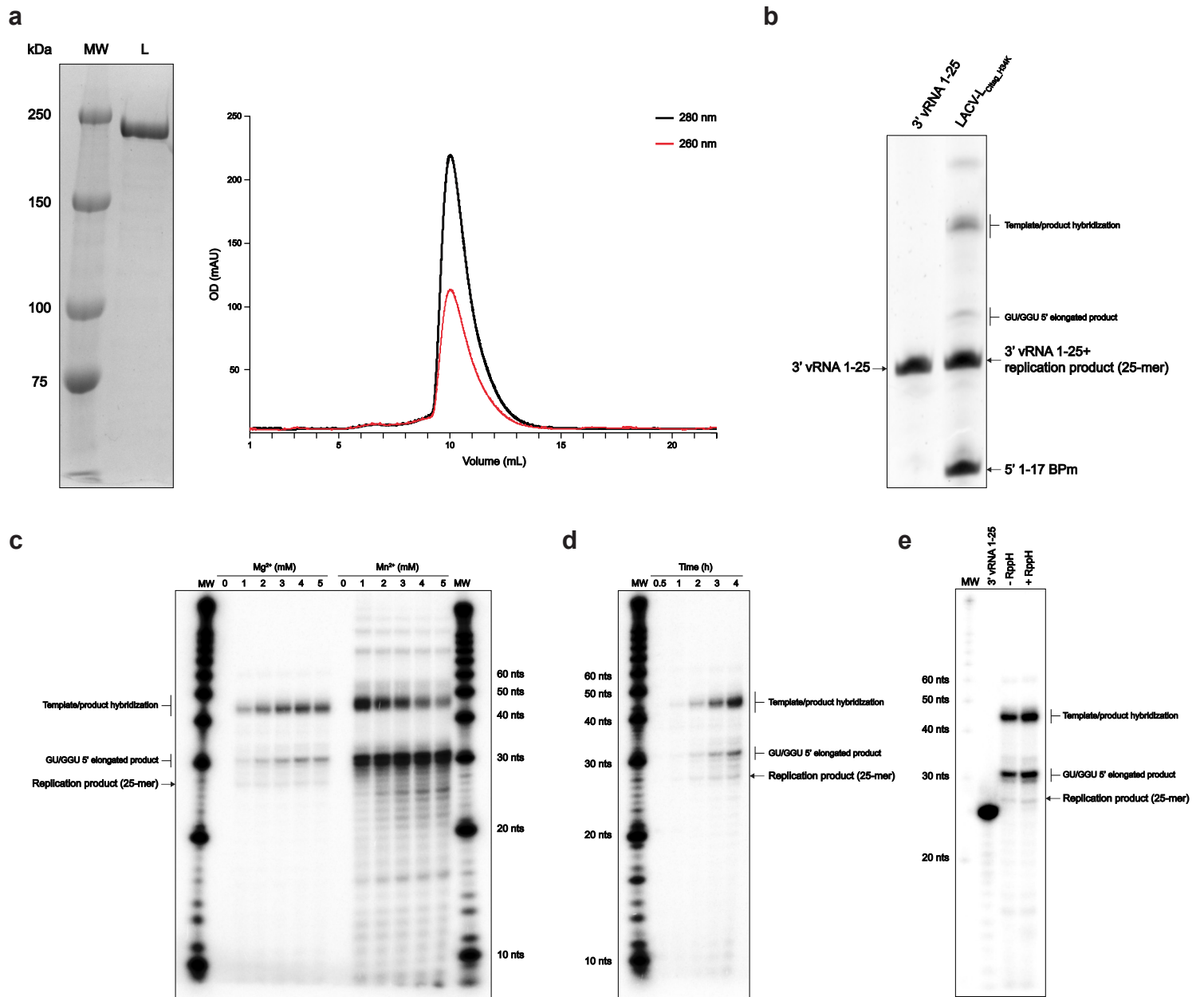
Replication

Type	Remarks	Search Settings	Search pattern
Reads containing the replication products	filter for length between 22-32	-k 4s	AGTAGTGTACTACCAAGTATAGATA
Replication Products		-k 0	^AGTAGTGTA
GGU/GU 5' extended replication products		-k 0	^(GT GGT)AGTAGTGTA
Replication products with a G mismatch in 5'		-k 0	^GGTAGTGT

Transcription

Type	Remarks	Search Settings	Search pattern
Transcript, no realignment		-k 4s	"^GGAATGCTATAATAGTAGTGTACTACCAAGTATAGATA"; "ATA.{0,6}\$"
Transcript, one realignment	three rounds of search and a length filter >40 <48	-k 4s; NA ; -k 0	"^GGAATGCTATAATAGTAGTAGTGTACTACCAAGTATAGATA"; "ATA.{0,6}\$"; ".+AGTAGTAGTGTA.+"
Transcript, two realignments	three rounds of search and a length filter >40	-k 4s; NA ; -k 0	"^GGAATGCTATAATAGTAGTAGTAGTGTACTACCAAGTATAGATA"; "ATA.{0,6}\$"; ".+AGTAGTAGTAGTGTA.+"
Transcript, three realignments	three rounds of search and a length filter >40	-k 4s; NA ; -k 0	"^GGAATGCTATAATAGTAGTAGTAGTAGTGTACTACCAAGTATAGATA"; "ATA.{0,6}\$"; ".+AGTAGTAGTAGTAGTGTA.+"
Rest	length filter >36 <41	NA	NA

SUPPLEMENTARY FIGURE 1



Supplementary Fig. 1. LACV-L_{C1tag_H34K} purification and replication activity optimization

a, 6% SDS-PAGE gel of LACV-L_{C1tag_H34K} (MW: Molecular weight; L: LACV-L_{C1tag_H34K}) and the corresponding gel filtration profile. Absorbance curves at 280 and 260 nm are indicated and respectively colored in black and red.

b, Urea-PAGE gel stained with SybrGold showing the 3' vRNA 1-25 next to LACV-L_{C1tag_H34K} replication assay in presence of 3' vRNA 1-25, 5'-1-17 BPm and 4 nucleotides.

c, Optimization of the divalent metal ion concentration for LACV-L_{C1tag_H34K} replication activity. Assessment of replication product formation using concentration ranging from 0 to 5 mM of MgCl₂ or MnCl₂. The molecular weight marker (MW) corresponds to the decade marker.

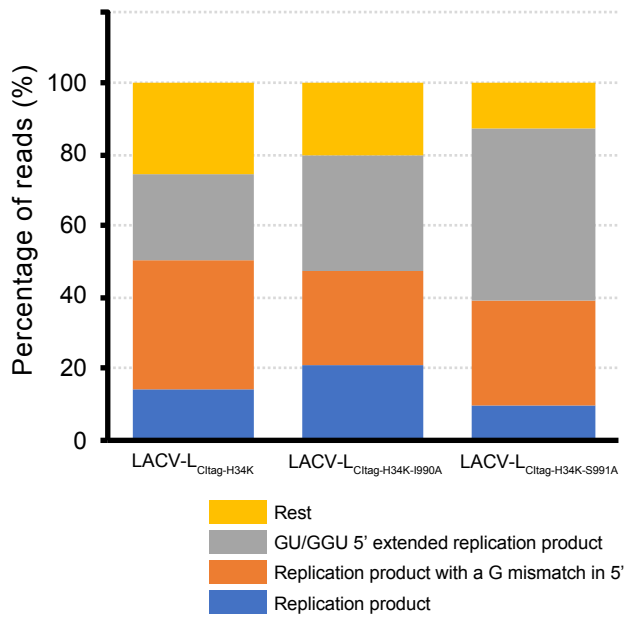
d, Time course of LACV-L_{C1tag_H34K} replication activity. Reactions were stopped after 0.5, 1, 2, 3 or 4h.

e, LACV-L replication reaction products incubated or not with 5' Terminator exonuclease that specifically degrades 5' mono-phosphate RNA shown on a 20% Urea-PAGE gel.

SUPPLEMENTARY FIGURE 2

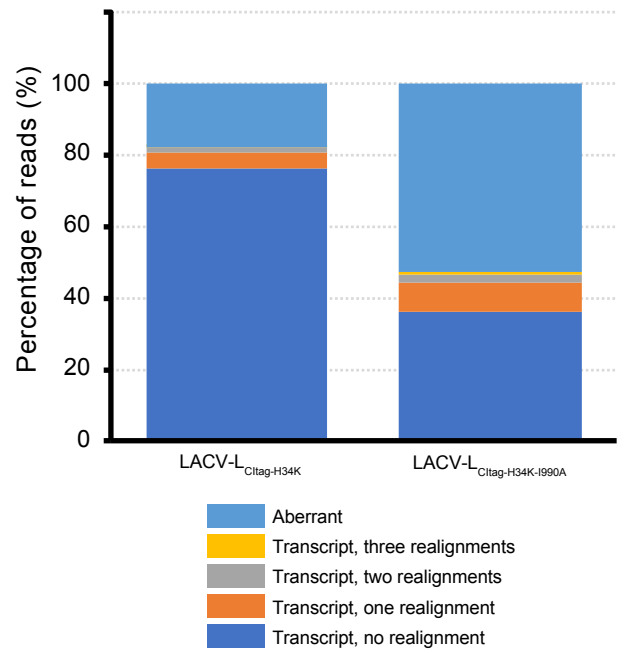
a

Replication



b

Transcription



Supplementary Fig. 2. RNA-sequencing

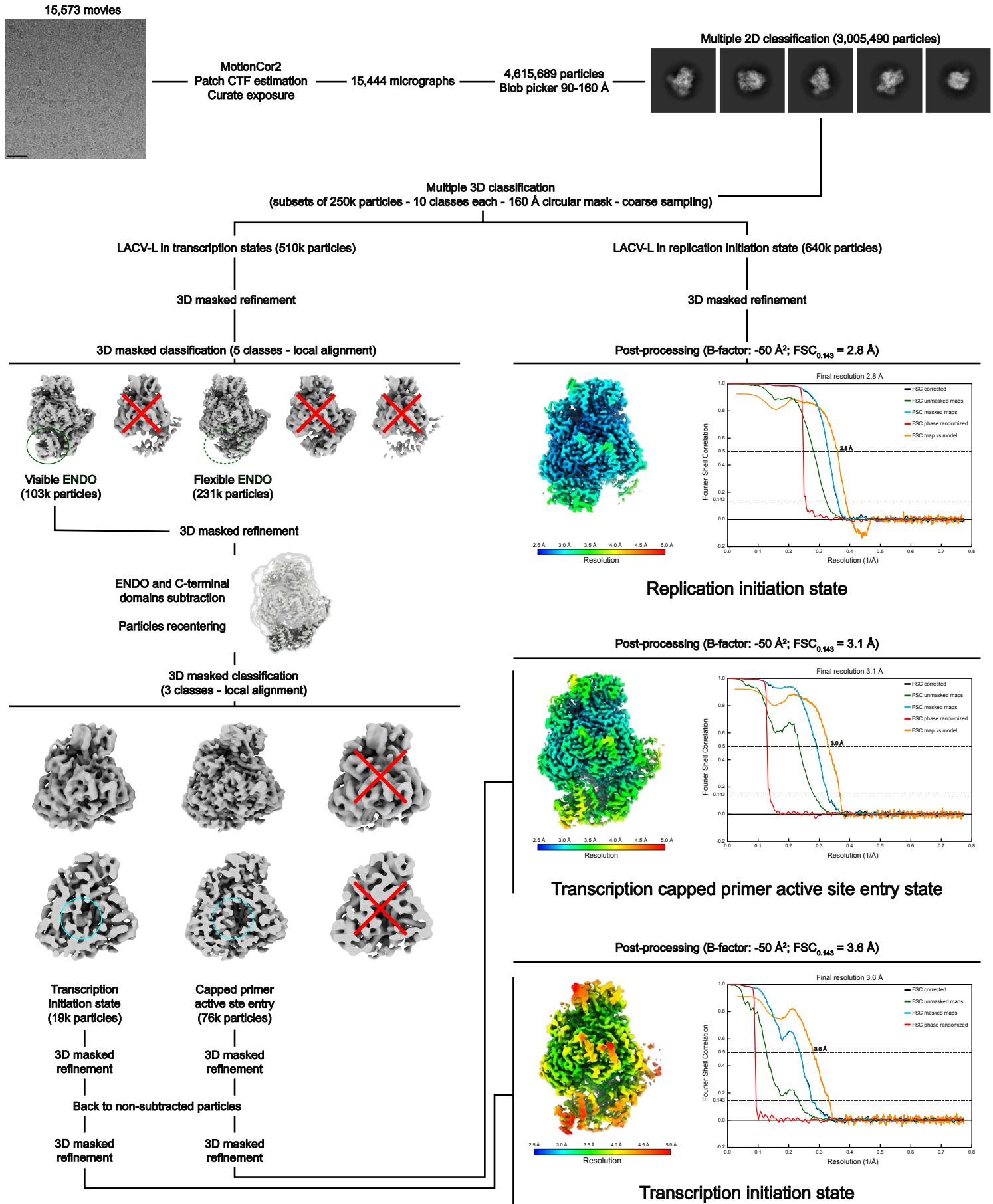
a, NGS analysis of reads containing the replication products for LACV-L_{Citag-H34K}, LACV-L_{Citag-H34K-I990A} and LACV-L_{Citag-H34K-S991A}. The percentage of read categories is shown. The replication products displayed in blue correspond to the expected 25-mer products. Replication products where the expected 5' terminal nucleotide is replaced by G are displayed in orange. Replication products that are extended in 5' by GU or GGU sequences are shown in gray. The other reads that contain the replication products but do not fall in the above categories are shown in green.

b, NGS analysis of reads containing the capped primer with a length compatible with transcription products. The percentage of read categories is shown. The 37-mer transcripts containing the capped primer and the expected product are shown in blue. The realigned transcripts that display one, two and three AGU nucleotide triplet(s) inserted between the primer and the product are shown in orange, gray and yellow, respectively. The aberrant products containing the capped primer and have a length compatible with transcription products (with or without realignment) but do not contain the full product are shown in light blue. Most of these latter reads correspond to the capped primer followed by multiple AGU repeats likely obtained by successive realignments that did not lead to product formation.

SUPPLEMENTARY FIGURE 3

cryoSPARC v3.0

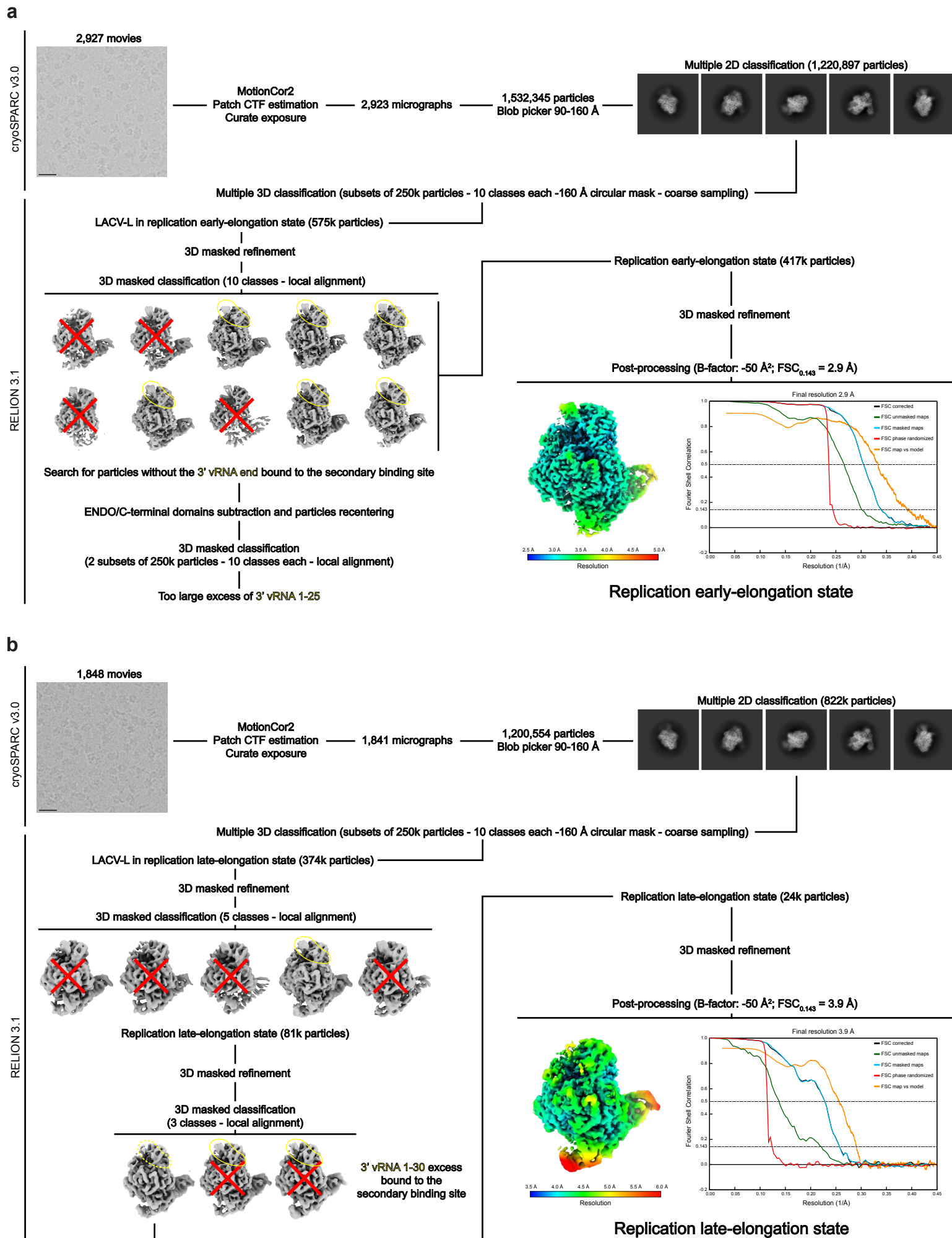
RELION 3.1



Supplementary Fig. 3. Image processing strategy to obtain the replication initiation state, the transcription capped primer active site entry state and the transcription initiation state

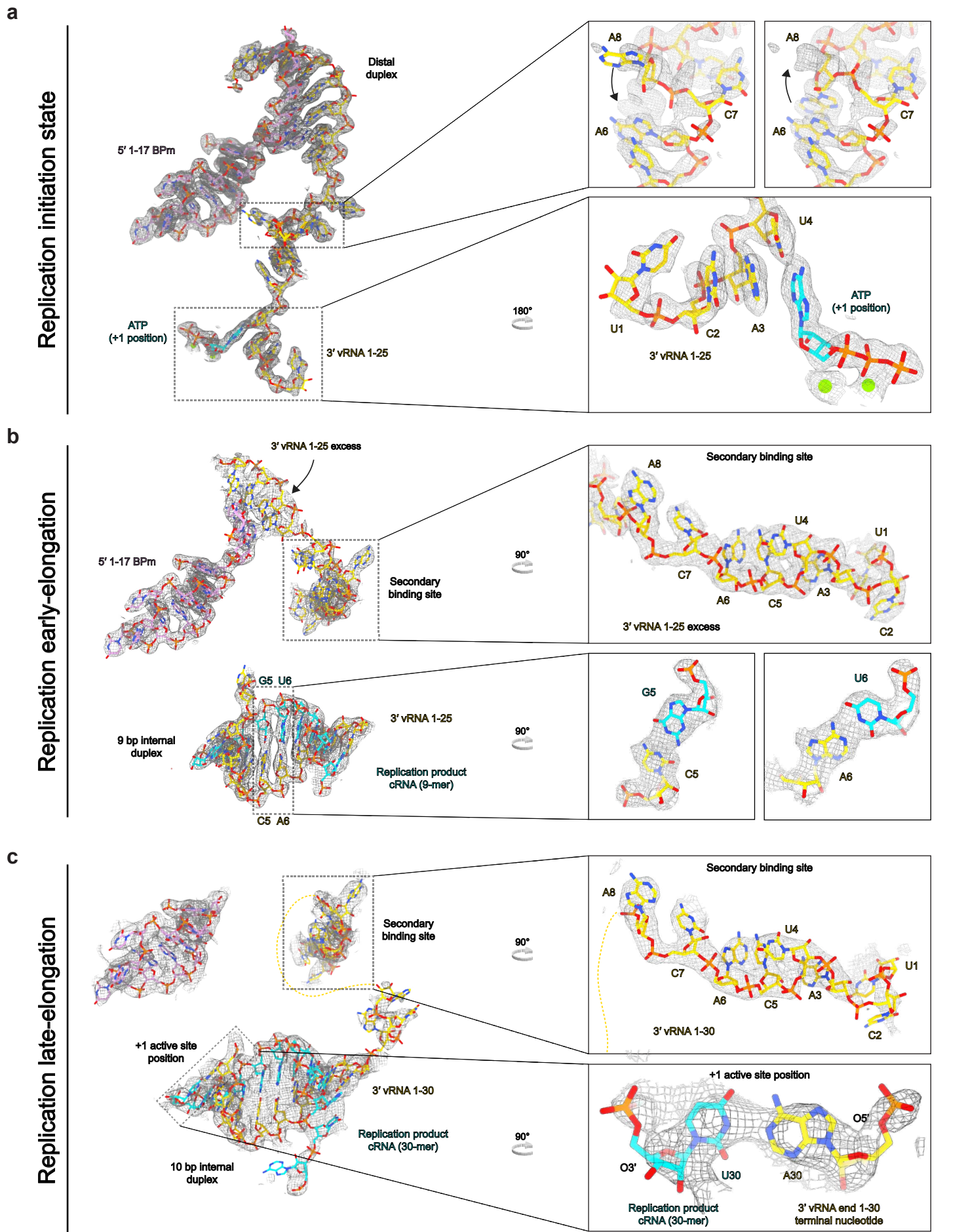
Schematics of the image processing strategy used with the data collected on a Titan Krios equipped a K3 direct electron detector. Representative micrograph, 2D class averages, 3D class averages are displayed. Local resolution EM maps colored according to resolution are shown. Fourier shell correlation curves are displayed. Scale bar = 200 Å.

SUPPLEMENTARY FIGURE 4



Supplementary Fig. 4. Image processing strategy to obtain the replication early-elongation state and the replication late-elongation state
a, b, Schematics of the image processing strategy used with the data collected on a Glacios cryo-TEM equipped a K2 direct electron detector to obtain the replication early-elongation state (**a**) and the replication late-elongation state (**b**). Representative micrographs, 2D class averages, 3D class averages are displayed. Local resolution EM maps colored according to resolution are shown. Fourier shell correlation curves are displayed. Scale bar = 200 Å.

SUPPLEMENTARY FIGURE 5



Supplementary Fig. 5. RNA densities and models in the different replication states.

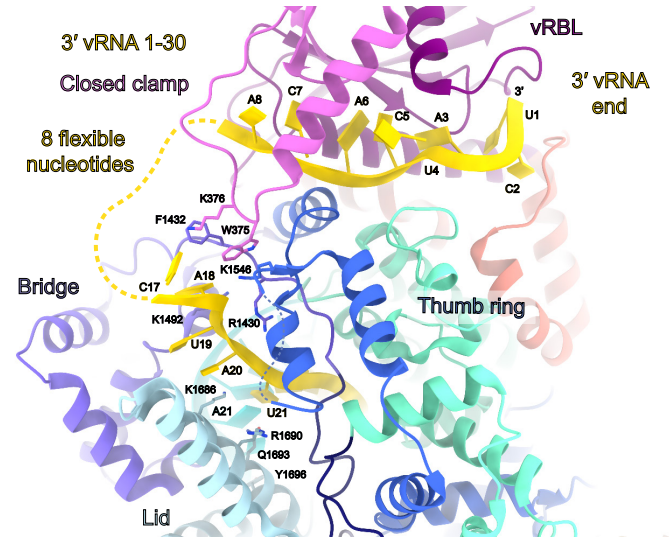
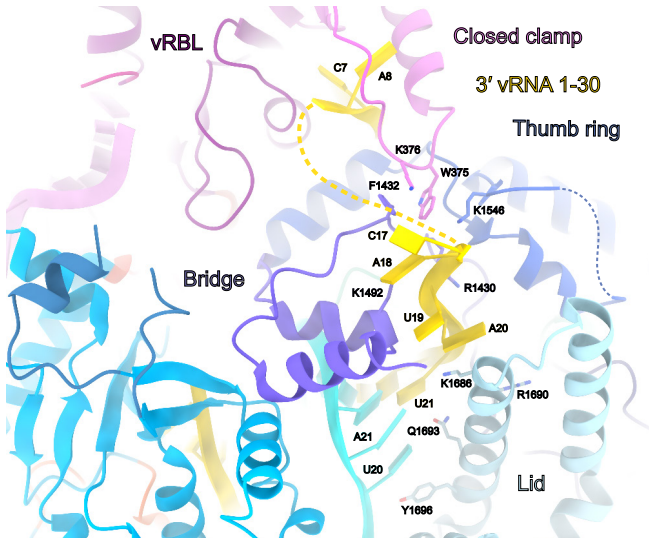
a, (left) General view of the RNA density and model in the replication initiation state. (right, top) Focused view of nucleotides A6, C7, A8 densities in the replication initiation state. Alternative conformations are visible in the density. A8 is built in one conformation in each zoom. An arrow indicates A8 movement. (right, bottom) Focused view of the RNA density focusing on the ATP, the nucleotides and the magnesium ions present in the active site.

b, (left) General view of the RNA density and model at replication early-elongation. (right, top) Focused view on the 3' vRNA excess that binds in the 3' secondary binding site. (right, bottom) Focused view on some nucleotides of the RNA template/product duplex.

c, (left) General view of the RNA density and model at replication late-elongation. (right, top) Zoom of the template 3' end that binds to the 3' secondary binding site. (right, bottom) Zoom on the +1 active site position that identifies the template and product terminal nucleotides.

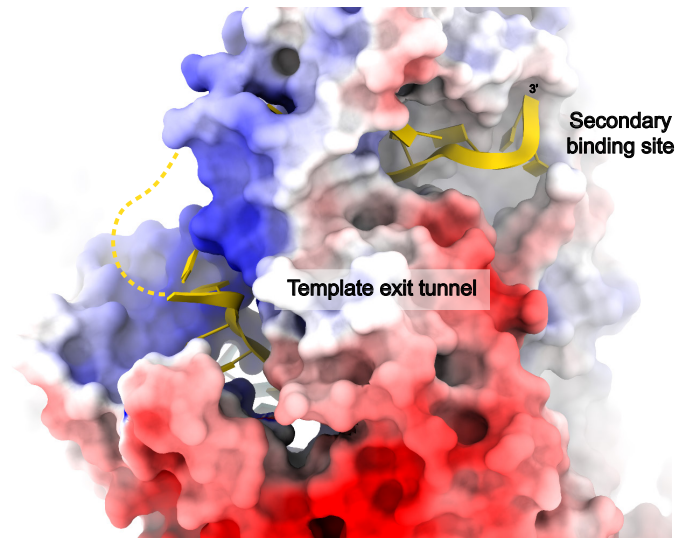
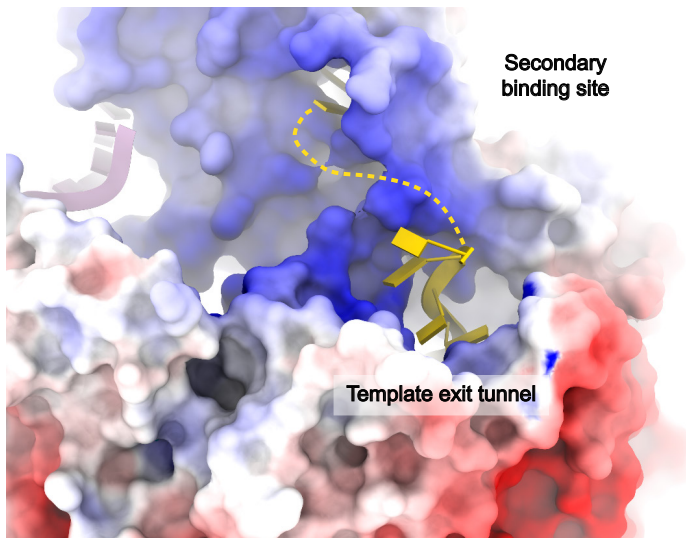
SUPPLEMENTARY FIGURE 6

a



50°

b



Electrostatic potential (kT)

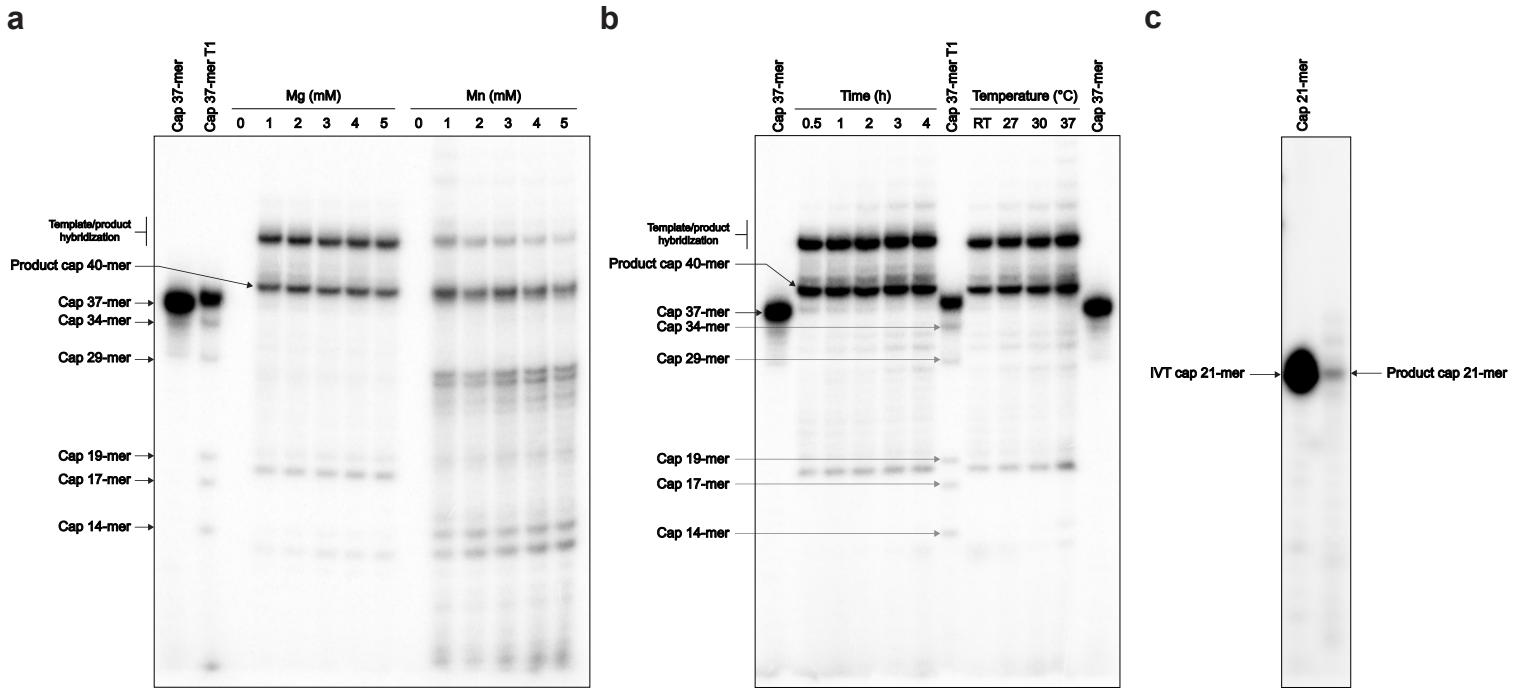


Supplementary Fig. 6. Path of the template RNA from the template exit channel to the 3'-vRNA secondary binding site

a, The 3'-vRNA is displayed in yellow. Interacting residues in the template exit channel are indicated.

b, LACV-L_{Clitag_H34K} surface colored according to its electrostatic potential showing that the path from the template exit channel towards the secondary binding site is positively charged.

SUPPLEMENTARY FIGURE 7



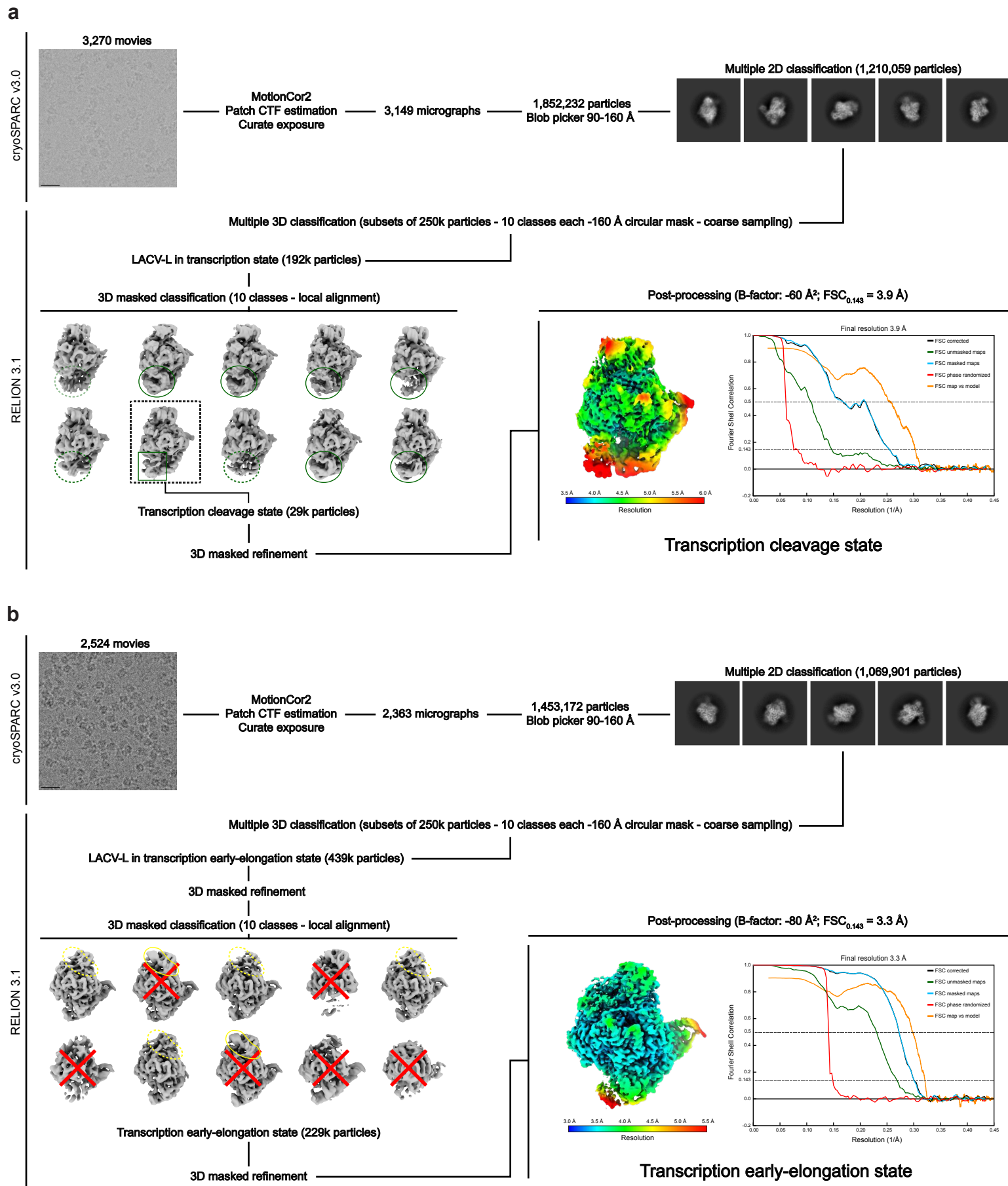
Supplementary Fig. 7. Transcription activity optimization of LACV-L_{C1tag_H34K}

a, Optimization of the divalent metal ion concentration for LACV-L_{C1tag_H34K} transcription activity. Assessment of transcription product formation using concentration ranging from 0 to 5 mM of MgCl₂ or MnCl₂. The Cap37-mer lane corresponds to a capped 37-mer RNA identical in sequence to the theoretical transcription product of LACV-L_{C1tag_H34K} with the cap14AG and the 3'-vRNA1-25 in the absence of realignment. The lane indicated as Cap37-mer T1 corresponds to an RNase T1 cleavage of the cap37-mer.

b, Time course and temperature effect on LACV-L_{C1tag_H34K} transcription activity. Reactions were stopped after 0.5, 1, 2, 3 or 4h. Reactions were run at room temperature (RT), 27, 30 or 37°C.

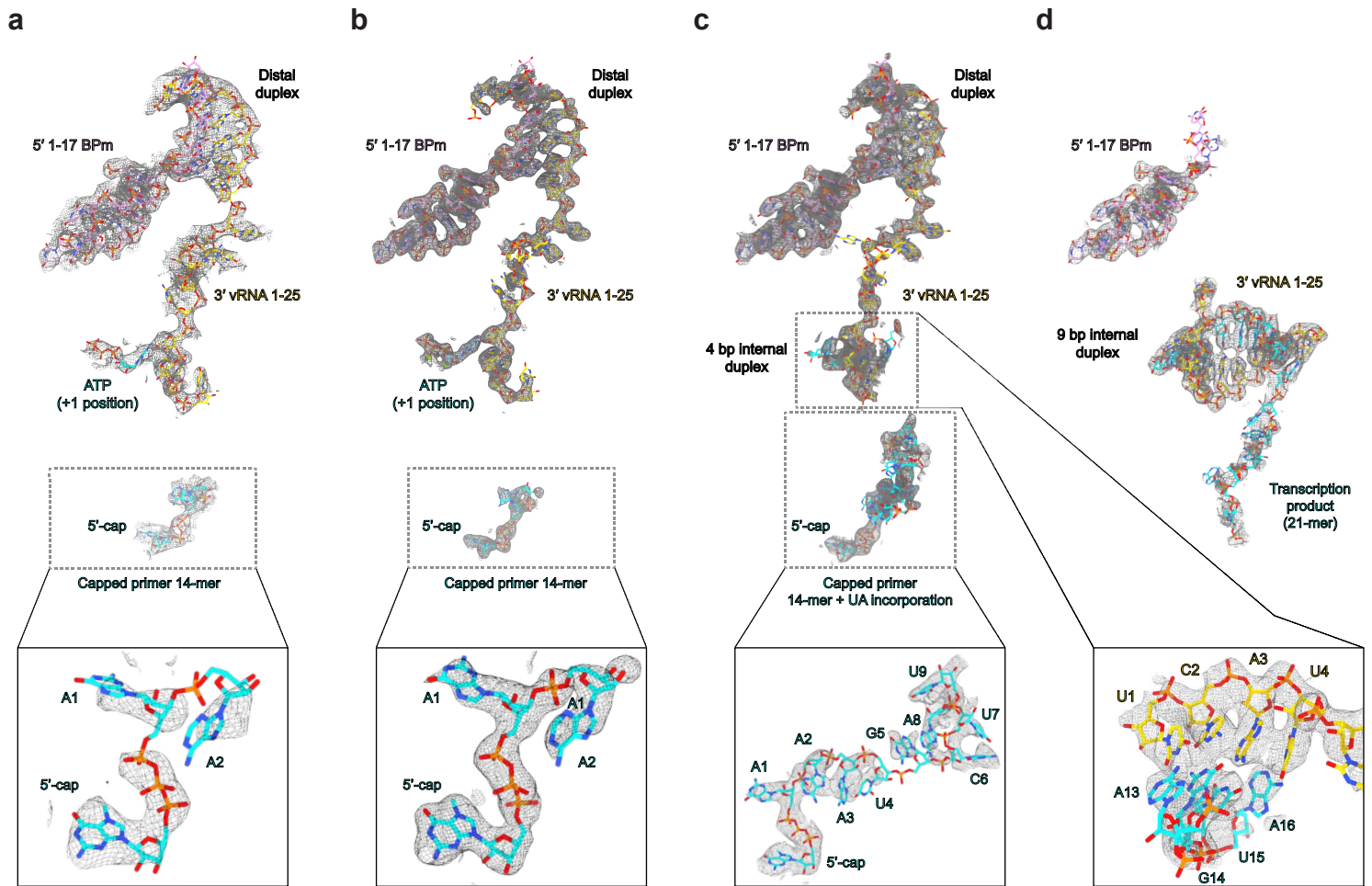
c, Transcription reaction of LACV-L when incubated with cap14AG, 3'-vRNA1-25, 5'-1-17Bpm and ATP/GTP/UTP. The molecular weight marker is a capped RNA with a sequence identical to the expected LACV-L un-realigned transcript product.

SUPPLEMENTARY FIGURE 8



Supplementary Fig. 8. Image processing strategy to obtain the transcription capped primer cleavage state and the transcription early-elongation state a,b. Schematics of the image processing strategy used with the data collected on a Glacios cryo-TEM equipped with a K2 direct electron detector to obtain the transcription capped primer cleavage state (a) and the transcription early-elongation state (b). Representative micrographs, 2D class averages, 3D class averages are displayed. Local resolution EM maps colored according to resolution are shown. Fourier shell correlation curves are displayed. Scale bar = 200 Å.

SUPPLEMENTARY FIGURE 9



Supplementary Fig. 9. RNA densities and models in the different transcription states.

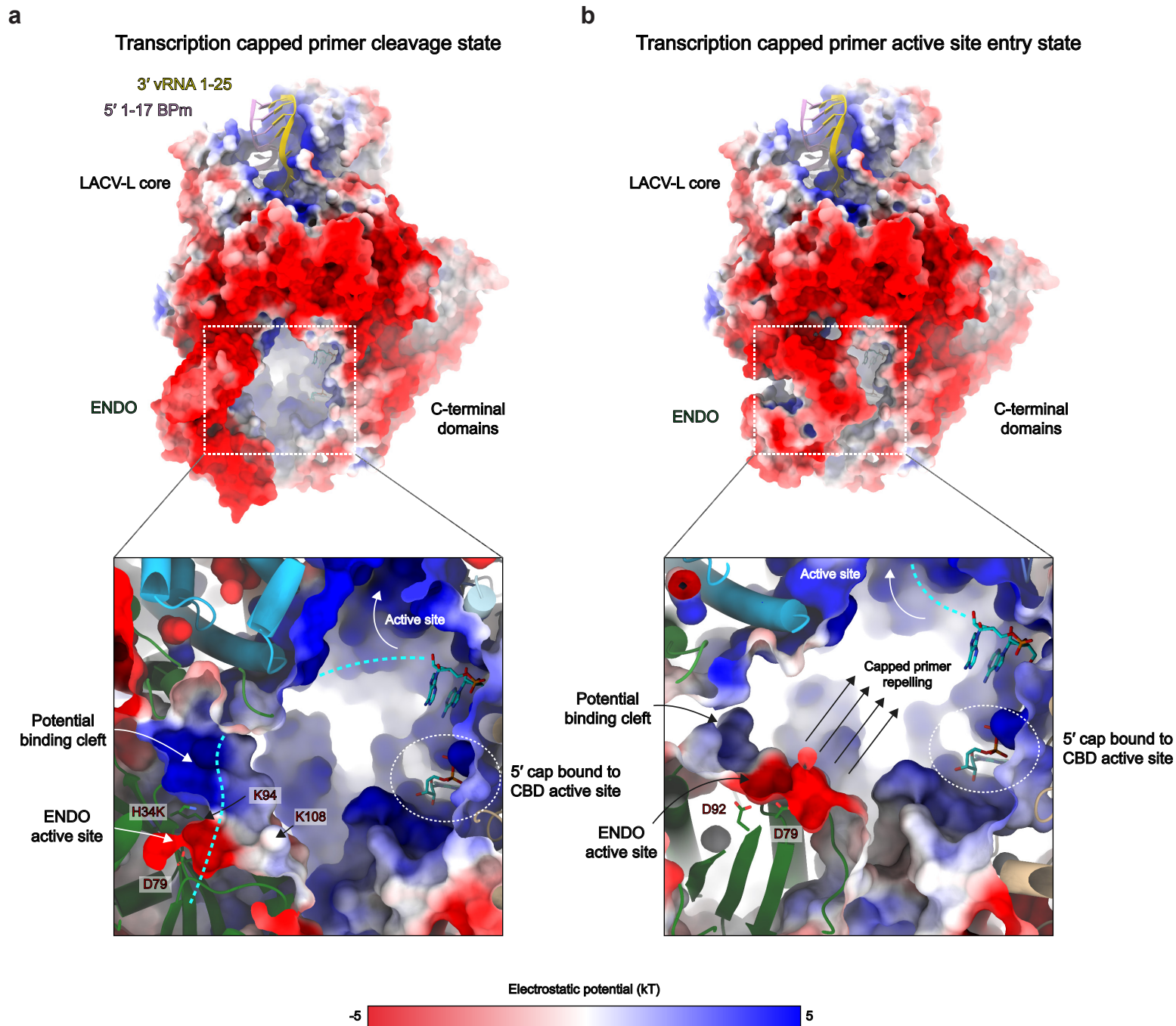
a, General view of the RNA density and model in the putative endonuclease cleavage conformation. Additional zoom on both capped RNA density and model is shown.

b, General view of the RNA density and model in the capped primer active site entry conformation. Additional zoom on both capped RNA density and model is shown.

c, General view of the RNA density and model in the transcription initiation conformation. Additional zoom on the capped RNA density, the 4 base-pair internal duplex density and their corresponding models are shown.

d, General view of the RNA density and model in the transcription early-elongation conformation.

SUPPLEMENTARY FIGURE 10



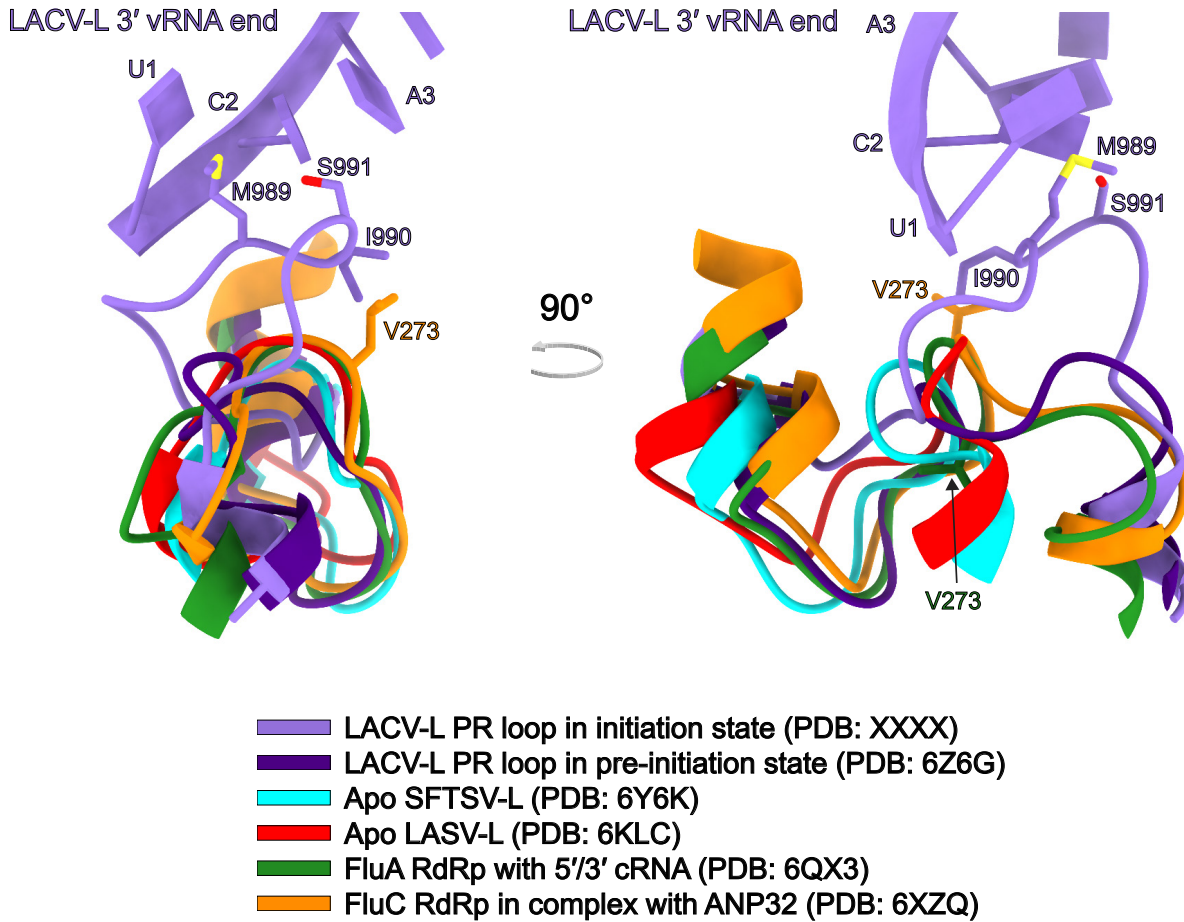
Supplementary Fig. 10. Electrostatic surface analysis of LACV-L_{C1tag_H34K} in the transcription capped primer cleavage state and the transcription capped primer active site entry state conformations

a, Electrostatic surface of the capped primer cleavage conformation. Top: global view, bottom: slabbed and zoomed view focusing on the charged environment close to the capped RNA. 5'-1-17Bpm and 3'-vRNA-1-25 are displayed and respectively colored in pink and gold. RNAs tunnels are positively charged. The ENDO is presenting a potential binding cleft for the capped primer (dotted line in cyan) surrounded by K94 and K108 leading to the ENDO active site.

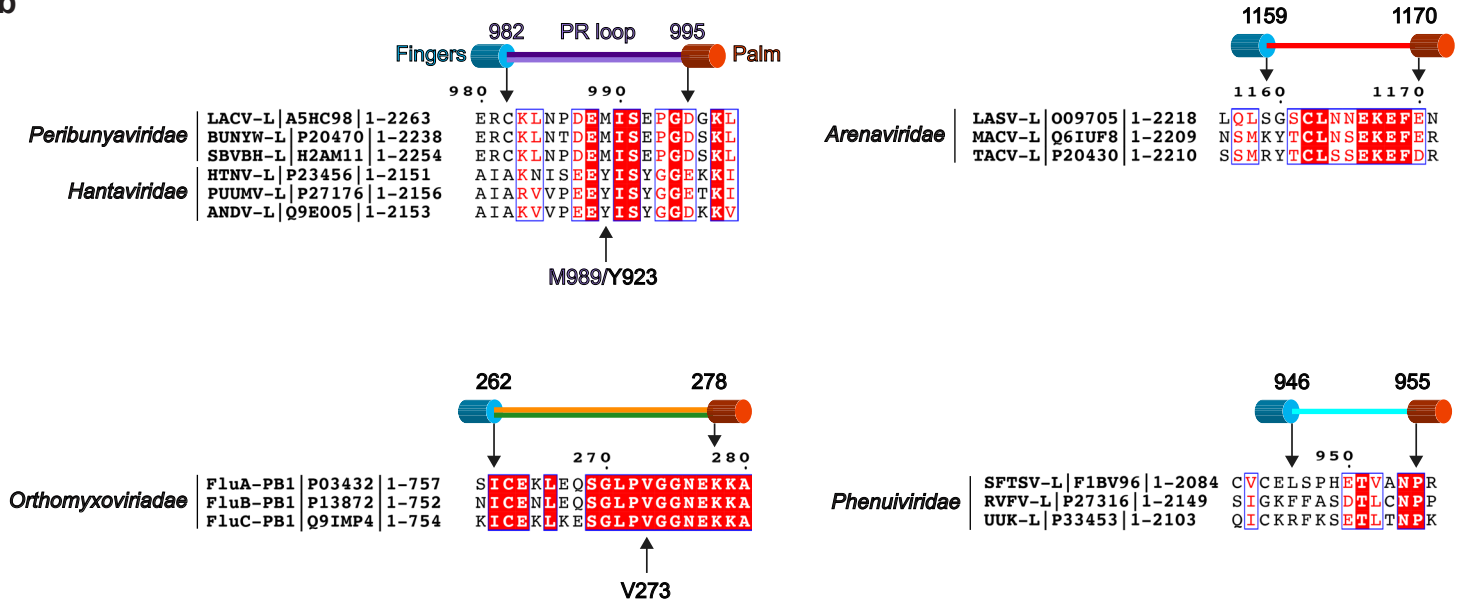
b, Electrostatic surface of the capped primer active site entry conformation as in (a). RNAs tunnels are positively charged. The ENDO active site is repelling the capped primer (dotted line in cyan) into the RdRp active site.

SUPPLEMENTARY FIGURE 11

a



b

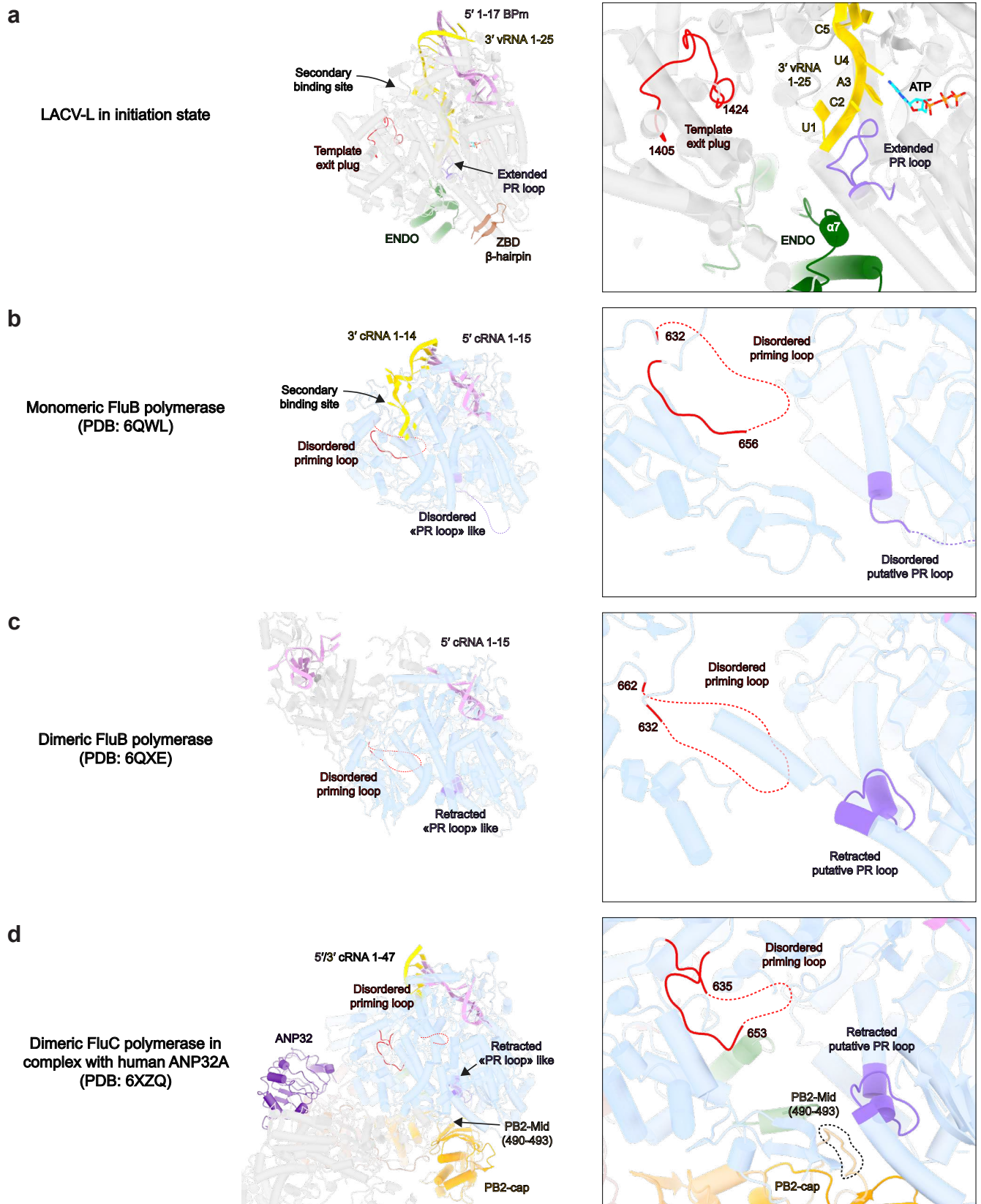


Supplementary Fig. 11. Putative Prime-and-realign loops in sNSV

a, Superposition of LACV-L PR loop in initiation state (purple), pre-initiation state (dark blue, PDB: 6Z6G), apo SFTSV-L (946-955) (light blue, PDB: 6Y6K), apo LASV-L (1159-1170) (red, PDB: 6KLC), FluA RdRp (262-278) in complex with 5'- and 3'-cRNA (green, PDB: 6QX3) and FluC RdRp (262-278) in complex with ANP32 (orange, PDB: 6XZQ). Interacting residues of the LACV-L PR loop with the 3'-vRNA end (M989, I990, S991) and the Flu RdRp (V273) are displayed.

b, Sequence alignment of the putative PR loop sequence in sNSV. *Peribunyaviridae* (LACV: La Crosse virus, BUNYW: Bunyamwera virus, SBVBH: Schmallenberg virus) and *Hantaviridae* (HTNV: Hantaan virus, PUUMV: Puumala virus, ANDV: Andes virus) families share some identity regarding LACV-L PR loop residues interacting with the 3'-vRNA template (E988, I990, S991). M989 in LACV-L is replaced by an aromatic residue (Y926) in *Hantaviridae* L protein that could play a similar role in template stabilization. Orthomyxoviruses (Influenza A, B, C) share a conserved PR loop on PB1 subunit with V273 playing an important role in prime-and-realign mechanism (Oymans and Te Velthuis, 2018). In *Arenaviridae* (LASV: Lassa virus, MACV: Machupo virus, TACV: Tacaribe virus), a conserved loop is also present between residues 1159 and 1170 contrary to *Phenuiviridae* (SFTSV: Sever Fever with Thrombocytopenia Syndrome virus, UUK: Uukuniemi virus) in which only T951 is conserved. For each displayed sequence, the virus name, the UniProt number and the viral polymerase length are indicated.

SUPPLEMENTARY FIGURE 12

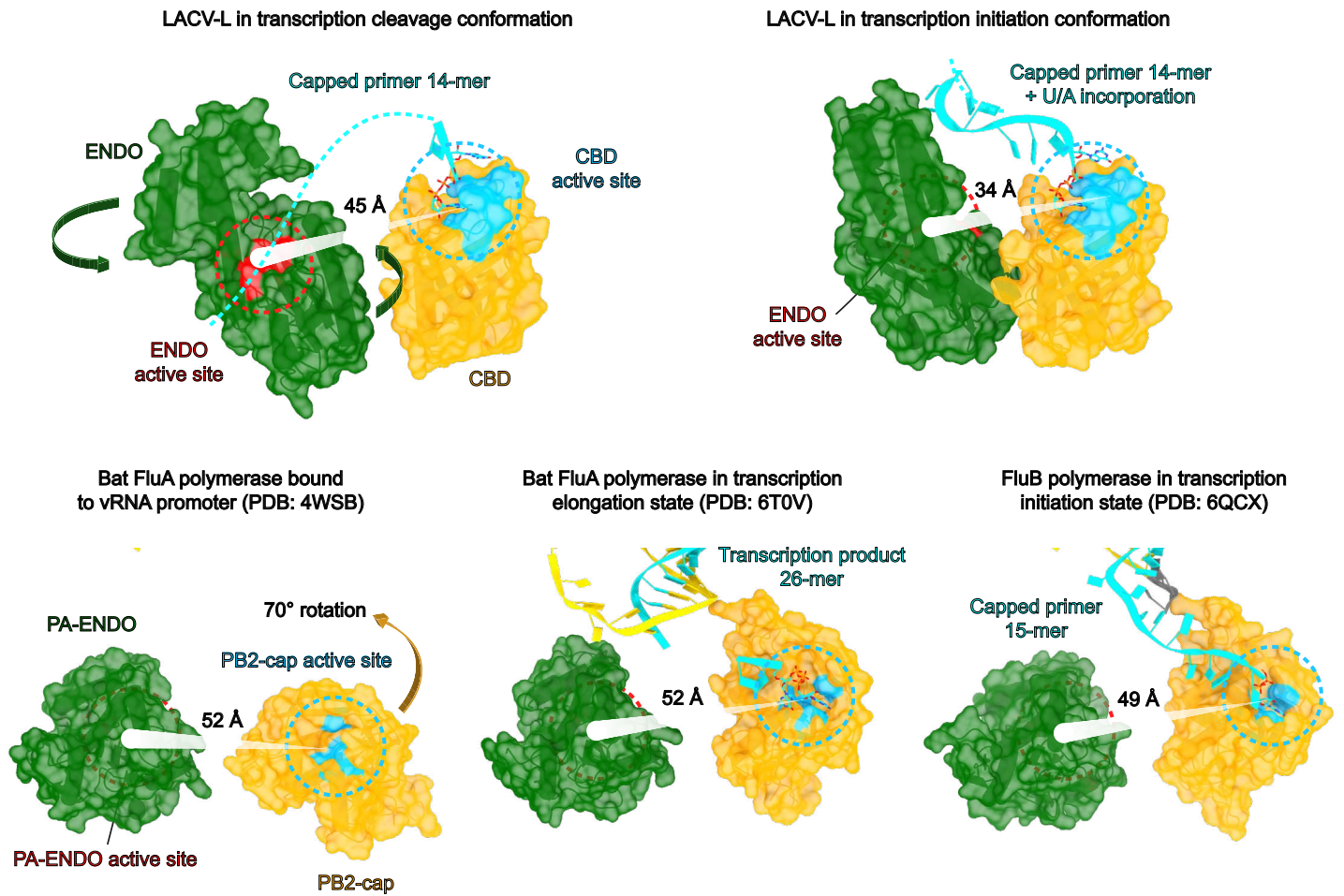


Supplementary Fig. 12. Prime-and-realign loops, LACV-L template exit plug and influenza polymerase priming loop

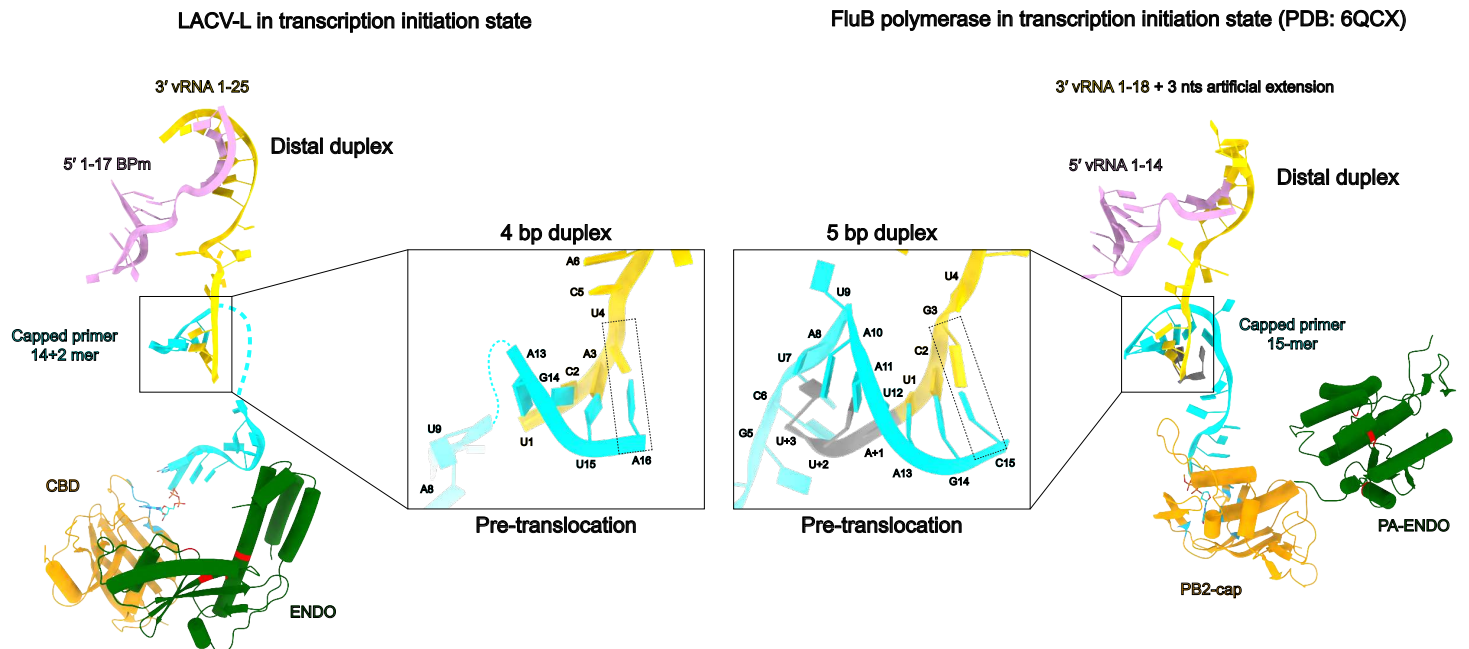
a-d. Cartoon representation of LACV-L in initiation state (a), monomeric influenza B polymerase bound to cRNA (PDB: 6QWL) (b), dimeric influenza B polymerase bound to cRNA (PDB: 6QXE) (c), dimeric influenza C polymerase in complex with human ANP32 (PDB: 6XZQ) (d). LACV-L core is transparent and colored in grey. Flu RdRp core is transparent and colored in blue. 3'-vRNA/cRNA end and 5'-vRNA/cRNA end present in each structure are respectively colored in gold and pink. The LACV-L template exit plug and the influenza priming loop are colored in red, if disordered represented as a dotted line. PR loops are colored in orange, the PB2-mid in beige, both the PB2-627 and the LACV-L ZBD β -hairpin in brown. On the right panels, close-up view of the polymerase internal cavity.

SUPPLEMENTARY FIGURE 13

a



b



Supplementary Fig. 13. ENDO and CBD positions in LACV-L and influenza polymerase during transcription initiation

a, Surface representation of the ENDO and the CBD of LACV-L (top) and influenza polymerase (bottom). ENDOS are colored in green with active sites in red. CBDs are colored in yellow with cap-binding active sites in cyan. Distances between ENDO active sites and cap-binding active sites are indicated. The capped primer is shown in cyan and its possible path as a dotted line. Rotations between the cleavage conformation and the transcription initiation conformation are indicated with arrows.

b, Comparison of the paths taken by the 3'-vRNA and the capped RNA primers between LACV-L (left) and influenza B polymerase (right). Both CBD and ENDO position are displayed. A zoom is made on the nascent template-product duplexes.

3.3. Discussion

L'ensemble de ces structures de polymérases, couplé aux tests d'activité *in vitro*, apportent une vision précise des différents changements de conformations que subit la LACV-L tout au long de la réplication, qui est initiée *de novo*, et de la transcription, qui est initiée à l'aide d'un ARN coiffé. Elles apportent également des informations cruciales sur la compréhension de l'initiation de chacun de ces deux mécanismes par amorçage et réalignement (*prime and realign*) et m'ont permis d'identifier et de révéler l'implication d'un nouvel élément structural : la *prime and realign loop* (*PR loop*) qui est également conservée chez la HTNV-L (Figure 50).

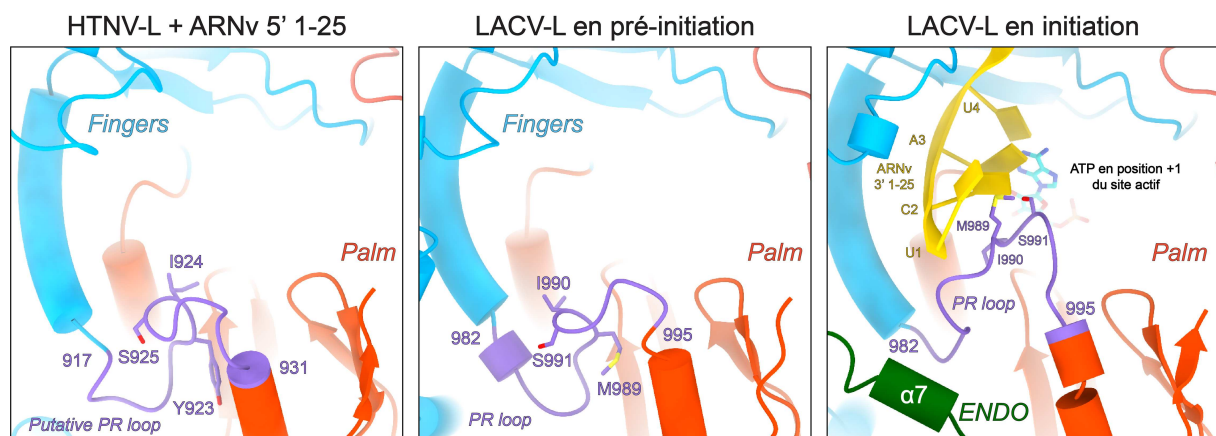


Figure 50. Comparaison de la *PR loop* entre la HTNV-L en complexe avec l'ARNv 5' 1-25, la LACV-L en pré-initiation (PDB : 6Z6G) (Arragain *et al.*, 2020) et la LACV-L en initiation (Arragain *et al.*, en cours de soumission). Les domaines sont indiqués et colorés de la manière identique. Les résidus de la pointe de la *PR loop* sont indiqués pour chacun des modèles.

3.3.1. Comparaison de la LACV-L et de la HTNV-L, un mécanisme d'amorçage et de réalignement similaire ?

Les résidus localisés à la pointe de la *PR loop* sont conservés chez les *Peribunyaviridae* et les *Hantaviridae* à l'exception de la méthionine 989 qui correspond, chez la HTNV-L, à la tyrosine 926. De par la nature de cet acide aminé, il est fort probable qu'il joue un rôle identique chez la HTNV-L en s'intercalant entre deux nucléotides de l'extrémité 3' de l'ARNv, stabilisant l'ensemble par empilement d'aromatique. De plus, dans les structures préalablement déterminées de la HTNV-L, la *PR loop* se trouve en position rétractée, analogue à un état de

pré-initiation chez la LACV-L (Gerlach *et al.*, 2015; Arragain *et al.*, 2020), ce qui suggère une potentielle extension de cette dernière lors de l'initiation de la réplication du génome viral au cours de la translocation de l'extrémité 3' de l'ARNv dans le site actif de la HTNV-L (Figure 50).

En 1995, Garcin *et al.* suggéraient que le mécanisme d'initiation chez les *Hantaviridae* consistait en l'hybridation d'un UTP sur le nucléotide A3 de l'extrémité 3' de l'ARNv (3'-UCAUCA...). La chaîne naissante serait ensuite synthétisée jusqu'au nucléotide A6 (3'-UCAUCA...) et l'amorce pppUAGU ultérieurement réalignée sur le nucléotide U1, laissant le pppU initiateur dépasser de l'extrémité 3' de l'ARNv et compromettant la conservation de l'intégrité du génome copié par la RdRp. De ce fait, afin de préserver l'information génétique du virus, le pppU initiateur serait alors clivé, laissant l'extrémité 5' du brin produit mono-phosphate et parfaitement complémentaire à l'ARN modèle (Garcin *et al.*, 1995).

Ce mécanisme de *prime and realign* aboutissant à la formation d'une extrémité 5' mono-phosphate du génome viral du HTNV permet de diminuer l'activation des senseurs intracellulaires comme RIG-I et dans une moindre mesure MDA5, qui sont responsables du déclenchement de la réponse immunitaire en induisant la production d'interférons et de cytokines au cours de l'infection virale. Ce mécanisme d'évasion à la réponse immunitaire innée serait également partagé chez les *Nairoviridae* qui disposent également d'une extrémité 5' mono-phosphate à la différence des *Phenuiviridae* et des *Peribunyaviridae* qui disposent d'une extrémité 5' triphosphate (Habjan *et al.*, 2008; voir Article n°3).

Chez le HTNV, aucune explication plausible n'existe quant au mécanisme de clivage du nucléotide additionnel en 5' qui surviendrait après l'étape d'amorçage et de réalignement. L'une des hypothèses pouvant être envisagées consisterait en l'intervention du domaine ENDO, mais aux vues de sa conformation dans les différentes structures de LACV-L obtenues, cette option reste relativement improbable. Afin de comprendre le mécanisme exact impliqué, il serait donc nécessaire de disposer de différentes structures de la HTNV-L entière.

Afin de comprendre exactement comment la HTNV-L initie la réplication de son génome, il sera vraisemblablement nécessaire d'utiliser une approche similaire à ce qui a été réalisé chez la LACV-L, à savoir muter l'extrémité 5' de l'ARNv dans le but de diminuer sa complémentarité avec l'extrémité 3' et une fois l'activité *in vitro* de la HTNV-L testée, d'estimer s'il est nécessaire ou non d'internaliser le tag de purification, qui est actuellement localisé en N-terminal de la HTNV-L. Ces différentes expériences apporteront *in fine* de nouvelles informations cruciales quant au mécanisme d'initiation de la réplication par amorçage et réalignement chez les *Hantaviridae*.

3.3.2. Synthèse de l'ARNc et de l'ARNv chez les Bunyavirales

Les tests d'activités *in vitro* de la LACV-L ont été réalisés dans le cadre de la synthèse de l'ARNc, identique au sens de transcription des ARNm codant pour les protéines virales. A l'heure actuelle, chez les Bunyavirus, aucune donnée structurale n'est disponible sur la réaction permettant au virus de synthétiser de nouvelles copies de l'ARNv. De ce fait, en utilisant des extrémités 5' et 3' correspondant à l'ARNc, une analyse fonctionnelle et structurale de la LACV-L permettrait de révéler les mécanismes qui empêchent la protéine L de transcrire les ARNc mais également d'apporter des informations structurales quant aux changements de conformations induits par leur fixation sur la RdRp. Ces données permettraient de mettre en lumière le rôle structural du mésappariement du 9^{ème} nucléotide des extrémités 5' et 3', conservé sur chacun des segments du génome viral et jouant un rôle dans la promotion de la transcription (Barr and Wertz, 2005). Cette analyse permettrait de révéler si le mécanisme d'initiation de la synthèse de l'ARNv fait appel à la *PR loop*, qui permet de stabiliser l'extrémité 3' de l'ARN modèle au cours de la synthèse de l'ARNc, et permettrait *in fine* de compléter les arrêts sur images du film du cycle de réplication des *Peribunyaviridae*.

Enfin, il est important de noter que la complémentarité des extrémités 5' et 3' dépend de chacun des segments codant pour les différentes protéines virales (Guu, Zheng and Tao, 2012). De ce fait, il serait également intéressant de tester des longueurs d'extrémités 5' et 3' de l'ARNv différentes pouvant induire la

formation d'un duplex distal plus ou moins important et permettant de révéler des détails sur la régulation de la transcription et/ou de la réplication des gènes viraux S, M et L.

3.3.3. Signal de terminaison de la transcription chez les Bunyavirales

L'ensemble de ces structures offre un aperçu général de la translocation de l'extrémité 3' de l'ARNv au sein de la LACV-L au cours de la réplication. En pré-initiation, l'extrémité 3' est liée à son site secondaire puis transloquée au niveau du site actif de la RdRp afin d'initier l'élongation du produit. Au cours de l'élongation initiale, l'extrémité 3' traverse la cavité interne pour ressortir du cœur de la RdRp et vient se loger sur son site secondaire, de manière similaire à l'état de pré-initiation. En transcription, le chemin suivi par l'extrémité 3' est identique à la réplication et a notamment été démontré chez la polymérase du virus influenza (Wandzik *et al.*, 2020). L'étape suivant l'élongation finale est la terminaison de la transcription. Chez les virus influenza, elle est caractérisée par la polyadénylation en 3' des ARNm viraux grâce à une séquence poly-U localisée en 5' du génome viral et réalisée par la RdRp via un mécanisme de bégayement (*stuttering*) (Poon *et al.*, 1999; Wandzik *et al.*, 2020).

Contrairement aux virus influenza pour lesquels le mécanisme de transcription s'effectue dans le noyau de la cellule infectée, la transcription des Bunyavirus s'effectue dans le cytoplasme et les ARNm viraux ne contiennent pas de queue poly-A. En effet, aucune séquence riche en U n'a été universellement identifiée chez les Bunyavirus et il a de plus été constaté que les signaux de terminaison de transcription prennent de nombreuses formes différentes, en termes de séquence ou de structure secondaire formée par l'ARN. Ces signaux de terminaison sont généralement situés à environ 100 nucléotides en amont de l'extrémité 5' du génome et sont également spécifiques de la famille virale et des différents segments (Eshita *et al.*, 1985; Collett, 1986; Bouloy *et al.*, 1990; Jin and Elliott, 1993). Dans l'état actuel des connaissances, deux motifs nucléotidiques se chevauchant et présents sur la région 5' NTR du segment S chez les *Peribunyaviridae* (BUNV) et les *Phenuiviridae* (RVFV) semblent jouer un rôle critique dans la signalisation de la terminaison, ces derniers consistant

respectivement en une séquence 3'-...UGUCG-5' et 3'-CGACGA...-5' (Barr, Rodgers and Wertz, 2006; Lara *et al.*, 2011; Blakqori, Lowen and Elliott, 2012).

Dû au peu d'informations disponibles quant aux signaux de terminaison propre aux Bunyavirus mais également en raison d'un potentiel mécanisme de transcription couplée à la traduction des ARNm viraux, la détermination de la structure d'un état de terminaison par cryo-ME reste donc complexe à mettre en place, bien que l'utilisation d'un ARN unique qui inclue les séquences des régions promotrice 5', 3' additionnées d'un signal de terminaison reste envisageable dans le futur.

3.3.4. Dimérisation et recrutement de partenaires chez les polymérase des sNSVs

Dans le cadre de la caractérisation structurale et fonctionnelle de la LACV-L, aucune multimérisation de la polymérase n'a été observée au sein des multiples jeux de données collectés et les produits de l'activité de réplication et de transcription ont été obtenus sans l'ajout de partenaires viraux ou cellulaires.

Au contraire de la famille des *Peribunyaviridae*, la dimérisation des RdRps a été récemment observée chez la famille des *Arenaviridae*. La formation de dimères de protéines L semble permettre la stabilisation des domaines C-terminaux en comparaison des RdRps monomériques et leur état de dimérisation semble également être augmenté en présence de l'extrémité 3' de l'ARNv. En mutant la boucle présente à l'interface formée entre les deux protéines L au sein du dimère, l'activité de réplication et de transcription du génome viral est grandement diminuée laissant supposer que la synthèse d'ARN nécessite la dimérisation des protéines L chez les *Arenaviridae* au cours de ces deux mécanismes (Peng *et al.*, 2020). Malheureusement, aucune structure obtenue en cours d'élongation ne vient supporter ces résultats et rien ne prouve que les mutations induites n'affectent tout simplement pas l'activité de la RdRp des *Arenaviridae*.

Chez le virus influenza, plusieurs modes de multimérisations ont également été rapportés et inclus au sein d'un modèle de réplication complet qui est le suivant (Figure 51) :

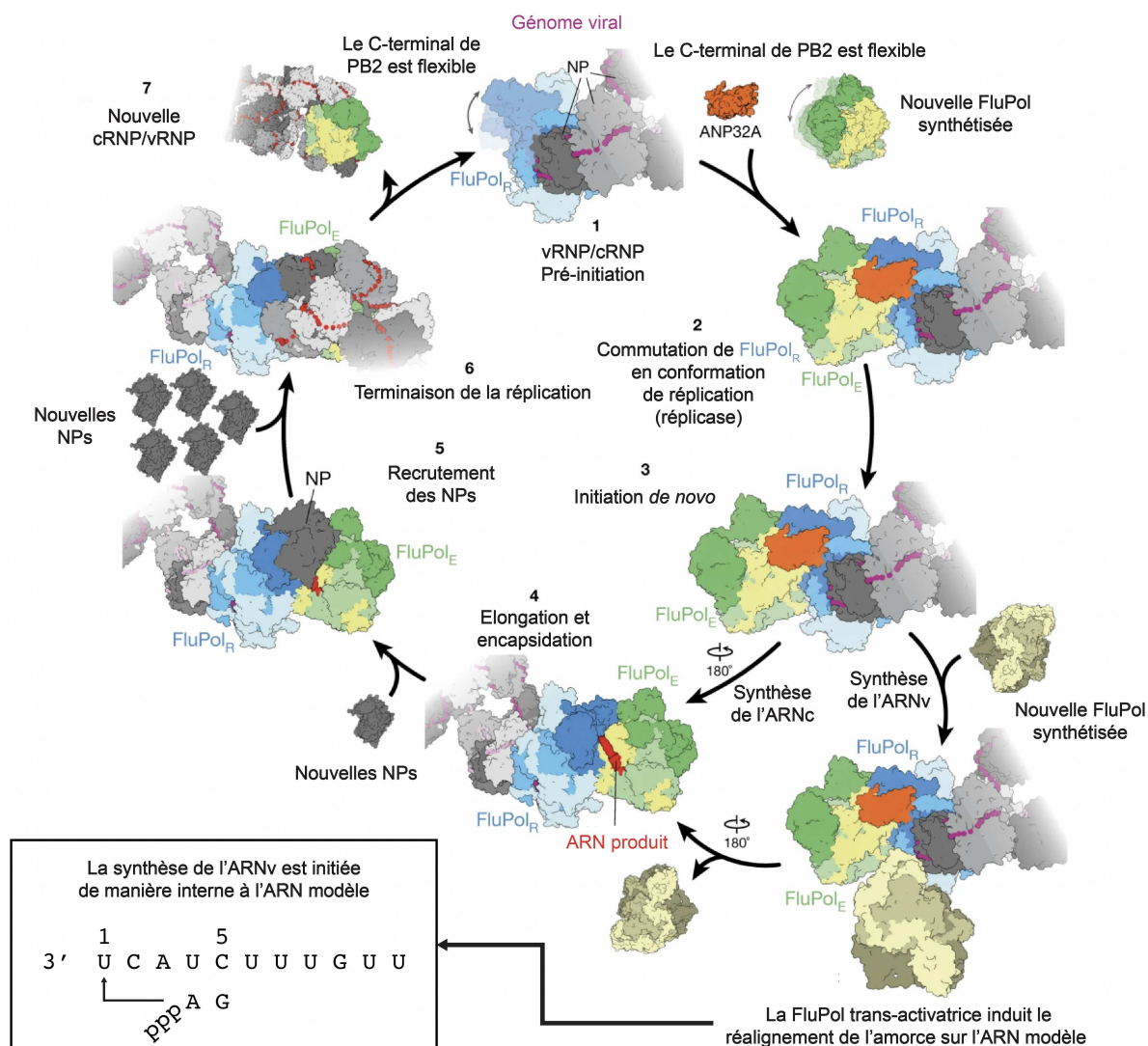


Figure 51. Modèle de la répliation du génome viral chez le virus influenza. Figure adaptée de Carrique *et al.*, 2020.

Au sein d'une vRNP ou d'une cRNP, le domaine PB2 de la polymérase virale (FluPol_R) est flexible. Ce dernier est alors stabilisé par (i) l'interaction d'une deuxième polymérase (FluPol_E) et (ii) du partenaire cellulaire ANP32 promouvant la formation d'un dimère. Ces deux éléments vont agir comme un commutateur et permettre à la FluPol_R de passer en mode de répliation, ou répliacase. Deux options sont alors possibles :

- Dans le cadre de la synthèse de l'ARNv, une troisième polymérase, dite trans-activatrice, va favoriser le mécanisme de réalignement au niveau du site actif de la FluPol_R pour *in fine* initier la répliation *de novo* interne à l'ARNc.

- Dans le cadre de la synthèse de l'ARNc, la réplication est initiée sans l'aide d'une polymérase trans-activatrice car la réplication *de novo* est terminale.

Dans les deux cas, lorsque la nouvelle extrémité 5' du produit de réplication nouvellement synthétisé quitte la FluPol_R, elle est immédiatement capturée par la FluPol_E et au fur et à mesure de l'élongation du produit, il sera progressivement encapsidé par le recrutement des NPs orchestrée en partie par ANP32. Les ARNv ou ARNc naissants s'assemblent alors en vRNP ou cRNP et sont libérés au cours de la terminaison de la réplication de FluPol_R, FluPol_E devenant une nouvelle unité fonctionnelle du virus (Oymans and Te Velhuis, 2018; Fan *et al.*, 2019; Carrique *et al.*, 2020) .

Chez la LACV-L, une plateforme potentielle d'interaction avec des partenaires cellulaires nécessaires à la réalisation du vol de coiffe en transcription ou d'une dimérisation au cours de la réplication, résideraient dans sa partie C-terminale. Cette dernière est en effet relativement exposée au solvant et est notamment composée d'un domaine de liaison au zinc (ZBD) dont la fonction reste inconnue. Pour conclure, de par le manque d'informations disponibles quant à de potentiels candidats cellulaires interagissant avec les protéines L, des études supplémentaires sont à envisager afin d'obtenir un tableau complet des différents régulateurs des mécanismes de réplication et de transcription que peuvent subir les protéines L des *Bunyavirales* au cours de l'infection virale.

Chapitre IV : Conclusion et perspectives générales

La cryo-ME a connu un développement impressionnant au cours de ces dix dernières années. D'une technique qualifiée de blobologie, extrêmement exigeante et pratiquée par quelques experts qui produisaient occasionnellement, et avec beaucoup d'efforts, des structures à moyenne résolution, elle est aujourd'hui devenue un pilier du domaine de la biologie structurale. De plus en plus accessible, elle ne cesse également de conquérir de nombreux laboratoires tant les possibilités qu'offre cette technique sont importantes. En effet, la cryo-ME permet d'étudier dans un environnement natif des objets de taille extrêmement variable, allant d'une protéine de 50 kDa à des cellules entières, ou de nature complètement différente, allant d'assemblages hélicoïdaux à des virus entiers. Ainsi, appliquée à l'étude de particules isolées, elle permet de résoudre les structures d'objets biologiques à une échelle quasi-atomique et de manière routinière, à condition que notre échantillon se prête au jeu de la congélation. Par collecte tomographique et moyenne de sous-tomogrammes, elle permet d'étudier ces mêmes objets dans un contexte cellulaire.

Cette technique est donc particulièrement prometteuse aujourd'hui et le sera encore plus dans les années à venir, en particulier aux vues des différentes améliorations encore réalisables en ce qui concerne le développement instrumental, avec des sources lumineuses de plus en plus cohérentes, des caméras de plus en plus rapides et sensibles, de nouveaux supports d'échantillons ou de nouvelles techniques de vitrification, qui *in fine*, permettront de surpasser les limites du possible actuel.

La révolution de la résolution atteignable est également attribuable à une autre facette de la cryo-ME : le développement de programmes de traitement d'images toujours plus performants. Ces derniers permettent désormais d'estimer et corriger les aberrations du microscope, les mouvements des particules induits par le faisceau d'électrons ou de déterminer les orientations et les translations à appliquer à nos particules de manière toujours plus précise. Les récents algorithmes de classification 3D permettent également de séparer de multiples conformations d'un même échantillon permettant d'obtenir de véritables « films moléculaires » de notre protéine en action.

En commençant ma thèse en Octobre 2017, j'ai eu la chance, tout au long de ces quatre années, d'évoluer au même rythme que ces développements, et la majorité d'entre eux ont été implémentés sur la plateforme de microscopie électronique au sein de l'Institut de Biologie Structurale ainsi que sur le cluster de calcul commun à l'IBS et à l'EMBL. Cette période a également vu la mise en place d'un Titan Krios à l'ESRF. J'ai donc pu tirer profit de toutes ces évolutions matérielles et logicielles en utilisant la cryo-ME en particules isolées comme outil afin de réaliser une analyse structurale et fonctionnelle de la réplication et de la transcription des *Bunyavirales*.

Cet ordre contient plus de 450 virus à ARN négatifs segmentés divisés en neuf familles principales. Certains d'entre eux sont considérés comme émergents et sont généralement associés à un taux de mortalité important. Durant cette thèse je me suis principalement concentré sur deux familles de cet ordre : les *Hantaviridae* et les *Peribunyaviridae*. Chacune d'elle contient de sérieux agents pathogènes comme le HTNV ou le LACV respectivement responsables de (i) fièvres hémorragiques à syndrome rénaux et pulmonaires ou des (ii) maladies neuro-invasives chez les jeunes adultes.

Le génome de ces virus est composé de trois segments codant pour quatre protéines structurales : le segment S, codant pour la NP, le segment M, codant pour les glycoprotéines de surface (Gn et Gc) et le segment L codant pour une RdRp, également appelée protéine L. Au sein de ces virus, chaque segment est couvert de nucléoprotéines NPs formant des assemblages hélicoïdaux appelés NCs. Les extrémités 5' et 3' de chaque segment sont spécifiquement liées à une protéine L formant des RNPs. Au sein d'un virion, il existe donc autant de RNPs que de segments. Ces assemblages macromoléculaires constituent l'unité fonctionnelle nécessaire à la réplication du génome viral et à sa transcription en ARNm. La réplication est réalisée *de novo* et nécessite deux étapes : la synthèse d'une copie de l'ARNv en ARNc de polarité positive, ultérieurement recopiée afin de produire de nouvelles copies du génome initial de polarité négative. La transcription est quant à elle réalisée à l'aide d'une amorce coiffée grâce à la combinaison de deux domaines présents au sein de la protéine L : un CBD qui permet de voler un ARNm appartenant à la cellule infectée et un domaine ENDO qui permet de cliver la coiffe

de cet ARNm et qui sera ultérieurement transloqué au niveau du site actif de la RdRp. Les objectifs initiaux de cette thèse consistaient à étudier chacun des différents acteurs impliqués dans ces deux mécanismes, à savoir les NCs et les protéines L.

Dans un premier temps, j'ai réussi à obtenir par cryo-ME et après optimisation des différents paramètres d'expression et de purification, la structure de la HTNV-NC à 3.3 Å de résolution. Cette structure 3D m'a permis de caractériser les différentes interactions entre les NPs, les résidus impliqués dans la liaison du génome viral et également de générer un modèle de déstabilisation de la NC au cours de la lecture de l'ARNv par la HTNV-L (Arragain *et al.*, 2019).

Fort de ces premiers résultats, je me suis donc naturellement penché sur l'étude structurale de la HTNV-L afin de disposer de la structure de chaque couple d'éléments qui composent l'unité fonctionnelle du virus : les NPs en complexe avec l'ARN et la HTNV-L liée aux extrémités 5' et 3' de l'ARNv. Ces analyses représentent le socle indispensable pour arriver au but final qui est de déterminer le fonctionnement de chacun de ces éléments au sein d'une RNP.

J'ai donc optimisé les paramètres d'expression et de purification de la HTNV-L afin d'obtenir une structure 3D en complexe avec les extrémités 5' et 3' de l'ARNv. Malgré les différents problèmes rencontrés tels que l'hétérogénéité de l'échantillon induite par le phénomène de congélation, la présence d'une orientation préférentielle et l'hybridation *in vitro* des extrémités du génome viral dû à leur quasi-complémentarité, j'ai finalement réussi à obtenir deux structures 3D à des résolutions de 3.4 et 3.6 Å dont l'une est liée à l'extrémité 5' de l'ARNv. Ces dernières m'ont permis de révéler l'architecture globale du cœur de la HTNV-L ainsi que le mode de fixation en *hook* de l'extrémité 5' du génome. De manière générale, la structure du cœur de la HTNV-L est analogue à la structure cristallographique de la LACV-L₁₋₁₇₅₀ ΔC-terminal (Gerlach *et al.*, 2015), ces deux protéines L partageant 47% de similarité de séquence et bien qu'il reste encore à déterminer l'organisation des domaines présents au C-terminal qui semble être spécifique à famille virale (Arragain *et al.*, 2020).

Afin de contourner les problèmes d'hétérogénéité et d'orientation préférentielle de la HTNV-L, je me suis donc attelé dans le même temps à la résolution de la structure de la LACV-L entière, qui jusque-là n'avait été cristallisée que sous sa forme tronquée (et entière mais à une résolution ne permettant pas de construire un modèle). De ce fait j'ai pu obtenir deux structures de la LACV-L. L'une dans un état de pré-initiation, me permettant de définir l'organisation des différents domaines présent en C-terminal, contenant notamment le CBD, et l'autre mimant un état d'élongation, me permettant de caractériser les différents changements de conformations que subit la LACV-L en présence d'un duplex d'ARN dans sa cavité interne (Arragain *et al.*, 2020).

Enfin, en couplant la mutation de l'extrémité 5' du génome viral afin de limiter son hybridation *in vitro* à l'extrémité 3' et en internalisant le tag de purification au sein de la séquence de la LACV-L, j'ai été en mesure de rétablir son activité *in vitro*. En optimisant cette dernière et en déterminant des conditions me permettant de bloquer la LACV-L dans différents états de réplication et de transcription, j'ai pu obtenir différents arrêts sur image du « film moléculaire » des différents changements de conformations que vit la LACV-L au cours de ses deux cycles de synthèse de l'ARNc et de l'ARNm. Cette caractérisation structurale et fonctionnelle m'a finalement permis d'établir un modèle d'initiation d'amorçage et de réalignement propre à la famille des *Peribunyaviridae* mais probablement conservé chez d'autres familles de l'ordre des *Bunyavirales* comme les *Hantaviridae* par exemple (Arragain *et al.*, en soumission).

L'ensemble de ces résultats, acquis au cours de ces quatre années de thèse, posent donc les bases d'une future compréhension complète, d'un point de vue structural et fonctionnel, des mécanismes de réplication et de transcription au sein de l'ordre des *Bunyavirales* et ouvrent évidemment la voie à de nouvelles expérimentations afin de parfaire notre connaissance de leur cycle viral.

Il reste maintenant notamment à déterminer les différentes étapes qui conduisent à la synthèse de l'ARNv à partir de l'ARNc mais également les vraisemblables changements de conformations induits au niveau des protéines L afin de ne pas effectuer la transcription d'ARN de polarité négative à partir d'une

matrice ARN de polarité positive. Il reste également à découvrir le mécanisme de terminaison de la synthèse des ARNm viraux, qui n'a été déterminé jusque-là que chez les virus influenza (Wandzik *et al.*, 2020), et qui diffère entre les *Articulavirales* et les *Bunyavirales* : la polymérase de grippe utilise un mécanisme de bégayement (*stuttering*) pour synthétiser un ARNm polyadénylé alors que les Bunyavirus reconnaissent un motif présent dans leur 5' NTR pour une terminaison sans ajout de queue poly-A. De plus, il reste à découvrir les partenaires cellulaires susceptibles d'interagir avec les RNPs et de caractériser leur implication dans les mécanismes de réplication et de transcription. Ces deux mécanismes se déroulant dans le cytoplasme de la cellule hôte chez les *Bunyavirales* et non dans le noyau comme chez les *Articulavirales* (Lukarska *et al.*, 2017; Carrique *et al.*, 2020), des partenaires cellulaires différents devraient être découverts. Enfin, il serait également intéressant de travailler sur le couplage de la transcription et de la traduction par la machinerie cellulaire de l'hôte, orchestré par les Bunyavirus (Barr, 2007).

Une fois notre compréhension des différents aspects structuraux et fonctionnels des acteurs individuels appartenant aux RNPs suffisamment avancée, il sera nécessaire de s'atteler à la reconstitution de ces unités virales *in vitro* ou à leur étude *in vivo* afin de pleinement caractériser leur fonctionnement au sein d'une cellule infectée. D'ici quelques années nous pourrons notamment compter sur le développement et l'accessibilité de nouveaux instruments, les cryo-microscopes à balayage couplés à des faisceaux d'ions focalisés (FIB-SEM) qui permettent de générer, à partir de cellules, des tranches d'une centaine de nm d'épaisseur. Ces techniques couplées aux récentes évolutions que connaît la microscopie optique-électronique corrélative (CLEM) ainsi que la cryo-tomographie, en termes de localisation d'évènements, stratégies de collecte et traitement d'images, permettront de développer une approche structurale multi-échelles visant à déterminer la structure de RNPs à partir de cellules infectées, dans de multiples états et tout au long de l'infection virale.

Dans le contexte pandémique actuel que nous traversons en raison de l'émergence du SARS-CoV-2, il est donc bon de rappeler qu'il est toujours capital de poursuivre ces études de recherche fondamentale entreprises sur des virus

pouvant être considérés comme émergents, à l'image de l'ordre des *Bunyavirales*. Premièrement, afin de comprendre leur mode de fonctionnement et d'infection et dans un deuxième temps, de manière plus appliquée, afin de développer de potentielles molécules antivirales à large spectre ou des candidats vaccins. Pour conclure, je n'ajouterai que cette phrase : la microscopie électronique et la virologie, c'est une histoire qui marche... et elle n'est pas prête de s'arrêter !

Références

Afonine, P.V. *et al.* (2018) ‘Real-space refinement in PHENIX for cryo-EM and crystallography’, p. 14.

Albornoz, A. *et al.* (2016) ‘Early Bunyavirus-Host Cell Interactions’, *Viruses*, 8(5). doi:10.3390/v8050143.

Amroun, A., Priet, S. and Querat, G. (2017) ‘Toscana virus cap-snatching and initiation of transcription’, *Journal of General Virology*, 98(11), pp. 2676–2688. doi:10.1099/jgv.0.000941.

Ariza, A. *et al.* (2013) ‘Nucleocapsid protein structures from orthobunyaviruses reveal insight into ribonucleoprotein architecture and RNA polymerization’, *Nucleic Acids Research*, 41(11), pp. 5912–5926. doi:10.1093/nar/gkt268.

Armstrong, M. *et al.* (2020) ‘Microscale Fluid Behavior during Cryo-EM Sample Blotting’, *Biophysical Journal*, 118(3), pp. 708–719. doi:10.1016/j.bpj.2019.12.017.

Arnold, S.A. *et al.* (2017) ‘Blotting-free and lossless cryo-electron microscopy grid preparation from nanoliter-sized protein samples and single-cell extracts’, *Journal of Structural Biology*, 197(3), pp. 220–226. doi:10.1016/j.jsb.2016.11.002.

Arragain, B. *et al.* (2019) ‘High resolution cryo-EM structure of the helical RNA-bound Hantaan virus nucleocapsid reveals its assembly mechanisms’, *eLife*. Edited by E.H. Egelman and J. Kuriyan, 8, p. e43075. doi:10.7554/eLife.43075.

Arragain, B. *et al.* (2020) ‘Pre-initiation and elongation structures of full-length La Crosse virus polymerase reveal functionally important conformational changes’, *Nature Communications*, 11(1), p. 3590. doi:10.1038/s41467-020-17349-4.

Arranz, R. *et al.* (2012) ‘The Structure of Native Influenza Virion Ribonucleoproteins’, *Science*, 338(6114), pp. 1634–1637. doi:10.1126/science.1228172.

Ashtiani, D. *et al.* (2018) ‘Delivery of femtolitre droplets using surface acoustic wave based atomisation for cryo-EM grid preparation’, *Journal of*

Structural Biology, 203(2), pp. 94–101. doi:10.1016/j.jsb.2018.03.012.

Baltimore, D. (1971) ‘Expression of animal virus genomes.’, *Microbiology and Molecular Biology Reviews*, 35(3), pp. 235–241.

Barr, J.N. *et al.* (2003) ‘Segment-specific terminal sequences of Bunyamwera bunyavirus regulate genome replication’, *Virology*, 311(2), pp. 326–338. doi:10.1016/s0042-6822(03)00130-2.

Barr, J.N. (2007) ‘Bunyavirus mRNA synthesis is coupled to translation to prevent premature transcription termination’, *RNA*, 13(5), pp. 731–736. doi:10.1261/rna.436607.

Barr, J.N., Rodgers, J.W. and Wertz, G.W. (2006) ‘Identification of the Bunyamwera bunyavirus transcription termination signal’, *The Journal of General Virology*, 87(Pt 1), pp. 189–198. doi:10.1099/vir.0.81355-0.

Barr, J.N. and Wertz, G.W. (2005) ‘Role of the conserved nucleotide mismatch within 3’- and 5’-terminal regions of Bunyamwera virus in signaling transcription’, *Journal of Virology*, 79(6), pp. 3586–3594. doi:10.1128/JVI.79.6.3586-3594.2005.

Battisti, A.J. *et al.* (2011) ‘Structural Studies of Hantaan Virus’, *Journal of Virology*, 85(2), pp. 835–841. doi:10.1128/JVI.01847-10.

Behrmann, E. *et al.* (2012) ‘Real-space processing of helical filaments in SPARX’, *Journal of Structural Biology*, 177(2), pp. 302–313. doi:10.1016/j.jsb.2011.12.020.

Beijerinck, M.W. (1898) ‘On a Contagium vivum fluidum causing the Spotted disease of the Tobacco-leaves’, *Koninklijke Nederlandse Akademie van Wetenschappen Proceedings Series B Physical Sciences*, 1, pp. 170–176.

Belloq, C. and Kolakofsky, D. (1987) ‘Translational requirement for La Crosse virus S-mRNA synthesis: a possible mechanism.’, *Journal of Virology*, 61(12), pp. 3960–3967.

Bepler, T. *et al.* (2020) ‘Topaz-Denoise: general deep denoising models for cryoEM and cryoET’, *Nature Communications*, 11(1), p. 5208. doi:10.1038/s41467-

020-18952-1.

Berriman, J. and Unwin, N. (1994) 'Analysis of transient structures by cryo-microscopy combined with rapid mixing of spray droplets', *Ultramicroscopy*, 56(4), pp. 241–252. doi:10.1016/0304-3991(94)90012-4.

Bishop, D.H. *et al.* (1982) 'The complete sequence and coding content of snowshoe hare bunyavirus small (S) viral RNA species.', *Nucleic Acids Research*, 10(12), pp. 3703–3713.

Blakqori, G., Lowen, A.C. and Elliott, R.M. (2012) 'The small genome segment of Bunyamwera orthobunyavirus harbours a single transcription-termination signal', *Journal of General Virology*, 93(7), pp. 1449–1455. doi:10.1099/vir.0.042390-0.

Bouloy, M. *et al.* (1990) 'Characterization of the 5' and 3' ends of viral messenger RNAs isolated from BHK21 cells infected with germiston virus (bunyavirus)', *Virology*, 175(1), pp. 50–58. doi:10.1016/0042-6822(90)90185-T.

Bowden, T.A. *et al.* (2013) 'Orthobunyavirus Ultrastructure and the Curious Tripodal Glycoprotein Spike', *PLoS Pathogens*, 9(5). doi:10.1371/journal.ppat.1003374.

Brilot, A.F. *et al.* (2012) 'Beam-Induced Motion of Vitrified Specimen on Holey Carbon Film', *Journal of structural biology*, 177(3), pp. 630–637. doi:10.1016/j.jsb.2012.02.003.

Brunotte, L. *et al.* (2011) 'Structure of the Lassa Virus Nucleoprotein Revealed by X-ray Crystallography, Small-angle X-ray Scattering, and Electron Microscopy', *The Journal of Biological Chemistry*, 286(44), pp. 38748–38756. doi:10.1074/jbc.M111.278838.

Cardone, G., Heymann, J.B. and Steven, A.C. (2013) 'One number does not fit all: Mapping local variations in resolution in cryo-EM reconstructions', *Journal of Structural Biology*, 184(2), pp. 226–236. doi:10.1016/j.jsb.2013.08.002.

Carragher, B. *et al.* (2000) 'Leginon: An Automated System for Acquisition of Images from Vitreous Ice Specimens', *Journal of Structural Biology*, 132(1), pp. 33–45. doi:10.1006/jsbi.2000.4314.

Carrique, L. *et al.* (2020) ‘Host ANP32A mediates the assembly of the influenza virus replicase’, *Nature*, 587(7835), pp. 638–643. doi:10.1038/s41586-020-2927-z.

Castro, C. *et al.* (2007) ‘Two proton transfers in the transition state for nucleotidyl transfer catalyzed by RNA- and DNA-dependent RNA and DNA polymerases’, *Proceedings of the National Academy of Sciences of the United States of America*, 104(11), pp. 4267–4272. doi:10.1073/pnas.0608952104.

Castro, C. *et al.* (2009) ‘Nucleic acid polymerases use a general acid for nucleotidyl transfer’, *Nature Structural & Molecular Biology*, 16(2), pp. 212–218. doi:10.1038/nsmb.1540.

Chang, S. *et al.* (2015) ‘Cryo-EM structure of influenza virus RNA polymerase complex at 4.3 Å resolution’, *Molecular Cell*, 57(5), pp. 925–935. doi:10.1016/j.molcel.2014.12.031.

Chen, S. *et al.* (2013) ‘High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy’, *Ultramicroscopy*, 135, pp. 24–35. doi:10.1016/j.ultramic.2013.06.004.

Cheng, A. *et al.* (2018) ‘High resolution single particle cryo-electron microscopy using beam-image shift’, *Journal of Structural Biology*, 204(2), pp. 270–275. doi:10.1016/j.jsb.2018.07.015.

Cheng, D. *et al.* (2012) ‘Practical Considerations in the Successful Preparation of Specimens for Thin-Film Cryo-Transmission Electron Microscopy’, in, pp. 880–890.

Cifuentes-Muñoz, N., Salazar-Quiroz, N. and Tischler, N.D. (2014) ‘Hantavirus Gn and Gc Envelope Glycoproteins: Key Structural Units for Virus Cell Entry and Virus Assembly’, *Viruses*, 6(4), pp. 1801–1822. doi:10.3390/v6041801.

Cochran, W., Crick, F.H. and Vand, V. (1952) ‘The structure of synthetic polypeptides. I. The transform of atoms on a helix’, *Acta Crystallographica*, 5(5), pp. 581–586. doi:10.1107/S0365110X52001635.

Collett, M.S. (1986) ‘Messenger RNA of the M segment RNA of Rift Valley fever virus’, *Virology*, 151(1), pp. 151–156. doi:10.1016/0042-6822(86)90114-5.

Crowther, R.A., DeRosier, D.J. and Klug, A. (1970) ‘The reconstruction of a three-dimensional structure from projections and its application to electron microscopy’, *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 317(1530), pp. 319–340. doi:10.1098/rspa.1970.0119.

Crowther, R.A., Huxley, H.E. and Klug, A. (1971) ‘Procedures for three-dimensional reconstruction of spherical viruses by Fourier synthesis from electron micrographs’, *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 261(837), pp. 221–230. doi:10.1098/rstb.1971.0054.

Dandey, V.P. *et al.* (2020) ‘Time-resolved cryo-EM using Spotiton’, *Nature Methods*, 17(9), pp. 897–900. doi:10.1038/s41592-020-0925-6.

Danev, R. *et al.* (2014) ‘Volta potential phase plate for in-focus phase contrast transmission electron microscopy’, *Proceedings of the National Academy of Sciences*, 111(44), pp. 15635–15640.

De Rosier, D.J. and Klug, A. (1968) ‘Reconstruction of Three Dimensional Structures from Electron Micrographs’, *Nature*, 217(5124), pp. 130–134. doi:10.1038/217130a0.

Deng, T., Vreede, F.T. and Brownlee, G.G. (2006) ‘Different De Novo Initiation Strategies Are Used by Influenza Virus RNA Polymerase on Its cRNA and Viral RNA Promoters during Viral RNA Replication’, *Journal of Virology*, 80(5), pp. 2337–2348. doi:10.1128/JVI.80.5.2337-2348.2006.

Desfosses, A. (2012) *Analyse d’images de microscopie électronique de biopolymères hélicoïdaux flexibles*. These de doctorat. Grenoble. Available at: <https://www.theses.fr/2012GRENV079> (Accessed: 27 June 2021).

Desfosses, A. *et al.* (2014) ‘SPRING – An image processing package for single-particle based helical reconstruction from electron cryomicrographs’, *Journal of Structural Biology*, 185(1), pp. 15–26. doi:10.1016/j.jsb.2013.11.003.

Diaz, R., Rice, W., J. and Stokes, D.L. (2010) ‘Fourier–Bessel Reconstruction of Helical Assemblies’, in *Methods in Enzymology*. Elsevier, pp. 131–165.

doi:10.1016/S0076-6879(10)82005-1.

D’Imprima, E. *et al.* (2019) ‘Protein denaturation at the air-water interface and how to prevent it’, *eLife*. Edited by A.T. Brunger *et al.*, 8, p. e42747. doi:10.7554/eLife.42747.

Donchet, A. *et al.* (2019) ‘The structure of the nucleoprotein of Influenza D shows that all Orthomyxoviridae nucleoproteins have a similar NP CORE , with or without a NP TAIL for nuclear transport’, *Scientific Reports*, 9(1), p. 600. doi:10.1038/s41598-018-37306-y.

Dubochet, J. *et al.* (1988) ‘Cryo-electron microscopy of vitrified specimens’, *Quarterly Reviews of Biophysics*, 21(2), pp. 129–228.

Dubochet, J. and McDowell, A.W. (1981) ‘Vitrification of Pure Water for Electron Microscopy’, *Journal of Microscopy*, 124(3), pp. 3–4. doi:10.1111/j.1365-2818.1981.tb02483.x.

Egelman, E.H. (2000) ‘A robust algorithm for the reconstruction of helical filaments using single-particle methods’, *Ultramicroscopy*, 85(4), pp. 225–234. doi:10.1016/s0304-3991(00)00062-0.

Egelman, E.H. (2007) ‘Single-particle reconstruction from EM images of helical filaments’, *Current Opinion in Structural Biology*, 17(5), pp. 556–561. doi:10.1016/j.sbi.2007.07.006.

Egelman, E.H. (2014) ‘Ambiguities in helical reconstruction’, *eLife*. Edited by W.I. Sundquist, 3, p. e04969. doi:10.7554/eLife.04969.

Eifan, S. *et al.* (2013) ‘Non-Structural Proteins of Arthropod-Borne Bunyaviruses: Roles and Functions’, *Viruses*, 5(10), pp. 2447–2468. doi:10.3390/v5102447.

Elliott, R. M. (1989) ‘Nucleotide sequence analysis of the large (L) genomic RNA segment of Bunyamwera virus, the prototype of the family Bunyaviridae’, *Virology*, 173(2), pp. 426–436. doi:10.1016/0042-6822(89)90555-2.

Elliott, R.M. (1989) *Nucleotide sequence analysis of the small (S) RNA segment of Bunyamwera virus, the prototype of the family Bunyaviridae*, *The*

Journal of general virology. J Gen Virol. doi:10.1099/0022-1317-70-5-1281.

Elliott, R.M. (1997) 'Emerging viruses: the Bunyaviridae', *Molecular Medicine (Cambridge, Mass.)*, 3(9), pp. 572–577.

Emsley, P. and Cowtan, K. (2004) 'Coot: model-building tools for molecular graphics', *Acta Crystallographica Section D: Biological Crystallography*, 60(12), pp. 2126–2132. doi:10.1107/S0907444904019158.

Erickson, H.P. and Klug, A. (1970) 'The Fourier Transform of an Electron Micrograph: Effects of Defocussing and Aberrations, and Implications for the Use of Underfocus Contrast Enhancement', *Berichte der Bunsengesellschaft für physikalische Chemie*, 74(11), pp. 1129–1137. doi:10.1002/bbpc.19700741109.

Eshita, Y. *et al.* (1985) 'Analyses of the mRNA transcription processes of snowshoe hare bunyavirus S and M RNA species', *Journal of Virology*, 55(3), pp. 681–689. doi:10.1128/jvi.55.3.681-689.1985.

Falk, B.W. and Tsai, J.H. (1998) 'Biology and molecular biology of viruses in the genus Tenuivirus', *Annual Review of Phytopathology*, 36, pp. 139–163. doi:10.1146/annurev.phyto.36.1.139.

Fan, G.Y. and Ellisman, M.H. (2000) 'Digital imaging in transmission electron microscopy', *Journal of Microscopy*, 200(1), pp. 1–13. doi:10.1046/j.1365-2818.2000.00737.x.

Fan, H. *et al.* (2019) 'Structures of influenza A virus RNA polymerase offer insight into viral genome replication', *Nature*, 573(7773), pp. 287–290. doi:10.1038/s41586-019-1530-7.

Faruqi, A.R. *et al.* (2005) 'Direct single electron detection with a CMOS detector for electron microscopy', *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 546(1), pp. 170–175. doi:10.1016/j.nima.2005.03.023.

Feathers, J.R., Spoth, K.A. and Fromme, J.C. (2021) 'Experimental evaluation of super-resolution imaging and magnification choice in single-particle cryo-EM', *Journal of Structural Biology: X*, 5, p. 100047. doi:10.1016/j.yjsbx.2021.100047.

Fermin, G. (2018) ‘Virion Structure, Genome Organization, and Taxonomy of Viruses’, *Viruses*, pp. 17–54. doi:10.1016/B978-0-12-811257-1.00002-4.

Fernández-García, Y. *et al.* (2016) ‘Atomic Structure and Biochemical Characterization of an RNA Endonuclease in the N Terminus of Andes Virus L Protein’, *PLOS Pathogens*, 12(6), p. e1005635. doi:10.1371/journal.ppat.1005635.

Ferron, F. *et al.* (2017) ‘Transcription and replication mechanisms of Bunyaviridae and Arenaviridae L proteins’, *Virus Research*, 234, pp. 118–134. doi:10.1016/j.virusres.2017.01.018.

Fontana, J. *et al.* (2008) ‘The unique architecture of Bunyamwera virus factories around the Golgi complex’, *Cellular Microbiology*, 10(10), pp. 2012–2028. doi:https://doi.org/10.1111/j.1462-5822.2008.01184.x.

Frank, J. (1975) ‘Averaging of low exposure electron micrographs of non-periodic objects’, *Ultramicroscopy*, 1(2), pp. 159–162. doi:10.1016/s0304-3991(75)80020-9.

Frank, J. *et al.* (1996) ‘SPIDER and WEB: Processing and Visualization of Images in 3D Electron Microscopy and Related Fields’, *Journal of Structural Biology*, 116(1), pp. 190–199. doi:10.1006/jsbi.1996.0030.

Freiberg, A.N. *et al.* (2008) ‘Three-Dimensional Organization of Rift Valley Fever Virus Revealed by Cryoelectron Tomography’, *Journal of Virology*, 82(21), pp. 10341–10348. doi:10.1128/JVI.01191-08.

Gallagher, J.R. *et al.* (2017) ‘Structural studies of influenza virus RNPs by electron microscopy indicate molecular contortions within NP supra-structures’, *Journal of Structural Biology*, 197(3), pp. 294–307. doi:10.1016/j.jsb.2016.12.007.

Garcin, D. *et al.* (1995) ‘The 5’ ends of Hantaan virus (Bunyaviridae) RNAs suggest a prime-and-realign mechanism for the initiation of RNA synthesis.’, *Journal of Virology*, 69(9), pp. 5754–5762.

Garcin, D. and Kolakofsky, D. (1990) ‘A novel mechanism for the initiation of Tacaribe arenavirus genome replication.’, *Journal of Virology*, 64(12), pp. 6196–6203.

Garcin, D. and Kolakofsky, D. (1992) 'Tacaribe arenavirus RNA synthesis in vitro is primer dependent and suggests an unusual model for the initiation of genome replication.', *Journal of Virology*, 66(3), pp. 1370–1376. doi:10.1128/JVI.66.3.1370-1376.1992.

Gavrilovskaya, I.N. *et al.* (1998) 'β3 integrins mediate the cellular entry of hantaviruses that cause respiratory failure', *Proceedings of the National Academy of Sciences*, 95(12), pp. 7074–7079. doi:10.1073/pnas.95.12.7074.

Gavrilovskaya, I.N. *et al.* (1999) 'Cellular Entry of Hantaviruses Which Cause Hemorrhagic Fever with Renal Syndrome Is Mediated by β3 Integrins', *Journal of Virology*, 73(5), pp. 3951–3959.

Gelderblom, H.R. and Krüger, D.H. (2014) '1 - Helmut Ruska (1908–1973): His Role in the Evolution of Electron Microscopy in the Life Sciences, and Especially Virology', in Hawkes, P.W. (ed.) *Advances in Imaging and Electron Physics*. Elsevier, pp. 1–94. doi:10.1016/B978-0-12-800146-2.00001-1.

Gerlach, P. *et al.* (2015) 'Structural Insights into Bunyavirus Replication and Its Regulation by the vRNA Promoter', *Cell*, 161(6), pp. 1267–1279. doi:10.1016/j.cell.2015.05.006.

Glaeser, R.M. (2016) 'Specimen Behavior in the Electron Beam', in *Methods in Enzymology*. Elsevier, pp. 19–50. doi:10.1016/bs.mie.2016.04.010.

Glaeser, Robert.M. *et al.* (2011) 'Review: Precise Beam-Tilt Alignment And Collimation Are Required To Minimize The Phase Error Associated With Coma In High-Resolution Cryo-EM', *Journal of structural biology*, 174(1), pp. 1–10. doi:10.1016/j.jsb.2010.12.005.

Gogrefe, N. *et al.* (2019) 'Structure of a functional cap-binding domain in Rift Valley fever virus L protein', *PLoS pathogens*, 15(5), p. e1007829. doi:10.1371/journal.ppat.1007829.

Goldsmith, C.S. *et al.* (1995) 'Ultrastructural characteristics of Sin Nombre virus, causative agent of hantavirus pulmonary syndrome', *Archives of Virology*, 140(12), pp. 2107–2122. doi:10.1007/BF01323234.

Grant, T. and Grigorieff, N. (2015) 'Measuring the optimal exposure for

single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6', *eLife*. Edited by W.I. Sundquist, 4, p. e06980. doi:10.7554/eLife.06980.

Grigorieff, N. (2007) 'FREALIGN: High-resolution refinement of single particle structures', *Journal of Structural Biology*, 157(1), pp. 117–125. doi:10.1016/j.jsb.2006.05.004.

Gubbens, A. *et al.* (2010) 'The GIF Quantum, a next generation post-column imaging energy filter', *Ultramicroscopy*, 110(8), pp. 962–970. doi:10.1016/j.ultramic.2010.01.009.

Guilligay, D. *et al.* (2008) 'The structural basis for cap binding by influenza virus polymerase subunit PB2', *Nature Structural & Molecular Biology*, 15(5), pp. 500–506. doi:10.1038/nsmb.1421.

Guo, Y. *et al.* (2012) 'Crimean–Congo hemorrhagic fever virus nucleoprotein reveals endonuclease activity in bunyaviruses', *Proceedings of the National Academy of Sciences*, 109(13), pp. 5046–5051. doi:10.1073/pnas.1200808109.

Guu, T.S.Y., Zheng, W. and Tao, Y.J. (2012) 'Bunyavirus: Structure and Replication', in Rossmann, M.G. and Rao, V.B. (eds) *Viral Molecular Machines*. Boston, MA: Springer US (Advances in Experimental Medicine and Biology), pp. 245–266. doi:10.1007/978-1-4614-0980-9_11.

Habjan, M. *et al.* (2008) 'Processing of genome 5' termini as a strategy of negative-strand RNA viruses to avoid RIG-I-dependent interferon induction', *PloS One*, 3(4), p. e2032. doi:10.1371/journal.pone.0002032.

Haddow, A.D. and Odoi, A. (2009) 'The incidence risk, clustering, and clinical presentation of La Crosse virus infections in the eastern United States, 2003-2007', *PloS One*, 4(7), p. e6145. doi:10.1371/journal.pone.0006145.

Hamaguchi, T. *et al.* (2019) 'A new cryo-EM system for single particle analysis', *Journal of Structural Biology*, 207(1), pp. 40–48. doi:10.1016/j.jsb.2019.04.011.

Hameau, J. (1779-1851) A. du texte (1895) *Étude sur les virus*, par Jean Hameau,... [Publiée par le Dr G. Hameau.] Préface par M. Grancher,... Available at: <https://gallica.bnf.fr/ark:/12148/bpt6k54038012> (Accessed: 12 March 2021).

He, S. and Scheres, S.H.W. (2017) 'Helical reconstruction in RELION', *Journal of Structural Biology*, 198(3), pp. 163–176. doi:10.1016/j.jsb.2017.02.003.

van Heel, M. *et al.* (1996) 'A new generation of the IMAGIC image processing system', *Journal of Structural Biology*, 116(1), pp. 17–24. doi:10.1006/jsbi.1996.0004.

van Heel, M. *et al.* (2000) 'Single-particle electron cryo-microscopy: towards atomic resolution', *Quarterly Reviews of Biophysics*, 33(4), pp. 307–369. doi:10.1017/S0033583500003644.

Henderson, R. (1995) 'The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules', *Quarterly Reviews of Biophysics*, 28(2), pp. 171–193. doi:10.1017/S003358350000305X.

Henderson, R. *et al.* (2012) 'Outcome of the First Electron Microscopy Validation Task Force Meeting', *Structure(London, England:1993)*, 20–330(2), pp. 205–214. doi:10.1016/j.str.2011.12.014.

Henderson, R. (2013) 'Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise', *Proceedings of the National Academy of Sciences*, 110(45), pp. 18037–18041.

Henning, S. and Adhikari, R. (2017) 'Scanning Electron Microscopy, ESEM, and X-ray Microanalysis', in *Microscopy Methods in Nanomaterials Characterization*. Elsevier, pp. 1–30. doi:10.1016/B978-0-323-46141-2.00001-8.

Hepojoki, J. *et al.* (2010) 'Cytoplasmic tails of hantavirus glycoproteins interact with the nucleocapsid protein', *The Journal of General Virology*, 91(Pt 9), pp. 2341–2350. doi:10.1099/vir.0.021006-0.

Herath, V. *et al.* (2020) 'Family Level Phylogenies Reveal Relationships of Plant Viruses within the Order Bunyavirales', *Viruses*, 12(9), p. 1010. doi:10.3390/v12091010.

Hoffmann, B. *et al.* (2012) 'Novel Orthobunyavirus in Cattle, Europe, 2011', *Emerging Infectious Diseases*, 18(3), pp. 469–472. doi:10.3201/eid1803.111905.

Hofmann, H. *et al.* (2013) 'Severe Fever with Thrombocytopenia Virus Glycoproteins Are Targeted by Neutralizing Antibodies and Can Use DC-SIGN as a Receptor for pH-Dependent Entry into Human and Animal Cell Lines', *Journal of Virology*, 87(8), pp. 4384–4394. doi:10.1128/JVI.02628-12.

Holm, T. *et al.* (2018) 'Biochemical and structural studies reveal differences and commonalities among cap-snatching endonucleases from segmented negative-strand RNA viruses', *The Journal of Biological Chemistry*, 293(51), pp. 19686–19698. doi:10.1074/jbc.RA118.004373.

Horne, K.M. and Vanlandingham, D.L. (2014) 'Bunyavirus-vector interactions', *Viruses*, 6(11), pp. 4373–4397. doi:10.3390/v6114373.

Hornung, V. *et al.* (2006) '5'-Triphosphate RNA Is the Ligand for RIG-I', *Science*, 314(5801), pp. 994–997. doi:10.1126/science.1132505.

Huiskonen, J.T. *et al.* (2009) 'Electron Cryo-Microscopy and Single-Particle Averaging of Rift Valley Fever Virus: Evidence for GN-GC Glycoprotein Heterodimers', *Journal of Virology*, 83(8), pp. 3762–3769. doi:10.1128/JVI.02483-08.

Huiskonen, J.T. *et al.* (2010) 'Electron Cryotomography of Tula Hantavirus Suggests a Unique Assembly Paradigm for Enveloped Viruses', *Journal of Virology*, 84(10), pp. 4889–4897. doi:10.1128/JVI.00057-10.

Hutchinson, K.L., Peters, C.J. and Nichol, S.T. (1996) 'Sin Nombre virus mRNA synthesis', *Virology*, 224(1), pp. 139–149. doi:10.1006/viro.1996.0515.

Ikegami, T. *et al.* (2005) 'Rift Valley Fever Virus NSs mRNA Is Transcribed from an Incoming Anti-Viral-Sense S RNA Segment', *Journal of Virology*, 79(18), pp. 12106–12111. doi:10.1128/JVI.79.18.12106-12111.2005.

Jakobi, A.J., Wilmanns, M. and Sachse, C. (2017) 'Model-based local density sharpening of cryo-EM maps', *eLife*. Edited by A.T. Brunger, 6, p. e27131. doi:10.7554/eLife.27131.

Jin, H. and Elliott, R.M. (1992) 'Mutagenesis of the L protein encoded by Bunyamwera virus and production of monospecific antibodies', *The Journal of General Virology*, 73 (Pt 9), pp. 2235–2244. doi:10.1099/0022-1317-73-9-2235.

Jin, H. and Elliott, R.M. (1993) 'Characterization of Bunyamwera virus S RNA that is transcribed and replicated by the L protein expressed from recombinant vaccinia virus', *Journal of Virology*, 67(3), pp. 1396–1404. doi:10.1128/jvi.67.3.1396-1404.1993.

Jones, R. *et al.* (2019) 'Structure and function of the Toscana virus cap-snatching endonuclease', *Nucleic Acids Research*, 47(20), pp. 10914–10930. doi:10.1093/nar/gkz838.

de Jong, A.F. and Van Dyck, D. (1993) 'Ultimate resolution and information in electron microscopy II. The information limit of transmission electron microscopes', *Ultramicroscopy*, 49(1–4), pp. 66–80. doi:10.1016/0304-3991(93)90213-H.

Jonsson, C.B., Figueiredo, L.T.M. and Vapalahti, O. (2010) 'A Global Perspective on Hantavirus Ecology, Epidemiology, and Disease', *Clinical Microbiology Reviews*, 23(2), pp. 412–441. doi:10.1128/CMR.00062-09.

Kasas, S. *et al.* (2003) 'Vitrification of cryoelectron microscopy specimens revealed by high-speed photographic imaging', *Journal of Microscopy*, 211(1), pp. 48–53. doi:10.1046/j.1365-2818.2003.01193.x.

Kaur, S. *et al.* (2021) 'Local computational methods to improve the interpretability and analysis of cryo-EM maps', *Nature Communications*, 12(1), p. 1240. doi:10.1038/s41467-021-21509-5.

Khoshouei, M. *et al.* (2017) 'Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate', *Nature Communications*, 8(1), p. 16099. doi:10.1038/ncomms16099.

Klug, A., Crick, F.H.C. and Wyckoff, H.W. (1958) 'Diffraction by helical structures', *Acta Crystallographica*, 11(3), pp. 199–213. doi:10.1107/S0365110X58000517.

Kohl, A. *et al.* (2004) 'Complementarity, sequence and structural elements within the 3' and 5' non-coding regions of the Bunyamwera orthobunyavirus S segment determine promoter strength', *The Journal of General Virology*, 85(Pt 11), pp. 3269–3278. doi:10.1099/vir.0.80407-0.

Koning, R.I., Koster, A.J. and Sharp, T.H. (2018) 'Advances in cryo-electron tomography for biology and medicine', *Annals of Anatomy - Anatomischer Anzeiger*, 217, pp. 82–96. doi:10.1016/j.aanat.2018.02.004.

Koppstein, D., Ashour, J. and Bartel, D.P. (2015) 'Sequencing the capsid-snatching repertoire of H1N1 influenza provides insight into the mechanism of viral transcription initiation', *Nucleic Acids Research*, 43(10), pp. 5052–5064. doi:10.1093/nar/gkv333.

Kouba, T., Drncová, P. and Cusack, S. (2019) 'Structural snapshots of actively transcribing influenza polymerase', *Nature Structural & Molecular Biology*, 26(6), pp. 460–470. doi:10.1038/s41594-019-0232-z.

Kucukelbir, A., Sigworth, F.J. and Tagare, H.D. (2014) 'Quantifying the local resolution of cryo-EM density maps', *Nature Methods*, 11(1), pp. 63–65. doi:10.1038/nmeth.2727.

Kühlbrandt, W. (2014) 'The Resolution Revolution', *Science*, 343(6178), pp. 1443–1444. doi:10.1126/science.1251652.

Lara, E. *et al.* (2011) 'Characterization of Wild-Type and Alternate Transcription Termination Signals in the Rift Valley Fever Virus Genome', *Journal of Virology*, 85(23), pp. 12134–12145. doi:10.1128/JVI.05322-11.

Lees, J.F., Pringle, C.R. and Elliott, R.M. (1986) 'Nucleotide sequence of the Bunyamwera virus M RNA segment: Conservation of structural features in the Bunyavirus glycoprotein gene product', *Virology*, 148(1), pp. 1–14. doi:10.1016/0042-6822(86)90398-3.

Li, S. *et al.* (2016) 'Acidic pH-Induced Conformations and LAMP1 Binding of the Lassa Virus Glycoprotein Spike', *PLoS Pathogens*, 12(2). doi:10.1371/journal.ppat.1005418.

Li, X. *et al.* (2013) 'Electron counting and beam-induced motion correction enable near atomic resolution single particle cryoEM', *Nature methods*, 10(6), pp. 584–590. doi:10.1038/nmeth.2472.

Liu, X. *et al.* (2018) 'Rice Stripe Tenuivirus Has a Greater Tendency To Use the Prime-and-Realign Mechanism in Transcription of Genomic than in

Transcription of Antigenomic Template RNAs', *Journal of Virology*, 92(1). doi:10.1128/JVI.01414-17.

Lozach, P.-Y. *et al.* (2011) 'DC-SIGN as a Receptor for Phleboviruses', *Cell Host & Microbe*, 10(1), pp. 75–88. doi:10.1016/j.chom.2011.06.007.

Lukarska, M. *et al.* (2017) 'Structural basis of an essential interaction between influenza polymerase and Pol II CTD', *Nature*, 541(7635), pp. 117–121. doi:10.1038/nature20594.

Lwoff, A. (1957) 'The concept of virus', *Journal of General Microbiology*, 17(2), pp. 239–253. doi:10.1099/00221287-17-2-239.

Lyumkis, D. (2019) 'Challenges and opportunities in cryo-EM single-particle analysis', *The Journal of Biological Chemistry*, 294(13), pp. 5181–5197. doi:10.1074/jbc.REV118.005602.

Maes, P. *et al.* (2018) 'Taxonomy of the family Arenaviridae and the order Bunyavirales: update 2018', *Archives of Virology*, 163(8), pp. 2295–2310. doi:10.1007/s00705-018-3843-5.

Mastrorarde, D.N. (2005) 'Automated electron microscope tomography using robust prediction of specimen movements', *Journal of Structural Biology*, 152(1), pp. 36–51. doi:10.1016/j.jsb.2005.07.007.

McJunkin, J.E. *et al.* (2009) *La Crosse Encephalitis in Children*, <http://dx.doi.org/10.1056/NEJM200103153441103>. Massachusetts Medical Society. doi:10.1056/NEJM200103153441103.

McMullan, G. *et al.* (2009) 'Detective quantum efficiency of electron area detectors in electron microscopy', *Ultramicroscopy*, 109(9), pp. 1126–1143. doi:10.1016/j.ultramic.2009.04.002.

McMullan, G., Faruqi, A.R. and Henderson, R. (2016) 'Direct Electron Detectors', in *Methods in Enzymology*. Elsevier, pp. 1–17. doi:10.1016/bs.mie.2016.05.056.

Mielke-Ehret, N. and Mühlbach, H.-P. (2012) 'Emaravirus: a novel genus of multipartite, negative strand RNA plant viruses', *Viruses*, 4(9), pp. 1515–1536.

doi:10.3390/v4091515.

Milazzo, A.-C. *et al.* (2005) ‘Active pixel sensor array as a detector for electron microscopy’, *Ultramicroscopy*, 104(2), pp. 152–159. doi:10.1016/j.ultramic.2005.03.006.

Moeller, A. *et al.* (2012) ‘Organization of the Influenza Virus Replication Machinery’, *Science*, 338(6114), pp. 1631–1634. doi:10.1126/science.1227270.

Mohl, B.-P. and Barr, J.N. (2009) ‘Investigating the specificity and stoichiometry of RNA binding by the nucleocapsid protein of Bunyamwera virus’, *RNA*, 15(3), pp. 391–399. doi:10.1261/rna.1367209.

Morin, B. *et al.* (2010) ‘The N-Terminal Domain of the Arenavirus L Protein Is an RNA Endonuclease Essential in mRNA Transcription’, *PLoS Pathogens*, 6(9). doi:10.1371/journal.ppat.1001038.

Müller, R. *et al.* (1994) ‘Rift valley fever virus L segment: correction of the sequence and possible functional role of newly identified regions conserved in RNA-dependent polymerases’, *Journal of General Virology*, 75(6), pp. 1345–1352. doi:10.1099/0022-1317-75-6-1345.

Murakami, K. and Trakselis, M.A. (eds) (2014) *Nucleic Acid Polymerases*. Berlin Heidelberg: Springer-Verlag (Nucleic Acids and Molecular Biology). doi:10.1007/978-3-642-39796-7.

Murphy, F.A., Harrison, A.K. and Whitfield, S.G. (1973) ‘Bunyaviridae: Morphologic and Morphogenetic Similarities of Bunyamwera Serologic Supergroup Viruses and Several Other Arthropod-Borne Viruses’, *Intervirology*, 1(4), pp. 297–316. doi:10.1159/000148858.

Nakane, T. *et al.* (2018) ‘Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION’, *eLife*. Edited by A.T. Brunger, 7, p. e36861. doi:10.7554/eLife.36861.

Nakane, T. *et al.* (2020) ‘Single-particle cryo-EM at atomic resolution’, *Nature*, 587(7832), pp. 152–156. doi:10.1038/s41586-020-2829-0.

Nasirian, H. (2020) ‘New aspects about Crimean-Congo hemorrhagic fever

(CCHF) cases and associated fatality trends: A global systematic review and meta-analysis', *Comparative Immunology, Microbiology and Infectious Diseases*, 69, p. 101429. doi:10.1016/j.cimid.2020.101429.

Naydenova, K. *et al.* (2019) 'CryoEM at 100 keV: a demonstration and prospects', *IUCrJ*, 6(6), pp. 1086–1098. doi:10.1107/S2052252519012612.

Naydenova, K., Jia, P. and Russo, C.J. (2020) 'Cryo-EM with sub-1 Å specimen movement', *Science*, 370(6513), pp. 223–226. doi:10.1126/science.abb7927.

Ng, A.K.-L. *et al.* (2012) 'Structural Basis for RNA Binding and Homo-Oligomer Formation by Influenza B Virus Nucleoprotein', *Journal of Virology*, 86(12), pp. 6758–6767. doi:10.1128/JVI.00073-12.

Noda, T. *et al.* (2006) 'Architecture of ribonucleoprotein complexes in influenza A virus particles', *Nature*, 439(7075), pp. 490–492. doi:10.1038/nature04378.

Noton, S.L. *et al.* (2009) 'Studies of an Influenza A Virus Temperature-Sensitive Mutant Identify a Late Role for NP in the Formation of Infectious Virions', *Journal of Virology*, 83(2), pp. 562–571. doi:10.1128/JVI.01424-08.

Olal, D. *et al.* (2014) 'Structural insights into RNA encapsidation and helical assembly of the Toscana virus nucleoprotein', *Nucleic Acids Research*, 42(9), pp. 6025–6037. doi:10.1093/nar/gku229.

Olal, D. and Daumke, O. (2016) 'Structure of the Hantavirus Nucleoprotein Provides Insights into the Mechanism of RNA Encapsidation', *Cell Reports*, 14(9), pp. 2092–2099. doi:10.1016/j.celrep.2016.02.005.

Olschewski, S., Cusack, S. and Rosenthal, M. (2020) 'The Cap-Snatching Mechanism of Bunyaviruses', *Trends in Microbiology*, 28(4), pp. 293–303. doi:10.1016/j.tim.2019.12.006.

Orlova, E.V. and Saibil, H.R. (2011) 'Structural Analysis of Macromolecular Assemblies by Electron Microscopy', *Chemical Reviews*, 111(12), pp. 7710–7748. doi:10.1021/cr100353t.

Överby, A.K. *et al.* (2008) 'Insights into bunyavirus architecture from electron cryotomography of Uukuniemi virus', *Proceedings of the National Academy of Sciences*, 105(7), pp. 2375–2379. doi:10.1073/pnas.0708738105.

Överby, A.K., Pettersson, R.F. and Neve, E.P.A. (2007) 'The Glycoprotein Cytoplasmic Tail of Uukuniemi Virus (Bunyaviridae) Interacts with Ribonucleoproteins and Is Critical for Genome Packaging', *Journal of Virology*, 81(7), pp. 3198–3205. doi:10.1128/JVI.02655-06.

Oymans, J. and Te Velthuis, A.J.W. (2018) 'A Mechanism for Priming and Realignment during Influenza A Virus Replication', *Journal of Virology*, 92(3). doi:10.1128/JVI.01773-17.

Palovcak, E. *et al.* (2018) 'A simple and robust procedure for preparing graphene-oxide cryo-EM grids', *Journal of Structural Biology*, 204(1), pp. 80–84. doi:10.1016/j.jsb.2018.07.007.

Papageorgiou, N. *et al.* (2020) 'Brothers in Arms: Structure, Assembly and Function of Arenaviridae Nucleoprotein', *Viruses*, 12(7). doi:10.3390/v12070772.

Payne, S. (2017) 'Introduction to DNA Viruses', in *Viruses*. Elsevier, pp. 231–236. doi:10.1016/B978-0-12-803109-4.00028-3.

Peacock, T.P. *et al.* (2019) 'Host Determinants of Influenza RNA Synthesis', *Annual Review of Virology*, 6(1), pp. 215–233. doi:10.1146/annurev-virology-092917-043339.

Peet, M.J., Henderson, R. and Russo, C.J. (2019) 'The energy dependence of contrast and damage in electron cryomicroscopy of biological molecules', *Ultramicroscopy*, 203, pp. 125–131. doi:10.1016/j.ultramic.2019.02.007.

Peng, Q. *et al.* (2019) 'Structural insight into RNA synthesis by influenza D polymerase', *Nature Microbiology*, 4(10), pp. 1750–1759. doi:10.1038/s41564-019-0487-5.

Peng, R. *et al.* (2020) 'Structural insight into arenavirus replication machinery', *Nature*, 579(7800), pp. 615–619. doi:10.1038/s41586-020-2114-2.

Pflug, A. *et al.* (2014) 'Structure of influenza A polymerase bound to the viral

RNA promoter', *Nature*, 516(7531), pp. 355–360. doi:10.1038/nature14008.

Pflug, A. *et al.* (2017) 'Structural insights into RNA synthesis by the influenza virus transcription-replication machine', *Virus Research*, 234, pp. 103–117. doi:10.1016/j.virusres.2017.01.013.

Pichlmair, A. *et al.* (2006) 'RIG-I-Mediated Antiviral Responses to Single-Stranded RNA Bearing 5'-Phosphates', *Science*, 314(5801), pp. 997–1001. doi:10.1126/science.1132998.

Plotch, S.J. and Krug, R.M. (1977) 'Influenza virion transcriptase: synthesis in vitro of large, polyadenylic acid-containing complementary RNA.', *Journal of Virology*, 21(1), pp. 24–34. doi:10.1128/JVI.21.1.24-34.1977.

Poch, O. *et al.* (1989) 'Identification of four conserved motifs among the RNA-dependent polymerase encoding elements', *The EMBO journal*, 8(12), pp. 3867–3874.

Poch, O. *et al.* (1990) 'Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains', *The Journal of General Virology*, 71 (Pt 5), pp. 1153–1162. doi:10.1099/0022-1317-71-5-1153.

Poon, L.L.M. *et al.* (1999) 'Direct Evidence that the Poly(A) Tail of Influenza A Virus mRNA Is Synthesized by Reiterative Copying of a U Track in the Virion RNA Template', *Journal of Virology*, 73(4), pp. 3473–3476. doi:10.1128/JVI.73.4.3473-3476.1999.

Punch, E.K. *et al.* (2018) 'Potassium is a trigger for conformational change in the fusion spike of an enveloped RNA virus', *The Journal of Biological Chemistry*, 293(26), pp. 9937–9944. doi:10.1074/jbc.RA118.002494.

Punjani, A. *et al.* (2017) 'cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination', *Nature Methods*, 14(3), pp. 290–296. doi:10.1038/nmeth.4169.

Punjani, A. and Fleet, D.J. (2021) '3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM', *Journal of Structural Biology*, 213(2), p. 107702. doi:10.1016/j.jsb.2021.107702.

Qi, X. *et al.* (2010) ‘Cap binding and immune evasion revealed by Lassa nucleoprotein structure’, *Nature*, 468(7325), pp. 779–783. doi:10.1038/nature09605.

Radermacher, M. *et al.* (1987) ‘Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of *Escherichia coli*’, *Journal of Microscopy*, 146(2), pp. 113–136. doi:10.1111/j.1365-2818.1987.tb01333.x.

Raju, R. and Kolakofsky, D. (1986) ‘Inhibitors of protein synthesis inhibit both La Crosse virus S-mRNA and S genome syntheses in vivo’, *Virus Research*, 5(1), pp. 1–9. doi:10.1016/0168-1702(86)90061-4.

Ramachandran, G.N. and Venkatachalam, C.M. (1970) ‘A new derivation of the Fourier transform of a helical structure’, *Zeitschrift für Kristallographie - Crystalline Materials*, 132(1–6), pp. 152–156. doi:10.1524/zkri.1970.132.16.152.

Ramírez-Aportela, E. *et al.* (2020) ‘Automatic local resolution-based sharpening of cryo-EM maps’, *Bioinformatics*, 36(3), pp. 765–772. doi:10.1093/bioinformatics/btz671.

Randall, R.E. and Goodbourn, S. (2008) ‘Interferons and viruses: an interplay between induction, signalling, antiviral responses and virus countermeasures’, *The Journal of General Virology*, 89(Pt 1), pp. 1–47. doi:10.1099/vir.0.83391-0.

Ravelli, R.B.G. *et al.* (2020) ‘Cryo-EM structures from sub-nl volumes using pin-printing and jet vitrification’, *Nature Communications*, 11(1), p. 2563. doi:10.1038/s41467-020-16392-5.

Raymond, D.D. *et al.* (2010) ‘Structure of the Rift Valley fever virus nucleocapsid protein reveals another architecture for RNA encapsidation’, *Proceedings of the National Academy of Sciences*, 107(26), pp. 11769–11774. doi:10.1073/pnas.1001760107.

Raymond, D.D. *et al.* (2012) ‘Phleboviruses encapsidate their genomes by sequestering RNA bases’, *Proceedings of the National Academy of Sciences*, 109(47), pp. 19208–19213.

Reguera, J. *et al.* (2013) 'Structural basis for encapsidation of genomic RNA by La Crosse Orthobunyavirus nucleoprotein', *Proceedings of the National Academy of Sciences*, 110(18), pp. 7246–7251. doi:10.1073/pnas.1302298110.

Reguera, J. *et al.* (2016) 'Comparative Structural and Functional Analysis of Bunyavirus and Arenavirus Cap-Snatching Endonucleases', *PLOS Pathogens*, 12(6), p. e1005636. doi:10.1371/journal.ppat.1005636.

Reguera, J., Cusack, S. and Kolakofsky, D. (2014) 'Segmented negative strand RNA virus nucleoprotein structure', *Current Opinion in Virology*, 5, pp. 7–15. doi:10.1016/j.coviro.2014.01.003.

Reguera, J., Weber, F. and Cusack, S. (2010) 'Bunyaviridae RNA polymerases (L-protein) have an N-terminal, influenza-like endonuclease domain, essential for viral cap-dependent transcription', *PLoS pathogens*, 6(9), p. e1001101. doi:10.1371/journal.ppat.1001101.

Reich, S. *et al.* (2014) 'Structural insight into cap-snatching and RNA synthesis by influenza polymerase', *Nature*, 516(7531), pp. 361–366. doi:10.1038/nature14009.

Renaud, J.-P. *et al.* (2018) 'Cryo-EM in drug discovery: achievements, limitations and prospects', *Nature Reviews Drug Discovery*, 17(7), pp. 471–492. doi:10.1038/nrd.2018.77.

Robb, N.C. *et al.* (2016) 'Single-molecule FRET reveals the pre-initiation and initiation conformations of influenza virus promoter RNA', *Nucleic Acids Research*, 44(21), pp. 10304–10315. doi:10.1093/nar/gkw884.

Rodriguez, C. *et al.* (2020) 'Fatal Encephalitis Caused by Cristoli Virus, an Emerging Orthobunyavirus, France - Volume 26, Number 6—June 2020 - Emerging Infectious Diseases journal - CDC'. doi:10.3201/eid2606.191431.

Rohou, A. and Grigorieff, N. (2014) 'FREALIX: Model-based refinement of helical filament structures from electron micrographs', *Journal of Structural Biology*, 186(2), pp. 234–244. doi:10.1016/j.jsb.2014.03.012.

Rohou, A. and Grigorieff, N. (2015) 'CTFFIND4: Fast and accurate defocus estimation from electron micrographs', *Journal of Structural Biology*, 192(2), pp.

216–221. doi:10.1016/j.jsb.2015.08.008.

Rosenthal, M. *et al.* (2017) ‘Structural insights into reptarenavirus capsid assembly machinery’, *PLoS pathogens*, 13(5), p. e1006400. doi:10.1371/journal.ppat.1006400.

Rosenthal, P.B. and Henderson, R. (2003) ‘Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy’, *Journal of Molecular Biology*, 333(4), pp. 721–745.

Rosenthal, P.B. and Rubinstein, J.L. (2015) ‘Validating maps from single particle electron cryomicroscopy’, *Current Opinion in Structural Biology*, 34, pp. 135–144. doi:10.1016/j.sbi.2015.07.002.

Rossier, C., Patterson, J. and Kolakofsky, D. (1986) ‘La Crosse virus small genome mRNA is made in the cytoplasm.’, *Journal of Virology*, 58(2), pp. 647–650.

Rossier, C., Raju, R. and Kolakofsky, D. (1988) ‘LaCrosse virus gene expression in mammalian and mosquito cells’, *Virology*, 165(2), pp. 539–548. doi:10.1016/0042-6822(88)90598-3.

Rotenberg, D. *et al.* (2015) ‘Thrips transmission of tospoviruses’, *Current Opinion in Virology*, 15, pp. 80–89. doi:10.1016/j.coviro.2015.08.003.

Rubinstein, J.L. *et al.* (2019) ‘Shake-it-off: a simple ultrasonic cryo-EM specimen-preparation device’, *Acta Crystallographica Section D: Structural Biology*, 75(12), pp. 1063–1070. doi:10.1107/S2059798319014372.

Ruigrok, R.W., Crépin, T. and Kolakofsky, D. (2011) ‘Nucleoproteins and nucleocapsids of negative-strand RNA viruses’, *Current Opinion in Microbiology*, 14(4), pp. 504–510. doi:10.1016/j.mib.2011.07.011.

Russo, C.J. and Passmore, L.A. (2014) ‘Ultrastable gold substrates for electron cryomicroscopy’, *Science*, 346(6215), pp. 1377–1380. doi:10.1126/science.1259530.

Russo, C.J. and Passmore, L.A. (2016) ‘Progress towards an optimal specimen support for electron cryomicroscopy’, *Current Opinion in Structural Biology*, 37, pp. 81–89. doi:10.1016/j.sbi.2015.12.007.

Sanchez-Garcia, R. *et al.* (2020) 'DeepEMhancer: a deep learning solution for cryo-EM volume post-processing', *bioRxiv*, p. 2020.06.12.148296. doi:10.1101/2020.06.12.148296.

Scheres, S.H. (2014) 'Beam-induced motion correction for sub-megadalton cryo-EM particles', *eLife*. Edited by W. Kühlbrandt, 3, p. e03665. doi:10.7554/eLife.03665.

Scheres, S.H.W. (2010) 'Maximum-likelihood methods in cryo-EM. Part II: application to experimental data', *Methods in enzymology*, 482, pp. 295–320. doi:10.1016/S0076-6879(10)82012-9.

Scheres, Sjors H. W. (2012) 'A Bayesian View on Cryo-EM Structure Determination', *Journal of Molecular Biology*, 415(2), pp. 406–418. doi:10.1016/j.jmb.2011.11.010.

Scheres, Sjors H.W. (2012) 'RELION: Implementation of a Bayesian approach to cryo-EM structure determination', *Journal of Structural Biology*, 180(3), pp. 519–530. doi:10.1016/j.jsb.2012.09.006.

Scheres, S. H. W. (2016) 'Chapter Six - Processing of Structurally Heterogeneous Cryo-EM Data in RELION', in Crowther, R.A. (ed.) *Methods in Enzymology*. Academic Press (The Resolution Revolution: Recent Advances In cryoEM), pp. 125–157. doi:10.1016/bs.mie.2016.04.012.

Scheres, S.H.W. (2016) 'Processing of Structurally Heterogeneous Cryo-EM Data in RELION', in *Methods in Enzymology*. Elsevier, pp. 125–157. doi:10.1016/bs.mie.2016.04.012.

Scheres, S.H.W. and Chen, S. (2012) 'Prevention of overfitting in cryo-EM structure determination', *Nature Methods*, 9(9), pp. 853–854. doi:10.1038/nmeth.2115.

Scholte, F.E.M. *et al.* (2017) 'Crimean-Congo Hemorrhagic Fever Virus Suppresses Innate Immune Responses via a Ubiquitin and ISG15 Specific Protease', *Cell Reports*, 20(10), pp. 2396–2407. doi:10.1016/j.celrep.2017.08.040.

Schorb, M. *et al.* (2019) 'Software tools for automated Transmission Electron Microscopy', *Nature methods*, 16(6), pp. 471–477. doi:10.1038/s41592-019-0396-9.

Schreur, P.J.W. and Kortekaas, J. (2016) 'Single-Molecule FISH Reveals Non-selective Packaging of Rift Valley Fever Virus Genome Segments', *PLOS Pathogens*, 12(8), p. e1005800. doi:10.1371/journal.ppat.1005800.

Serna, M. (2019) 'Hands on Methods for High Resolution Cryo-Electron Microscopy Structures of Heterogeneous Macromolecular Complexes', *Frontiers in Molecular Biosciences*, 6. doi:10.3389/fmolb.2019.00033.

Serris, A. *et al.* (2020) 'The Hantavirus Surface Glycoprotein Lattice and Its Fusion Control Mechanism', *Cell*, 183(2), pp. 442-456.e16. doi:10.1016/j.cell.2020.08.023.

Sgro, G.G. and Costa, T.R.D. (2018) 'Cryo-EM Grid Preparation of Membrane Protein Samples for Single Particle Analysis', *Frontiers in Molecular Biosciences*, 5. doi:10.3389/fmolb.2018.00074.

Shaikh, T.R. *et al.* (2008) 'SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs', *Nature protocols*, 3(12), pp. 1941–1974. doi:10.1038/nprot.2008.156.

Sholders, A.J. and Peersen, O.B. (2014) 'Distinct conformations of a putative translocation element in poliovirus polymerase', *Journal of Molecular Biology*, 426(7), pp. 1407–1419. doi:10.1016/j.jmb.2013.12.031.

Smithburn, K.C., Haddow, A.J. and Mahaffy, A.F. (1946) 'A Neurotropic Virus Isolated from Aedes Mosquitoes Caught in the Semliki Forest', *The American Journal of Tropical Medicine and Hygiene*, s1-26(2), pp. 189–208. doi:10.4269/ajtmh.1946.s1-26.189.

Stewart, A. and Grigorieff, N. (2004) 'Noise bias in the refinement of structures derived from single particles', *Ultramicroscopy*, 102(1), pp. 67–84. doi:10.1016/j.ultramic.2004.08.008.

Stewart, M. (1988) 'Computer image processing of electron micrographs of biological structures with helical symmetry', *Journal of Electron Microscopy Technique*, 9(4), pp. 325–358. doi:10.1002/jemt.1060090404.

Suda, Y. *et al.* (2016) 'Analysis of the entry mechanism of Crimean-Congo hemorrhagic fever virus, using a vesicular stomatitis virus pseudotyping system',

Archives of Virology, 161(6), pp. 1447–1454. doi:10.1007/s00705-016-2803-1.

Sun, Y. *et al.* (2018) ‘Bunyavirales ribonucleoproteins: the viral replication and transcription machinery’, *Critical Reviews in Microbiology*, 44(5), pp. 522–540. doi:10.1080/1040841X.2018.1446901.

Sutherland, L.J., Anyamba, A. and LaBeaud, A.D. (2013) ‘Emerging and Reemerging Human Bunyavirus Infections and Climate Change’, in *Viral Infections and Global Change*. John Wiley & Sons, Ltd, pp. 477–493. doi:10.1002/9781118297469.ch26.

Talmon, Y. *et al.* (1987) ‘Electron microscopy of vitrified-hydrated La Crosse virus.’, *Journal of Virology*, 61(7), pp. 2319–2321.

Tanaka, null *et al.* (1999) ‘A new 200 kV Omega-filter electron microscope’, *Journal of Microscopy*, 194(1), pp. 219–227. doi:10.1046/j.1365-2818.1999.00446.x.

Tang, G. *et al.* (2007) ‘EMAN2: An extensible image processing suite for electron microscopy’, *Journal of Structural Biology*, 157(1), pp. 38–46. doi:10.1016/j.jsb.2006.05.009.

Tang, Y.-S. *et al.* (2021) ‘Crystal structures of influenza nucleoprotein complexed with nucleic acid provide insights into the mechanism of RNA interaction’, *Nucleic Acids Research*, 49(7), pp. 4144–4154. doi:10.1093/nar/gkab203.

Te Velhuis, A.J.W. *et al.* (2016) ‘The role of the priming loop in Influenza A virus RNA synthesis’, *Nature Microbiology*, 1(5). doi:10.1038/nmicrobiol.2016.29.

Te Velhuis, A.J.W. and Oymans, J. (2018) ‘Initiation, Elongation, and Realignment during Influenza Virus mRNA Synthesis’, *Journal of Virology*, 92(3). doi:10.1128/JVI.01775-17.

Tegunov, D. and Cramer, P. (2019) ‘Real-time cryo-electron microscopy data preprocessing with Warp’, *Nature Methods*, 16(11), pp. 1146–1152. doi:10.1038/s41592-019-0580-y.

Terasaki, K. *et al.* (2011) ‘Mechanism of tripartite RNA genome packaging in Rift Valley fever virus’, *Proceedings of the National Academy of Sciences of the*

United States of America, 108(2), pp. 804–809. doi:10.1073/pnas.1013155108.

Thierry, E. *et al.* (2016) ‘Influenza Polymerase Can Adopt an Alternative Configuration Involving a Radical Repacking of PB2 Domains’, *Molecular Cell*, 61(1), pp. 125–137. doi:10.1016/j.molcel.2015.11.016.

Typke, D. *et al.* (2005) ‘High-throughput film-densitometry: an efficient approach to generate large data sets’, *Journal of Structural Biology*, 149(1), pp. 17–29. doi:10.1016/j.jsb.2004.09.003.

Van Heel, M. (1987) ‘Angular reconstitution: A posteriori assignment of projection directions for 3D reconstruction’, *Ultramicroscopy*, 21(2), pp. 111–123. doi:10.1016/0304-3991(87)90078-7.

te Velthuis, A.J.W. (2014) ‘Common and unique features of viral RNA-dependent polymerases’, *Cellular and molecular life sciences: CMLS*, 71(22), pp. 4403–4420. doi:10.1007/s00018-014-1695-z.

Venkataraman, S., Prasad, B.V.L.S. and Selvarajan, R. (2018) ‘RNA Dependent RNA Polymerases: Insights from Structure, Function and Evolution’, *Viruses*, 10(2). doi:10.3390/v10020076.

Vinothkumar, K.R. and Henderson, R. (2016) ‘Single particle electron cryomicroscopy: trends, issues and future perspective’, *Quarterly Reviews of Biophysics*, 49. doi:10.1017/S0033583516000068.

Wade, R.H. (1992) ‘A brief look at imaging and contrast transfer’, *Ultramicroscopy*, 46(1–4), pp. 145–156. doi:10.1016/0304-3991(92)90011-8.

Wallat, G.D. *et al.* (2014) ‘High-Resolution Structure of the N-Terminal Endonuclease Domain of the Lassa Virus L Polymerase in Complex with Magnesium Ions’, *PLOS ONE*, 9(2), p. e87577. doi:10.1371/journal.pone.0087577.

Wandzik, J.M. *et al.* (2020) ‘A Structure-Based Model for the Complete Transcription Cycle of Influenza Polymerase’, *Cell*, 181(4), pp. 877–893.e21. doi:10.1016/j.cell.2020.03.061.

Wang, Panpan *et al.* (2020) ‘Structure of severe fever with thrombocytopenia syndrome virus L protein elucidates the mechanisms of viral transcription

initiation', *Nature Microbiology*, 5(6), pp. 864–871. doi:10.1038/s41564-020-0712-2.

Weissenberger, G., Henderikx, R.J.M. and Peters, P.J. (2021) 'Understanding the invisible hands of sample preparation for cryo-EM', *Nature Methods*, 18(5), pp. 463–471. doi:10.1038/s41592-021-01130-6.

Welsch, S., Müller, B. and Kräusslich, H.-G. (2007) 'More than one door – Budding of enveloped viruses through cellular membranes', *Febs Letters*, 581(11), pp. 2089–2097. doi:10.1016/j.febslet.2007.03.060.

White, H.D. *et al.* (2003) 'A second generation apparatus for time-resolved electron cryo-microscopy using stepper motors and electrospray', *Journal of Structural Biology*, 144(1–2), pp. 246–252. doi:10.1016/j.jsb.2003.09.027.

Wu, C. *et al.* (2019) 'High-quality, high-throughput cryo-electron microscopy data collection via beam tilt and astigmatism-free beam-image shift', *Journal of Structural Biology*, 208(3), p. 107396. doi:10.1016/j.jsb.2019.09.013.

Yang, X. *et al.* (2012) 'Motif D of Viral RNA-Dependent RNA Polymerases Determines Efficiency and Fidelity of Nucleotide Addition', *Structure*, 20(9), pp. 1519–1527. doi:10.1016/j.str.2012.06.012.

Ye, Q., Krug, R.M. and Tao, Y.J. (2006) 'The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA', *Nature*, 444(7122), pp. 1078–1082. doi:10.1038/nature05379.

Yoneyama, M. and Fujita, T. (2008) 'Structural Mechanism of RNA Recognition by the RIG-I-like Receptors', *Immunity*, 29(2), pp. 178–181. doi:10.1016/j.immuni.2008.07.009.

Yu, X.-J. *et al.* (2011) 'Fever with Thrombocytopenia Associated with a Novel Bunyavirus in China', *New England Journal of Medicine*, 364(16), pp. 1523–1532. doi:10.1056/NEJMoa1010095.

Yuan, P. *et al.* (2009) 'Crystal structure of an avian influenza polymerase PA N reveals an endonuclease active site', *Nature*, 458(7240), pp. 909–913. doi:10.1038/nature07720.

Zhang, K. (2016) ‘Gctf: Real-time CTF determination and correction’, *Journal of Structural Biology*, 193(1), pp. 1–12. doi:10.1016/j.jsb.2015.11.003.

Zheng, S.Q. *et al.* (2017) ‘MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy’, *Nature Methods*, 14(4), pp. 331–332. doi:10.1038/nmeth.4193.

Zheng, W. and Tao, Y.J. (2013) ‘Structure and assembly of the influenza A virus ribonucleoprotein complex’, *FEBS Letters*, 587(8), pp. 1206–1214. doi:https://doi.org/10.1016/j.febslet.2013.02.048.

Zhu, D. *et al.* (2018) ‘Pushing the resolution limit by correcting the Ewald sphere effect in single-particle Cryo-EM reconstructions’, *Nature Communications*, 9(1), p. 1552. doi:10.1038/s41467-018-04051-9.

Zivanov, J. *et al.* (2018) ‘New tools for automated high-resolution cryo-EM structure determination in RELION-3’, *eLife*. Edited by E.H. Egelman and J. Kuriyan, 7, p. e42166. doi:10.7554/eLife.42166.

Zivanov, J., Nakane, T. and Scheres, S.H.W. (2019) ‘A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis’, *IUCrJ*, 6(1), pp. 5–17. doi:10.1107/S205225251801463X.

Zivanov, J., Nakane, T. and Scheres, S.H.W. (2020) ‘Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1’, *IUCrJ*, 7(2), pp. 253–267. doi:10.1107/S2052252520000081.

